

**BE640 – Intermediate Biostatistics**  
**Frequently Asked Questions**  
**Topic 2 FAQ 2 – Regression and Correlation**

---

This FAQ is about **confounding** and **effect modification**: a reminder of the intuition we already have, some examples, some formal definitions, some techniques for their discovery and, some notes on their interpretation in statistical analyses.

**Intuition and Some Examples.**

**Confounding** - When a relationship between X and Y is **distorted** because of an underlying influence of some other variable Z on each of X and Y, we say the X-Y relationship is **confounded** by Z. Quite likely, you already have an intuition of confounding. Consider alcohol and lung cancer.

**Alcohol and Lung Cancer**

A quick look around reveals that high levels of alcohol consumption is associated with lung cancer. However, we know to be suspicious because we're familiar with the interrelationships of alcohol and smoking and lung cancer and smoking. We appreciate that the apparent association between alcohol consumption and lung is spuriously non-null because of the effects of smoking are lurking in the background unaccounted.

- X-Y in this example is ALCOHOL – LUNG CANCER
- Z in this example is SMOKING
- In actuality, there is no association between alcohol and lung cancer
  - Among nonsmokers: no alcohol-lung cancer association
  - Among smokers: no alcohol-lung cancer association
- Thus, the spuriousness here is an example of “*positive confounding*”, with “positive” referring to “*spuriously significant*”
- Confounding occurred as a result of **sampling bias**.

**Another Familiar Example of Confounding**

- The apparent relationship between low parity and breast cancer is confounded by age. In actuality, increasing age is associated with increased parity and increasing age is associated with increased risk of breast cancer, but at every age there is no relationship between parity and breast cancer.

**Effect Modification** - When a relationship between X and Y is **different** depending on the level of some other variable Z, we say the X-Y relationship is **modified** by Z. Quite likely here, too, you already have an intuition. Consider surgery and mortality. And advancing age.

### **Bypass Surgery and Mortality**

Among the treatments for coronary artery disease, and especially significant atherosclerosis, is bypass surgery. However, the success or non-success of bypass surgery depends on age. Mortality outcome associated with bypass surgery is different for older patients than for younger patients.

- X-Y in this example is BYPASS SURGERY – MORTALITY
- Z in this example is AGE
- The relationship between surgery and mortality is different with age.
  - Among the young: surgery is associated with reduced mortality
  - Among the old: surgery is associated with increased mortality
- Thus, a crude analysis of the BYPASS SURGERY – MORTALITY relationship will be **uninformative** because it does not take into account the effect of age.
- Effect modification occurs as a result of **variations in nature**.

### **Another Familiar Examples of Effect Modification**

- Synergistic effects of drugs when taking in combination

## Formal Definitions of Confounding

- “... confounding may be considered a confusion of effects. Specifically, the apparent effect of the exposure of interest is distorted because the effect of an extraneous factor is mistaken for or mixed with actual exposure effect (which may be null). The distortion introduced by a confounding factor can be large, and it can lead to overestimation or underestimation of an effect, depending on the direction of the associations that the confounding factor has with exposure and disease. Confounding can even change the apparent direction of an effect”.

Source: Rothman KJ and Greenland S. *Modern Epidemiology, Second Edition*. Lippincott 1998. page 120.

- “... we define a confounder to be a ‘risk factor’ for the disease under study whose ‘control’ in some appropriate way (either singly or in conjunction with other variables) will reduce or completely correct a bias when estimating the (true) exposure-disease relationship”.

Source: Kleinbaum, DG, Kupper LL, and Morgenstern H. *Epidemiologic Research: Principles and Quantitative Methods*. Lifetime Learning 1982. page 244.

- “A confounder is a variable that (a) is causally related to the disease under study (or, as often occurs in practice, serves as a proxy measure for unknown or unmeasured causes), and (b) is associated with the exposure under study in the study population, but is not a consequence of this exposure. It follows from (a) that within each level of exposure under study, the confounder is related to risk for disease, or in probabilistic terms, the confounder is related to the disease conditional on exposure”.

Source: Kelsey JL, Thompson WD, and Evans AS *Methods in Observational Epidemiology* Oxford University Press 1986. page 12

## Formal Definitions of Effect Modification

- “... refers to variation in the magnitude of a measure of exposure effect across levels of another variable. The variable across which the effects measure varies is called an effect modifier. Effect-measure modification is also known as heterogeneity of effect, nonuniformity of effect, and effects variation. Absence of effect-measure modification is also known as homogeneity of effect, uniformity of effect, and commonality of effect across strata.

Source: Rothman KJ and Greenland S. *Modern Epidemiology, Second Edition*. Lippincott 1998. page 254.

- “Effect-measure modification differs from confounding in several ways. The most central difference is that, whereas confounding is a bias that the investigator hopes to prevent or remove from the effect estimate, effect-measure modification is a property of the effect under study. Thus, effect-measure modification is a finding to be reported rather than a bias to be avoided. In epidemiologic analysis one tries to eliminate confounding but one tries to detect and estimate effect-measure modification”

Source: Rothman KJ and Greenland S. *Modern Epidemiology, Second Edition*. Lippincott 1998. page 254.

## Techniques for Discovery of Confounding and Effect Modification in Regression

Our goal is to develop an understanding of the interrelationships among possibly several predictors (let's call these  $X_1, X_2, \dots, X_p$ ) of a single outcome  $Y$ .

### Guiding Principles:

- **Confounding** is a function of **study design (sampling)** and produces **bias**. The resultant confounded estimate of an X-Y relationship is one that is biased. It is **NOT** of interest.
- **Effect modification** is a function of **nature (biology)** and produces **variation** in the X-Y relationship. Reporting these variations is of **VERY MUCH** interest.
- Repeating Rothman, **“In epidemiologic analysis one tries to eliminate confounding but one tries to detect and estimate effect-measure modification”** Source: Rothman KJ and Greenland S. *Modern Epidemiology, Second Edition*. Lippincott 1998. page 254.

### Outline of Steps


- **1 – List potential effect modifiers and confounders.** Do your literature review and compile this subject matter understanding.
- **2 – Obtain estimates of crude associations with outcome.** Perform model free analyses of two-way associations (two group t-tests, 2x2 tables, etc). Follow with fits of one predictor models as appropriate. Report nature, strength, and significance levels of the crude associations.

- **3 – Stratify data according to potential confounders and effect modifiers and perform stratified analyses. Or, sometimes, it is appropriate to assess effect-measure modification through the evaluation of interactions.**

Assess nature and strength of effect-measure modification. Define strata according to values of the potential effect modifier. Obtain, separately for each stratum, fits of one predictor models of outcome. Or, define new predictors that are the interaction of the predictor of interest X with the potential effect modifier Z. Simply, the new predictor is the product of X and Z. (Eg.  $XZ = X*Z$ )


- Effect-measure modification **IS** present if (1) the nature of an X-Y relationship is different depending on the stratum level of the stratification variable, OR (2) the interaction variable XZ is significant.
- Effect-measure modification **IS NOT** present if (1) the nature of an X-Y relationship is the same in each stratum of the stratification variable and (2) the interaction XZ is not significant.

- **4 – Depends on results of step “3”.**



<b>Effect-Measure Modification</b>	<b>NO effect-measure modification</b>
<p style="text-align: center;"><b>Report!</b></p> <p>Report separate, stratum specific, estimates of the X-Y relationships</p>	<p style="text-align: center;"><b>Check for Confounding</b></p> <p>Perform model free analyses of two-way associations (two group t-tests, 2x2 tables, etc) of predictors X with potential confounders Z. Follow with fits of one predictor models of X on Z as appropriate.</p> <p><b>Z is a confounder if it is related to X as well as Y.</b></p>

- 5 – Depends on results of step “4”.



<b>Confounding</b>	<b>NO confounding</b>
<p><u>Need to adjust –</u> Estimate an adjusted X-Y relationship by fitting a model for Y that includes both X and confounder Z as predictors</p> <p><b>Report!</b> Report estimated X-Y relationship from this adjusted model.</p>	<p><u>No need to adjust –</u> Estimate X-Y relationship by fitting a crude model that does NOT include Z as one of the predictors..</p> <p><b>Report!</b> Report estimated X-Y relationship from this crude model.</p>

## Notes on Interpretation in Statistical Analyses.

### Confounding

#### Example – Coffee, Bladder Cancer and Smoking:

If you suspect that smoking (Z) is a potential confounder of the relationship between the exposure coffee drinking (X) and bladder cancer(Y), your first step is to fit a single predictor model depicting the hypothesized association between outcome and each suspected confounder:  $Y = B_0 + B_2Z$ . Note that if there is more than one suspected confounder more than one one-predictor model must be fit.

If  $p\text{-value} < .25$  (or some generous threshold), retain Z for further consideration

However, if  $p \geq 0.25$ , then this variable is likely not associated with the outcome and thus can not be a confounder. This suspected confounder can now be dropped from consideration *unless it is of a priori importance*,

Add suspected confounders to your exposure-outcome model ( $Y = B_0 + B_1X$ ). Thus, your new model is  $Y = B_0 + B_1X + B_2Z$ . Fit this new model to your data and assess the significance of the suspected confounder in this two predictor model via its associated p-value.

If this suspected confounder (Z) retains significance at a more stringent level i.e.  $p < 0.10$  or  $p < 0.05$  (or some other desired level of stringency),

Then evaluate its effect on the value of  $B_1$  for the coefficient of the exposure of interest X. Conclude confounding is present if the value of  $B_1$  changes by 15% (or 20% depending on stringency) or more.

## Effect Modification

### Example – Surgery, Mortality, and Age:

If you suspect that the relationship between the exposure surgery (X) and outcome mortality (Y) is different for different levels of a third variable age, the effect modifier age (W). In other words, there is an additional joint effect of surgery (X) and age (W) on mortality (Y) beyond their individual effects.

#### How to Detect:

Effect modification can be detected by adding a product term (interaction term) into the model. The interaction term is represented as the product of exposure and the effect modifier, X\*W. For example,  $Y = B_0 + B_1X + B_2W + B_3X*W$  depicts a basic model with interaction term.

#### How to Determine:

Fit your desired model including the hypothesized interaction terms. Assess the significance of the effect modification by comparing the p-value interaction term's coefficient ( $B_3$ ) to the level of significance. Generally,  $p < 0.05$  is considered statistically significant evidence of effect modification.