**Instructions**

Please complete the following lab assignment. You may work on the assignment in groups or on your own. However, to get credit, you must submit your own answers in Canvas. This lab is open note and open book. You may also ask the instructor and the TA questions. Please note that in most cases we will try to guide you towards answering your own question rather than directly providing you with an answer.

# QX. Feedback

Placeholder

Please make sure you understand why this is the correct answer. You may use the "Previous" button below to update your answer if your original answer was incorrect.

Click the "Next" button below to move on to the next question.

# Last Question Feedback

Please make sure you understand why the example answer is correct. You may use the "Previous" button to update your answer if you feel like you can now give a more correct/complete answer.

Click the "Submit" button below if you are ready to submit this lab.

# 

**Feedback**

**Optional**: Please feel free to leave any comments below about the usefulness of this lab. Which parts were helpful? What could I do to improve it? What is still unclear?

# Szklo Textbook:

1. In a case-control study of risk factors of colon cancer, 430 cases were compared with 551 controls. The investigators used a questionnaire to obtain information about demographic variables, socioeconomic variables (e.g., education), weight, and height, among other variables. Using the self-reported weight and height information, body mass index [BMI, weight (kg)/height (m) 2 ] values were calculated. Participants with BMI ≥ 30 kg/m 2 were considered “obese.” The association between obesity and colon cancer in this study is shown in the table.

|  | Cases | Controls |
| --- | --- | --- |
| Obese | 162 | 133 |
| Nonobese | 268 | 418 |

a. Calculate the odds ratio relating obesity and colon cancer in this study.

*Observed OR = (162 ÷ 268)/(133 ÷ 418) = 1.9*

Subsequently, the investigators obtained additional funds to conduct a validation study of some of the information obtained from the participants’ interviews. For the validation study, 100 participants (50 cases and 50 controls) were randomly selected and invited to attend a clinic, where diverse objective physical measurements and more extensive questionnaires were used in an attempt to estimate the validity of the self-reported information in the study. Despite intensive efforts for recruitment, only 60 of the 100 participants invited for the validation study agreed to the clinic visit. The participants who agreed to attend included a larger proportion of females and individuals of a higher educational level than those who declined. Using objectively measured weight and height, BMI was recalculated in the 60 individuals in the validation study. Among the individuals who were classified as obese using measured weight and height, 90% of the cases and 95% of the controls had also been classified as obese by the BMI based on self-reported information; 100% of those classified as nonobese using measured weight and height had been classified as such by the self-reported information.

b. Assuming that weight and height values did not change in the time between the interviews and the validation study, calculate the “corrected” odds ratio based on the estimates obtained from the validation study. That is, estimate the odds ratio that would have been obtained if no misclassification of obese status based on self-reported weight and height information had occurred.

*Total number of cases = 162 + 268 = 430*

*Sensitivity of self-report in cases = 0.90.*

*Thus, truly obese = 162/0.90 = 180, and truly nonobese = 430 – 180 = 250.*

*Total number of controls = 133 + 418 = 551*

*Sensitivity of self-report in controls = 0.95. Thus, truly obese = 133/0.95 = 140, and truly nonobese = 551 – 140 = 411.*

*“Corrected” OR = (180 / 250) / (140 /411) = 2.1*

c. How do you explain the difference between the observed and the corrected odds ratio obtained in this study?

*There is differential misclassification of obesity status resulting from the fact that the sensitivity of self-reported obesity is different between cases and controls. (When there is differential misclassification, the OR can be biased in any direction depending on what the patterns of misclassification in cases and controls are—for example, closer to 1.0, as in this case.)*

d. In addition to the need to assume no change in weight and height between interviews and validation assessment, what are, in your judgment, other important limitations of the use of the validation study (vis-à-vis the whole study) to estimate a corrected odds ratio, as you did in answer to Exercise 1b?

*Individuals in the validation study are not representative of the entire study population (e.g., validation study participants included more females and were higher educated, and perhaps they are also different with regard to other variables related to the self-report validity). Thus, it may be inappropriate to generalize the validation results to the entire study population. Another limitation is that validation results are based on a small sample and, therefore, are subject to large sample variability.*

2. It is estimated that about one-third of prostate cancer cases can be present in men in their fourth or fifth decade of life without any clinical symptoms or signs. \* Several observational studies (including cohort studies) have suggested that vasectomy may be related to prostate cancer risk. For example, in a meta-analysis of 5 cohort studies and 17 case-control studies, the pooled relative risk estimate was found to be 1.37 (95% confidence interval, 1.15, 1.62). † In these studies, prostate cancer was not systematically evaluated; rather, usual clinical diagnosis was the diagnostic strategy. Describe the possible threats to validity when inferring that this association is causal.

*Individuals who undergo vasectomy may have better access to health care and, therefore, have their subclinical disease diagnosed more often. In case-control studies, recall bias may occur whereby cases are more likely to recall past vasectomy than controls (although, given the fact that this exposure is objective, this is not likely). Publication bias whereby only positive results are reported in peer-reviewed journals. (Not discussed in this chapter: There may be differences between vasectomized and nonvasectomized men that are relevant to the observed findings, such as level of sexual activity. This would be an example of confounding and is discussed in detail in Chapter 5 .)*

3. Two clinical trials have examined the association of PSA testing with mortality from prostate cancer. In one trial (Schröder et al. ‡ ), the mortality for those tested for PSA was significantly lower than that for the control group at the end of the trial. In the other trial (Andriole et al. § ), a significant difference could not be found between the groups. Why have the authors examined mortality instead of cumulative survival in cases?

*The authors evaluated mortality in the total groups to avoid both lead time bias and overdiagnosis bias (see Chapter 4 , Section 4.4.3 ).*

4. A breast cancer screening program based on repeat free clinical breast examinations was implemented in a developing country for women aged 50–59 years. The program is open to all eligible women, but it is not compulsory.

a. Do you expect the incidence among the women who take advantage of the program to be the same as for the total eligible population? Why? In women who choose to take advantage of the program, the average annual incidence of previously undetected breast cancer is found to be about 100 per 100,000 on follow-up. In the first exam, point prevalence is found to be approximately 200 per 100,000. Assume that, after cases are confirmed by biopsy, no false negatives go undetected.

*No. Women with certain characteristics known to be related to breast cancer are more likely to participate in the program—for example, those with a family history of breast cancer, those with benign breast disease, and those with a higher socioeconomic status (and thus better educational level and health awareness). Incidence in those taking advantage of the program is therefore expected to be higher than in the total population of women aged 50–59 years.*

b. What is the average duration of the detectable preclinical phase in cases of breast cancer detected at the first exam?

*Cases detected at the first exam are point prevalent cases. Given the low point prevalence, the simplified formula expressing the relation between point prevalence and incidence can be used: Prevalence ≈ Incidence × Duration. Thus,*

*Duration ~ Point prevalence / incidence*

*Using the incidence and point prevalence values above:*

*Duration ~ (200/100,000) / (100/100,000) = 2 years*

*Using the more precise, correct formula, Point prevalence = Incidence x Duration x (1 - Point prevalence), and thus,*

*Duration = (Point prevalence) / [Incidence \* (1 - Point Prevalence)]*

*In this example, Duration = 0.002 / [0.001 \* (1 - 0.0020)] = 2.004 years*

*(Note the use of three decimal places for the duration of the detectable clinical phase of breast cancer when using the correct formula aimed at highlighting the fact that, when the prevalence is very low, the duration values using either the correct formula or the simplified formula are virtually identical.)*

c. Define lead time bias in the context of evaluation of a screening program or procedure.

*Lead time is the time between early diagnosis using a screening test (followed by a confirmatory diagnostic test) and the time when the disease would have been diagnosed by usual clinical practice (that is, if screening had not been done).*

d. How does lead time bias affect estimation of average survival time?

*Survival time appears to be longer in those who undergo the screening procedure (above and beyond any possible true benefits brought about by the screening). This longer survival reflects the fact that diagnosis was advanced by the application of the screening test (vis-à-vis when it would have occurred without screening).*

e. Estimate the average lead time for prevalent cases in this example, and state the assumption underlying this estimation.

*The average lead time for point prevalent cases is about one-half of the duration of the detectable preclinical phase (i.e., 2 years ÷ 2 = 1 year). This estimation is based on the assumption that the sensitivity of the test is homogeneous throughout the duration of the detectable preclinical phase.*

f. As the interval between screening exams becomes shorter, what is the tendency of the average lead time value for incident cases that are detected after the initial screening?

*The average lead time for incident cases will increasingly approximate the duration of the detectable preclinical phase when screenings are done more and more frequently. This is because cases are more likely to be detected earlier in the detectable preclinical phase.*

5. The sensitivity of high levels of prostate specific antigen (PSA ≥ 4.0 ng/ml) to identify prostate cancer has ranged from 35% to 71% and specificity from 63% to 91%, with resulting false-positive rates of 20% to 68%. \*\* A cohort study to evaluate the relationship of serum retinol (vitamin A) levels to prostate cancer was conducted by Mondul and collaborators. \*\*\* They included close to 30,000 smoking men aged 50–69 years at baseline. The hazard ratio\*\* (relative risk) of prostate cancer associated with the upper quintile of serum retinol—compared with that of the lowest quintile—was found to be 1.13 as measured in the 3-year follow-up visit. This exercise will assume that this is the true relative risk for the association of the 5th vs the 1st quintile of retinol. The table summarizes some of the study results. For the purposes of this exercise, the number of person-years given in the paper by Mondul et al. was considered as number of persons. \*\*\*

| Incidence rates and relative risks of prostate cancer according to extreme quintiles of serum retinol, 1985–2006, from the Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study. | | | | |
| --- | --- | --- | --- | --- |
| Retinol levels | No. of cases | No. of persons | Incidence/1,000 | Relative Risk |
| Quintile 1 | 312 | 66,010 | 4.73 | 1.00 |
| Quintile 2 | 377 | 70,747 | 5.33 | 1.13 |
| Data from Mondul AM, Watters JL, Männistö S, et al. Serum retinol and risk of prostate cancer. Am J Epidemiol. 2011;173:813-821. | | | | |

a. Assuming sensitivity and specificity of 0.70 and 0.90 for the ascertainment of prostate cancer in this study, and further assuming that these levels are the same for both quintiles, calculate the biased relative risk using the data in the table above as the gold standard (true values).

| Quintile 1 | | | | Quintile 5 | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Study | Gold Standard | | Total | Study | Gold Standard | | Total |
| Disease | No Disease | Disease | No Disease |
| Disease | 218 | 6,570 | 6,788 | Disease | 264 | 7,037 | 7,301 |
| No Disease | 94 | 59,128 | 59,222 | No Disease | 113 | 63,333 | 63,445 |
| Total | 312 | 65,698 | 66,010 | Total | 377 | 70,370 | 70,747 |
| Incidence = 6,788 / 66,010 = 0.103 (10.3%) | | | | Incidence = 7,301 / 70,747 = 0.103 (10.3%) | | | |
| Relative Risk = 1.00 | | | | | | | |

*Illustration for a couple of the calculations:*

*Quintile 1, Disease/Disease cell: 312 × 0.70 = 218*

*Quintile 5, Disease/No disease cell: (70,747 − 377) × (1 − 0.90) = 7037*

b. Define the type of misclassification based on the answer to Exercise 5a.

*Nondifferential*

# Doug’s exercises

BACKGROUND: Non-response bias is a type of selection bias that can occur in case-control studies. Its analogue in prospective studies is loss-to-follow-up bias. The following hypothetical data are intended to illustrate the circumstances under which bias will be introduced in the odds ratio because of non-response bias.

Reminder: The approximate 95% confidence interval of the natural log of the OR is:

ln OR ± 1.96 Ö (1/a + 1/b + 1/c + 1/d)

**PART 1 – EQUAL RESPONSE PROPORTIONS IN CASES AND CONTROLS**

1. Suppose that we are doing a case-control study in which we select 400 cases of multiple sclerosis and 400 controls. We collected our exposure information by personal interview and had an 80% response percentage in each group. Based on interview data, the percentage of cases exposed to multiple viral infections was 30% and the percentage of controls exposed was 15%.

a. Construct the appropriate 2x2 table to represent these data and calculate the odds ratio and 95% confidence interval associated with exposure. How would you interpret the results of this study?

|  |  | Cases | Controls |
| --- | --- | --- | --- |
|  | + | (320 x 0.3=) 96 | (320 x 0.15 =) 48 |
| Exposed |  |  |  |
|  | \_ | 224 | 272 |
|  |  | 320 | 320 |

OR = 2.43 (95% CI = 1.65-3.58)

% exposed cases = 96/320 x 100 = 30%

% exposed controls = 48/320 x 100 = 15%

The odds of having had multiple viral infections were 2.43 times higher in people with multiple sclerosis than in people without multiple sclerosis. Or

There was a significant, 2.43-fold increase in the odds of a history of multiple viral infections in those with multiple sclerosis than in those without multiple sclerosis.

b. In order to estimate the impact of non-response on our study results, we selected a 50% random sample of non-responders and made intensive efforts to obtain exposure information from this group. The following data represent the information regarding exposure history obtained from the random sample of non-responders (assume we got them all). Calculate the percent exposed in cases and controls and the odds ratio associated with exposure.

Cases Controls

+ 12 6

Exposed OR=2.43

- 28 34 (0.81-7.30)

40 40

% exposed cases = 12/40 x 100 = 30%

% exposed controls = 6/40 x 100 = 15%

Note that the 95% CI now includes 1.0 even though the OR is the same as for the respondents. This is because of the smaller sample size, which makes this estimate LESS precise.

c. Based on the results in parts 1.a. and b. above, is there bias in the estimate of association that you reported in your study (i.e., that based only on observed data from the responders)? Explain your answer.

No, because the exposure frequencies in non-respondents were the same as for respondents in both cases and controls. Thus, no bias was introduced by the fact that 20% did not respond since the association of disease status and exposure was the same in non-respondents as in respondents.

Using the same 2x2 table you generated in part 1.a. above (i.e., the study results as you observed them based on those who took part in your study), respond to the following questions.

d. Suppose that in this example, we again selected a 50% random sample of non-responders and made intensive efforts to obtain exposure information from this group. The following data represent the information regarding exposure history obtained from the random sample of non-responders. Calculate the percent exposed in cases and controls and the odds ratio associated with exposure.

Cases Controls

+ 20 7

Exposed OR=4.71

- 20 33 (1.69-13.13)

40 40

% exposed cases = 20/40 x 100 = 50%

% exposed controls = 7/40 x 100 = 18%

Note that the 95% CI is very wide (as it is above) because of the small sample size. The estimate is very imprecise.

e. Based on the results in parts 1.a. and d. above, is there bias in the estimate of association that you reported in your study (i.e., that based only on observed data from the responders)? Explain your answer. How could you use the information from your follow-up with non-respondents to obtain a better estimate of association in your study?

Yes, because non-respondent cases were more likely to be exposed than respondent cases. Thus, using the observed data (the data from the respondents) you would have under-estimated exposure frequency in the cases, thus under-estimating the odds ratio, assuming the exposure frequency in the controls was about the same.

Because you have a random sample estimate of exposure frequency in non-respondents, you could use these estimates to estimate exposure in all non-respondents and then add them back into the total group as follows:

Case non-respondents = 80 x .50 (exposure frequency in non-respondents) = 40 exposed. 80 total – 40 exposed = 40 unexposed

Control non-respondents = 80 x .18 (exposure frequency in non-respondents) = 14 exposed. 80 total – 14 exposed = 66 unexposed

If we add this information back into the original respondents (table 1) we get the following:

|  |  | Cases | Controls |
| --- | --- | --- | --- |
|  | + | 96 + 40 = 136 | 48 + 14 = 62 |
| Exposed |  |  |  |
|  | \_ | 224 + 40 = 264 | 272 + 66 = 338 |
|  |  | 320 + 80 = 400 | 320 + 80 = 400 |

OR = (136 x 338) ÷ (62 x 264) = 2.81 (95% CI = 2.00-3.95)

As you can see, the corrected measure of association is slightly larger than the original based only on responded. This is expected given the higher OR in the non-respondents.

**PART II – UNEQUAL RESPONSE PROPORTIONS IN CASES AND CONTROLS**

2. Suppose that we are doing another case-control study in which we select 400 cases of multiple sclerosis and 400 controls. We collected our exposure information by personal interview and had a 75% response percentage in cases and an 80% response percentage in controls. Based on interview data, the percentage of cases exposed to multiple viral infections was 30% and the percentage of controls was 15%.

a. Construct the appropriate 2x2 table to represent these data and calculate the odds ratio and 95% confidence interval associated with exposure. How would you interpret the results of this study?

|  |  | Cases | Controls |
| --- | --- | --- | --- |
|  | + | 90 | 48 |
| Exposed |  |  |  |
|  | \_ | 210 | 272 |
|  |  | 300 | 320 |

OR = 2.43 (95% CI = 1.64-3.60)

% exposed cases = 90/300 x 100 = 30%

% exposed controls = 48/320 x 100 = 15%

The odds of having had multiple viral infections were 2.43 times higher in people with multiple sclerosis than in people without multiple sclerosis. Or

There was a significant, 2.43-fold increase in the odds of a history of multiple viral infections in those with multiple sclerosis than in those without multiple sclerosis.

b. In order to estimate the impact of non-response on our study results, we selected a 50% random sample of non-responders and made intensive efforts to obtain exposure information from this group. The following data represent the information regarding exposure history obtained from the random sample of non-responders. Calculate the percent exposed in cases and controls and the odds ratio associated with exposure.

Cases Controls

+ 15 6

Exposed OR = 2.43

- 35 34 (0.84-6.99)

50 40

% exposed cases = 15/50 x 100 = 30%

% exposed controls = 6/40 x 100 = 15%

c. Based on the results in parts 2.a. and b. above, is there bias in the estimate of association that you reported in your study (i.e., that based only on data observed from the responders)? Explain your answer.

No, because the exposure frequencies in non-respondents were the same as for respondents in both cases and controls. Thus, no bias was introduced even though there were different response percentages in cases and controls. There is no bias since the association of disease status and exposure was the same in non-respondents as in respondents.

Using the same 2x2 table you generated in part 2.a. above (i.e., the study results as you observed them, based on those who took part in your study), respond to the following questions.

d. Suppose that in this example, we again selected a 50% random sample of non-responders and made intensive efforts to obtain exposure information from this group. The following data represent the information regarding exposure history obtained from the random sample of non-responders. Calculate the percent exposed in cases and controls and the odds ratio associated with exposure.

Cases Controls

+ 28 8

Exposed

- 22 32

50 40

OR = 5.09 (1.96-13.23)

% exposed cases = 28/50 x 100 = 56%

% exposed controls = 8/40 x 100 = 20%

e. Based on the results in parts 2.a. and d. above, is there bias in the estimate of association that you reported in your study (i.e., that based only on data observed from the responders)? Explain your answer. How could you use the information from your follow-up with non-respondents to obtain a better estimate of association in your study?

Yes, because non-respondent cases were more likely to be exposed than respondent cases. Thus, using the observed data (the data from the respondents) you would have under-estimated exposure frequency in the cases, thus under-estimating the odds ratio, assuming the exposure frequency in the controls was about the same.

Because you have a random sample estimate of exposure frequency in non-respondents, you could use these estimates to estimate exposure in all non-respondents and then add them back into the total group as follows:

Case non-respondents = 100 x .56 (exposure frequency in non-respondents) = 56 exposed. 100 total – 56 exposed = 44 unexposed

Control non-respondents = 80 x .20 (exposure frequency in non-respondents) = 16 exposed. 80 total – 16 exposed = 64 unexposed

If we add this information back into the original respondents we get the following:

|  |  | Cases | Controls |
| --- | --- | --- | --- |
|  | + | 90 + 56 = 146 | 48 + 16 = 64 |
| Exposed |  |  |  |
|  | \_ | 210 + 44 = 254 | 272 + 64 = 336 |
|  |  | 300 + 100 = 400 | 320 + 80 = 400 |

OR = (146 x 336) ÷ (64 x 254) = 3.02 (95% CI = 2.16-4.22)

As you can see, the corrected measure of association is slightly larger than the original based only on responded. This is expected given the higher OR in the non-respondents.

f. What inference can you make about the presence of non-response bias based on the response percentages alone? NONE

g. What kinds of things might you do as an investigator to determine whether there is or is not non-response bias in your study?

1. select a random sample from non-respondents and obtain exposure history from them

2. compare characteristics of respondents and non-respondents within case and control groups in regard to factors that are known to influence exposure frequency

3. compare ORs for those who were easy to recruit and for those who were finally recruited, but with some difficulty to see if they are similar or different. This might tell you something about the non-respondents.

h. What might you do to minimize or avoid non-response bias?

Increase response percentages to as high as possible (i.e., decrease non-

response)

We might also conduct a “sensitivity” analysis – e.g., use a series of reasonable assumptions re: the percent exposed in non-respondent cases or non-respondent controls and re-calculate the observed odds ratio estimate based on these assumptions to see how much impact they have on the observed odds ratio.

# Doug’s Lab 3

**Lab 3. Selection Biases**

Today’s lab will be working through a various calculations regarding selection biases. These calculations are meant to help you understand what is actually happening mathematically when these biases occur. We will also look at some ways to correct some of these biases in our calculations.

Reminder: The approximate 95% confidence interval of the natural log of the OR is:

ln OR ±1.96 √ (1/a + 1/b + 1/c + 1/d)

**Part 1. Incidence Prevalence Bias**

The Framingham Heart Study was a Population-based prospective cohort study of risk factors for coronary heart disease (CHD) that was initiated in the late 1940's in Framingham, MA. Participants were examined every two years and risk factors for CHD as well as CHD events were tracked. Below is a series of contingency tables from two analyses that were done with the Framingham data. The first shows a cross-sectional estimate of the effect of LDL on the occurrence of CHD 12 years into the study. The second shows a longitudinal examination of the 12 year cumulative incidence since the start of data collection.

Q1. Calculate the population prevalence ratio for the effect of High LDL on having CHD from study 1.

Q2. Calculate the rate ratio for the effect of High LDL on having CHD from study 2.

Q3. Why are they different? Why do we see an effect of high LDL in the cohort study but not the cross-sectional study?

Q4. So how would you go about preventing incidence-prevalence bias in this case?

**Part 2. Non-Response Bias**

Suppose that we are doing a case-control study in which we select 400 cases of multiple sclerosis and 400 controls. We collected our exposure information by personal interview and had an 80% response percentage in each group. Based on interview data, the percentage of cases exposed to multiple viral infections was 30% and the percentage of controls exposed was 15%.

Q5. Construct the appropriate 2x2 table to represent these data and calculate the odds ratio and 95% confidence interval associated with exposure. How would you interpret the results of this study?

Q6. . In order to estimate the impact of non-response on our study results, we selected a 50% random sample of non-responders and made intensive efforts to obtain exposure information from this group. The following data represent the information regarding exposure history obtained from the random sample of non-responders (assume we got them all). Calculate the percent exposed in cases and controls and the odds ratio associated with exposure.

Cases Controls

+ 20 33

Exposed

- 20 37

40 40

Q7. Based on the results above, is there bias in the estimate of association that you reported in your study (i.e., that based only on observed data from the responders)? Explain your answer.

Q8. How could you use the information from your follow-up with non-respondents to obtain a better estimate of association in your study? Keep in mind that your sample of non-responder was a random sample.

# Doug’s Lab 4

**Lab 4. Measurement Error.**

In today’s lab, we will work through the impact of measurement error on epidemiological studies.

**Part 1. Reliability**

Below are the results for a potential self-administered screening test for HPV. We are curious if the self-administered swab is a reliable test and decide to examine the agreement between the self-administered swabs and clinician-administered swabs.

|  | **Self-Collected** | |
| --- | --- | --- |
| **Clinician Collected** | Positive | Negative |
| Positive | 170 | 132 |
| Negative | 128 | 985 |

What is the percent agreement between these measures? Would you consider this a reliable measure?

Is this agreement beyond what would be expected by chance? Calculate the kappa statistic to help answer this question.

**Part 2. Misclassification**

The following are hypothetical examples that show the impact of misclassification of both exposure and disease measures.

*Misclassification of disease status.*

Target Population (no misclassification)

Disease No Disease Incidence Proportion

Exposed 400 600 400/1000=0.4

(N=1,000)

Unexposed 200 800 200/1000=0.2

(N=1,000)

RR = 2.0 (95% CI 1.7-2.3)

Suppose that disease incidence is classified with a sensitivity of 85% in both exposed and unexposed and a specificity of 92% in both exposed and unexposed, and complete the following 2 x 2 tables. Assume no misclassification of exposure. D = true disease positive; no D = true disease negative; D' = study disease positive; no D' = study disease negative

EXPOSED UNEXPOSED

True Status True Status

D no D D no D

Study D' Study D'

Status Status

no D' no D'

Using the data from these tables, which corresponds to the subjects classified as diseased and non-diseased among the exposed and un-exposed, complete the following table, which would represent what you would have actually observed in your study, with the degree of misclassification described.

OBSERVED STUDY RESULTS

Disease No Disease

Exposed

(N=1,000)

Unexposed

(N=1,000)

RR 95% CI =

Was this an example of differential or non-differential misclassification?

What effect did misclassification have in this example?

Is this result biased toward or away from the null hypothesis?

*Misclassification of exposure status*

Target Population (no misclassification)

Cases Controls

YES 600 300

Exposure

Status NO 400 700

OR = 3.5 (95% CI 2.9-4.2)

Assume that exposure status is classified with a sensitivity of 95% and a specificity of 75% in cases and with a sensitivity of 65% and a specificity of 90% in controls and complete the following 2 x 2 tables. Assume no misclassification of disease status.

CASES CONTROLS

True Status True Status

E Ē E Ē

Study E' Study E'

Status Status

Ē' Ē'

Using the data from these tables that correspond to the subjects classified as exposed and un-exposed among the cases and controls, complete the following table, which would represent what you would have actually observed in your study, with the degree of misclassification described.

OBSERVED STUDY RESULTS

Cases Controls

YES

Exposure

Status NO

1,000 1,000

OR, 95% CI =

Was this an example of differential or non-differential misclassification?

What effect did misclassification have in this example?

Is this result biased toward or away from the null hypothesis?

**~ 95% confidence interval of OR = exp (ln (OR) ± 1.96Ö (1/a + 1/b + 1/c + 1/d)**

**~ 95% confidence interval of RR = exp (ln (RR) ± 1.96Ö [b/(a(a+b)) + d/(c(c+d))]**