Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

_____ _____

Student Signature Date

Thesis Approval Sheet

Student's Name: Haley M. Cionfolo

Student ID Number: 02737239

Thesis Title: "Effect Measure Modification or Confounding: A Case Study of Applied Epidemiological Methods for Beginners"

Statement of Approval:

I, _____, approve of the contents in this submission and certify their originality and adequacy to fulfill all necessary requirements for graduation from the Rollins School of Public Health.

_____ _____

Advisor Signature Date

Effect Measure Modification or Confounding: A Case Study of Applied Epidemiological Methods for Beginners

By

Haley M. Cionfolo
B.S., Duke University, 2023
B.A., Duke University, 2023

Advisor: Allison Chamberlain, M.S., Ph.D.

An abstract of
A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Global Epidemiology
2025

ABSTRACT

Effect Measure Modification or Confounding: A Case Study of Applied Epidemiological Methods for Beginners
By Haley M. Cionfolo

For epidemiology students and early-stage public health professionals, learning how to distinguish between effect measure modification (EMM) and confounding in real-world applications is often a challenge. After all, both methods require thinking beyond the standard exposure-outcome relationship to consider other factors—like biological mechanisms—at play. In this paper, we use a peer-to-peer perspective to guide these beginners through the process of evaluating a research question and accompanying analyses with these methodologies using an example of examining the influence of nativity to brucellosis-endemic countries on the relationship between key risk behaviors, like consuming unpasteurized dairy products, and severe disease. Although we find that neither of these methodologies are necessary in this example—and that there is no notable influence of nativity in this population—we provide step-by-step directions to determine when EMM and confounding are applicable and how to use them with real data.

Effect Measure Modification or Confounding: A Case Study of Applied Epidemiological Methods for Beginners

By

Haley M. Cionfolo
B.S., Duke University, 2023
B.A., Duke University, 2023

Advisor: Allison Chamberlain, M.S., Ph.D.

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Global Epidemiology
2025

I. Introduction

If you're an epidemiology master's student reading this, chances are you could use a little help to more deeply understand differences between effect measure modification and confounding. There's no shame in that—graduate school itself is a grind, and the basic tenants of epidemiology make us question every little detail about the potentially causal relationships we are interested in. Part of this struggle may include wrestling with the differences between effect measure modification and confounding. If it is, don't worry. You're not alone!

Needing help isn't a weakness. You care about learning these confusing concepts and learning them correctly. Recognizing the challenge and asking for guidance is half the battle. As a fellow epidemiology student, our field is difficult. Dizzying, even. As some say, epidemiologists are disease detectives, searching for patterns, while constantly remaining vigilant of sneaky variables that might distort the truth. It is a constant battle between what we can observe and what lurks beneath the surface.

Effect measure modification and confounding throw many epidemiology students for a loop. A quick visit to Reddit shows just how common this confusion is (1). Even some of your professors struggled with these concepts back in the day, but they might not admit it. The reason is that these two concepts are very similar: they both attempt to address the impact of external factors on a cause-and-effect relationship that you are interested in measuring (2,3).

Sometimes, the best way to learn is from a friend or peer who has already gone through untangling the "knots" of understanding. At the very least, it might save you a few headaches from going down Google rabbit holes. Throughout this piece, I will walk you through effect measure modification and confounding, peer-to-peer, covering what they are, when we use them, and how to interpret them.

a. What is Effect Measure Modification?

Effect measure modification, which we will refer to as its shorter and less scary abbreviation EMM, is one of epidemiology's favorite ways to say, "It depends". We use EMM when we suspect there may be a third variable shaking up the relationship between exposure and outcome. Our hunch may stem from our literature review, our observations, or something else about our data and topic.

For example, say we suspect that the relationship between an exposure and a disease may differ between males and females. EMM helps us understand each group separately, as we have doubt that our relationship of interest is the same across both sexes. Here, the relationship *depends* on sex, as we suspect different magnitudes of association for males and females.

When you were first introduced to EMM in a classroom, you likely also learned about stratification. And what's *stratification* again? It's when you break a population up into groups, or *strata,* based upon a specific characteristic or variable to examine a specific exposure-outcome relationship within each subgroup. When we stratify for EMM, we split our population into groups based upon the third variable—in this case, sex—and analyze males and females separately. This results in two measurements of association—one for males and one for females—which we can then compare to each other. All of this to say, any time we include EMM in our analysis, we expect to generate more than one measure of association, one for each level or category of the EMM variable we stratified our population by. The results that are generated allow us to clearly see how our relationship of interest varies by levels or categories of the EMM variable (4,5).

Now, let's take it one step further. We account for EMM in any regression model—from unconditional logistic regression to survival analysis to Poisson regression—with "product

terms". Product terms multiply the exposure and the third variable to quantify the combined effect of the presence of both variables**.** Not only are our exposure and third variable both present in the model separately, but we also include this third term in which we multiple their values together. It's this product term that allows us to create and examine separate measures of association for each level or category of the EMM term. This helps us answer key questions, like in the aforementioned example: What is the combined effect between the exposure and being male? Being female? Our model can now give us these answers.

When thinking about EMM, keep phrases like "What is the difference between…?", "How do X and Y differ…?", and "Across X and Y groups…" in mind. If you want to compare different subpopulations for your research question, it is a good sign you need to use EMM.

b. What is Confounding?

Let's shift gears to talk about confounding, another way epidemiologists say, "It depends." Confounding, however, does this in a very different way. We use confounding when we believe a stealthy third variable may distort the relationship between the exposure and the outcome, either making the association appear stronger or weaker than it actually is.

Let's use an example to put it in perspective. Have you ever heard that air conditioner usage is linked to ice cream sales? There may be truth to this correlation, sure, but it isn't causal. Cranking up the air conditioning isn't *causing* people to buy more ice cream. There's a different culprit: outdoor temperature. As it gets hotter outside, more people turn on their air conditioning to cool down. To beat the heat, people also crave cold treats, like ice cream, more. As such, the relationship between exposure (like air conditioner use) and outcome (like ice cream sales) *depends* on if we consider the influence of other, more causal variables (like temperature).

What makes a confounder? This variable must be associated with the exposure and the outcome independently. If you have reason to suspect a confounder, from the literature, observations, or context, you can't just ignore it. It is warping the reality represented by your analysis and must be controlled for promptly.

This is where students often get EMM and confounding mixed up, because both methods involve stratification to clarify the true relationship between exposure and outcome. Confounding uses stratification differently than EMM. When we stratify for confounding, we ask, "Does the relationship change when we take this third, possibly confounding, variable into account?" If the association stays about the same when you control for that third variable, great—no confounding. If it doesn't stay the same, we must account for it. While this still divides our data into two groups, like in EMM, we are now looking at the effect of one variable's inclusion in our analysis, compared to its absence (4,5).

So how do we account for confounders in any kind of regression? They stand alone in our equations, much like the exposure. We can think of it like a sponge, absorbing their share of the influence on the exposure-outcome relationship so the true magnitude of association we're actually interested in can be detected. Without them "standing guard" in our models, our results might be misleadingly large or small.

When you are dealing with confounding, listen for phrases like "controlling for", "adjusting for", or even "regardless of". These are clues that you might need to account for these variables as confounders in your analysis. Simply put, confounding is all about making sure we are measuring the actual relationship we think we are measuring, not one that is muddied by other factors that also might have associations with our exposure and our outcome.

II. The Problem

a. The Goal of This Piece

How do you choose whether to use EMM or confounding in the real world? No one will tell you how to handle your variable of interest in relation to the larger exposure-outcome relationship when you leave the classroom or are otherwise set free to pursue your own projects. How do you decide what to do?

Even seasoned epidemiologists run into this debate too, each with strong arguments for either approach for answering their research question. For example, a researcher from the Netherlands experienced a similar problem, documenting her struggle to determine whether sex should be an EMM or confounder in her research on cardiac surgery outcomes. Long hailed exclusively a confounder in this field, growing research regarding physiological differences between men and women have suggested it could very well be an understudied EMM (6).

Again, *it depends.*

In this piece, I will spare you the trouble of trial and error by walking you through my own struggles with this decision. After all, sometimes there is no better teacher than each other. Together, we will explore the relationship between different forms of exposure to animal and animal products and severe brucellosis as it relates to whether a patient was born in an endemic country, as well as how to handle this "nativity" variable.

a. The Research Question

Specifically, our research question is "How does nativity affect the risk of severe brucellosis?" Here, we have 2 exposures for which we are interested in quantifying relationships, listed and defined in the table below.

Exposure	Definition

Animal Fluids	Contact with any bodily fluid from an animal through slaughter, birthing, hunting, or general contact
Unpasteurized Dairy Consumption	Ingesting any dairy product derived from animals, including raw milk and unpasteurized cheese

b. What is Brucellosis?

In the early 2000s, the World Health Organization declared brucellosis the "top dog" of zoonotic diseases, earning the dual distinction of being the world's most widespread zoonosis and the most common laboratory-acquired infection (7). Caused by various species of sneaky *Brucella* bacteria, this disease knows no boundaries—infecting humans and animals alike, including cattle, goats, sheep, pigs, and wild hogs (8). Disease presentation is similar across both humans and animals, including early, nondifferential symptoms, like fever, fatigue, nausea, vomiting, weakness, high blood pressure, and decreased appetite, and hallmarks of later-stage infections, like necrosis, inflammation and/or enlargement of lymph nodes, liver, lungs, and spleen, lesions, and immune dysfunction.

How does brucellosis jump from animals to humans? Typically, it hides in raw meat or unpasteurized dairy that might be shed by animals or consumed when humans touch them or their bodily fluids (9). For example, in the United States, feral swine are prime suspects. Hunters, in particular, risk contracting brucellosis while handling infected carcasses, a harsh reminder that even hobbies can have public health implications (10).

Brucellosis is considered endemic in most countries, with somewhere between 500,000 to 2.1 million new human cases each year worldwide (11). Many countries have implemented policies to prevent transmission of *Brucella*, such as restrictions on the commercialization of unpasteurized dairy, but brucellosis continues to trouble many populations (12).

In lower- and middle-income countries (LMICs) brucellosis isn't just a health concern—it's a daily threat. Without robust infection control measures for livestock, these regions bear the brunt of the disease. Resource limitations mean prevention is often out of the question for some countries in the Middle East, Africa and Latin America, creating the perfect circumstances for *Brucella* to thrive in humans and livestock (11).

While efforts like the 1934 Brucellosis Eradication Program have significantly reduced the incidence of brucellosis within the United States, the flow of people and animals across country borders presents global challenges for brucellosis control (10). The combined effects of zoonosis and human movement have often led to brucellosis outbreaks in LMICs (7,13). On the other hand, brucellosis is a rare occurrence in high-income countries (HICs), including those in Europe and North America, but the combined effects of zoonosis and human movement do not make these settings exempt from brucellosis case detection and strain (14).

Brucellosis surveillance in the United States has revealed some intriguing—and concerning--patterns. The cases detected in the last 30 years have often come with a story. Between 2000 and 2009, for example, 18 patients were diagnosed with brucellosis in Houston, Texas (15). Many of these patients also reported immigration and/or travel from Central America (n=13), as well as consumption of raw milk products. Such patterns and stories hint at a larger puzzle. How do nativity and associated cultural landscapes intersect with exposure to shape patient risk?

c. The Problem of Nativity

When we talk about where someone was born, we use the term *nativity*. As the W.H.O.'s Social Determinants of Health Framework reminds us, factors outside of medicine and chance, such as where someone lives, can shape their susceptibility to disease and overall health

outcomes (16). This may be due to heredity and the biological traits passed down from generation to generation, or due to exposures from their cultural environment.

It's not uncommon for nativity to play a role in infectious disease risk. For example, a 2017 study from Michigan suggests that residents born in tuberculosis-endemic countries were diagnosed with active disease in the United States at *7 times* the rate of those of domestic birth, though the role of biological mechanisms here is not clear (17). While some patients may have been exposed after immigrating due to social vulnerability, the prevailing theory is that most infections were contracted in their country of birth, remained latent, and reactivated after their move to the United States. This is particularly common for those from low-resource settings, who often have weaker immune systems, hence why the BCG vaccine is an important intervention in these locations.

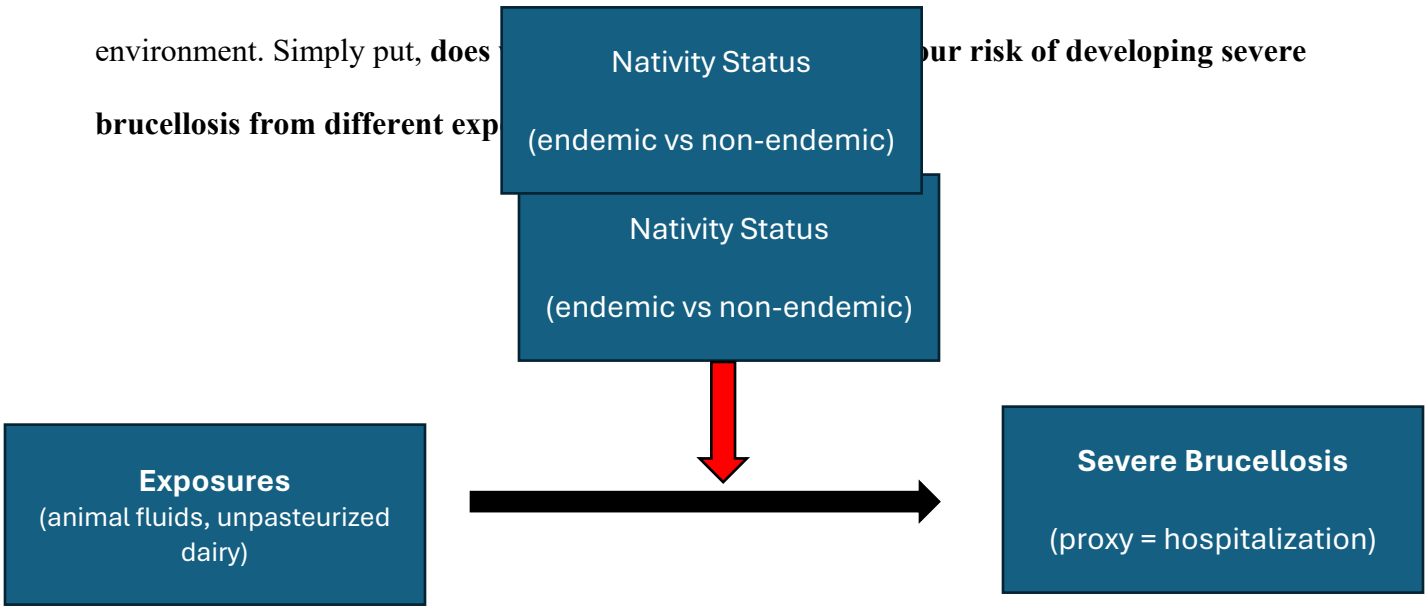
There is reason to suspect that U.S. brucellosis cases experience a similar phenomenon . After all, as previously mentioned, brucellosis is not endemic to much of the United States. Moreover, those born outside of the United States may travel to their country of birth and engage with cultural norms, and, in an endemic country, activate any immunological mechanism that either exacerbates or reduces the risk of infection. On top of this, research from as far back as 1995 suggests that brucellosis infections can relapse years later, especially when treatment isn't adhered to or properly prescribed or when the immune system is facing other adversities. Some studies propose that brucellosis can stay dormant for anywhere from 2 to even *40 years* (18–20). Could that mean that someone infected in a brucellosis-endemic country could experience disease reactivation long after moving to a non-endemic country, like the United States?

Does nativity influence brucellosis risk in the United States similarly to domestic tuberculosis cases, either due to hereditary or biological mechanisms or due to environmental

exposure? The answer isn't clear. While some evidence suggests that nativity is not directly associated with diagnosis with a Nationally Notifiable Disease, there has been little focus on the influence of nativity and specific notifiable infections, like brucellosis (21).

1. **Could Nativity be an EMM?**

Perhaps, we think nativity could be an EMM, and we want to evaluate the differential impact of being born in a brucellosis-endemic country compared to being born in a non-endemic country. After all, different cultures may have unique dishes, agricultural practices, traditions, attitudes, and implementation of existing brucellosis-control recommendations that may impact someone's susceptibility to brucellosis. For example, queso fresco made from unpasteurized cheese is a common delicacy in Mexico, where brucellosis is endemic. Moreover, by definition, having endemic disease suggests higher-than-average circulation of the pathogen in the environment. Simply put, **does** **ur risk of developing severe brucellosis from different exp**

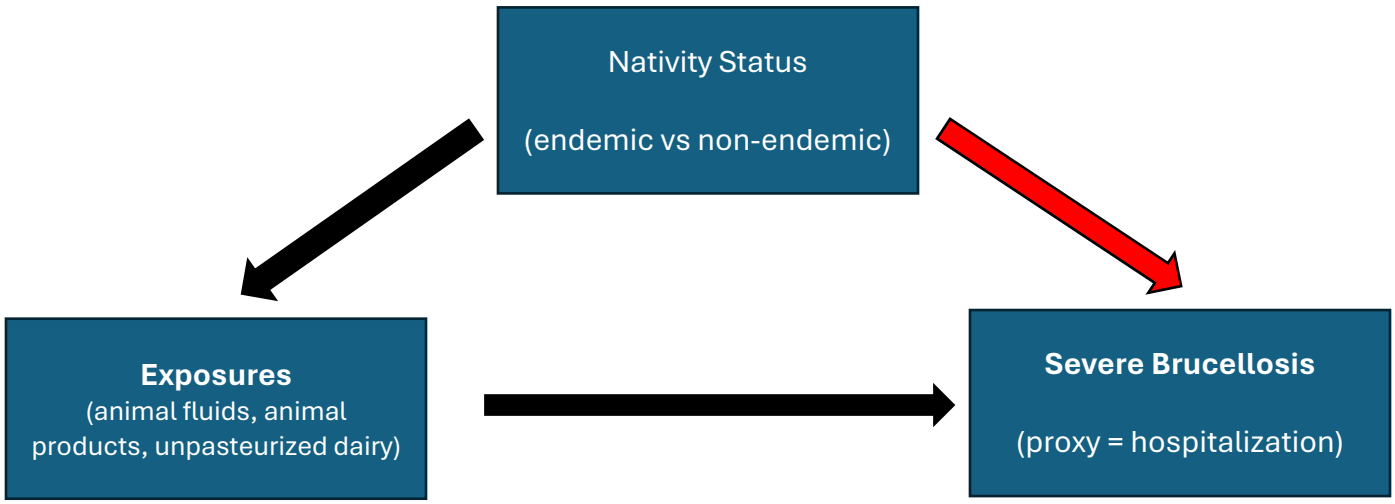


2. **Could Nativity be a Confounder?**

Maybe we think that nativity could be a confounder, instead. Being from an endemic country may be associated with more frequent exposures to animal and animal products, for

similar reasons considered above. Moreover, it's feasible that a specific nativity could translate to biological susceptibility, like we see in other diseases. For example, sickle cell anemia acts as biological adaptation protecting these individuals from severe malarial disease (22). Conversely, immunological mechanisms can lead to worse health outcomes with each additional exposure, Namely, studies have suggested that, with each consecutive dengue infection, a patient is more likely to experience severe disease (23). Therefore, we may want to account for nativity in our calculations and level the field across our population, instead of splitting it in two. Simply put, **what is your risk of developing severe brucellosis from different exposure pathways, regardless of where you were born?**

Once again, *it depends*. Each analysis hinges on similar logic, but each have different questions and generate different answers.



d. The Dataset

To dig into the possible relationships between different transmission sources, severe disease, and the impact of being born in an endemic country, we'll dive into surveillance data from the CDC (24). This dataset pulls together information collected through brucellosis case

report forms (CRFs) submitted between 2010 and 2024 for cases identified in the United States. It's worth noting that this data does not represent the complete picture of brucellosis burden during this timeframe, as CRF submission to the CDC is voluntary. On the other hand, the Nationally Notifiable Disease Surveillance System (NNDSS) provides a more comprehensive view of all reported cases, but it does not include details on animal exposures or nativity—and often, these cases cannot be appropriately matched to CRF data.

To explore the association between each exposure and severe brucellosis (as measured by whether a case was hospitalized or not), especially when considering nativity, we need to choose the right study design and analysis approach. Since we are not asking about changes over time here, but rather the point prevalence of a non-rare outcome, a cross-sectional design is key. Using this design, we'll use log-binomial regression to answer our questions.

III. Application

Using our favorite statistical software, we can easily evaluate the log-binomial regression relationship between our exposures—animal fluids and unpasteurized dairy—and our outcome of severe brucellosis infection by proxy of hospitalization.

Before we factor in nativity, or any other variable for that matter, we must evaluate the crude relationship between just the exposures and outcome. These unadjusted models should be entered like this, with no other variables involved:

HOSPITALIZED ~ ANIMAL FLUID

HOSPITALIZED ~ DAIRY

Where:

Patient had severe brucellosis if HOSPITALIZED = 1, else 0.

Patient was exposed if ANIMAL FLUID or DAIRY = 1, else 0.

We should have one measure of association, a prevalence ratio for each. Our interpretations, since just the exposures and outcome are involved, have no bells or whistles. Write them like this:

1. The prevalence of severe infection in those who were exposed to animal fluids is 0.84 times the prevalence of severe infection in those not who were not (95%CI: 0.74, 0.98).
2. The prevalence of severe infection in those who consumed unpasteurized dairy is 1.13 times the prevalence of severe infection in those not who did not (95%CI: 1.03, 1.25).

Okay! Now that we've done that, we can start playing around with nativity! Starting with nativity as an EMM, we evaluate these relationships with nativity as both a stand-alone variable and a product term. This product term is what will generate the two prevalence ratios we want from evaluating nativity as an EMM. As such, we enter our models as:

HOSPITALIZED ~ ANIMAL FLUID + NATIVITY + ANNIMAL FLUID*NATIVITY

HOSPITALIZED ~ DAIRY + NATIVITY + DAIRY*NATIVITY

Where:

Patient had severe brucellosis if HOSPITALIZED = 1, else 0.

Patient was exposed if ANIMAL FLUID or DAIRY = 1, else 0.

Patient was born in a brucellosis-endemic country if NATIVITY = 1, else 0.

As previously mentioned, when we run these regressions, the output we report will include two prevalence ratios per exposure, as shown below: one for the exposure among those born in an endemic country and one for the exposure among those not born in an endemic country. Therefore, when we interpret our measures of association, we must specify which part

of the population we are referring to. For example, for unpasteurized dairy exposure, we may report our prevalence ratios like this.

3. Among those born in a country where brucellosis is endemic, the prevalence of severe infection in those who consumed unpasteurized dairy is 1.08 times the prevalence of severe infection in those not who did not (95%CI: 0.79, 1.51).
4. Among those not born in a country where brucellosis is endemic, the prevalence of severe infection in those who consumed unpasteurized dairy is 1.13 times the prevalence of severe infection in those not exposed (95%CI: 1.00, 1.27).

It should be noted here that, when stratified by nativity as an effect modifier, unpasteurized dairy consumption seems harmful for those who were not born in an endemic country but has no relationship for those born in an endemic country, as the corresponding 95% confidence interval straddles the null value of 1.00. You make think this is odd. Hold onto that thought! We will revisit it shortly.

Nativity as an Effect Measure Modifier				
Model	**Point Estimate**	**2.5%**	**97.5%**	**Direction**
Animal Fluid	-	-	-	-
➔ Nativity = 1	0.92	0.54	1.56	Inconclusive
➔ Nativity = 0	0.86	0.73	1.00	Protective
Unpasteurized Dairy	-	-	-	-
➔ Nativity = 1	1.08	0.79	1.51	Inconclusive
➔ Nativity = 0	1.13	1.00	1.27	Harmful

Now, we can repeat this process with our handy-dandy statistical software to evaluate nativity as a confounder. As previously described, confounders appear in regression equations alone, unlike EMM. Therefore, our models and variable designations here would be the same as before, just without the product terms.

HOSPITALIZED ~ ANIMAL FLUID + NATIVITY

HOSPITALIZED ~ DAIRY + NATIVITY

Now, because we no longer have our product term, we will only have one measure of association in our output, just for our relationship between exposure and outcome when we account for differences in patient nativity. We must mention this specification when we report our single prevalence ratio, like in this example for unpasteurized dairy exposure.

5. The prevalence of severe brucellosis infection in those who consumed unpasteurized dairy was 1.11 times the prevalence of severe brucellosis in those who did not, regardless of where patients were born (95%CI: 1.01, 1.23).

Here, like we noted earlier, instead of the phrase "regardless of", we could have also said "controlling/adjusting/accounting for patient nativity", whichever floats your boat. But that phrase is what tells your reader that you analyzed the relationship with nativity, or your variable of interest, as a confounder. We can also note here, that when we considered nativity, we found a modest, but definitively harmful relationship between unpasteurized dairy consumption and severe brucellosis.

Nativity as a Confounder				
Model	**Point Estimate**	**2.5%**	**97.5%**	**Direction**
Animal Fluid	0.88	0.75	1.00	Protective

Unpasteurized Dairy	1.11	1.01	1.23	Harmful

IV. Putting It All Together

As we can see in this example, both EMM and confounding methods output similar values of similar magnitudes, but they each have very different meanings.

Let's look at those interpretations, again, this time side by side.

1. **EMM (Nativity = 1):** Among those born in a country where brucellosis is endemic, the prevalence of severe infection in those who consumed unpasteurized dairy is 1.08 times the prevalence of severe infection in those not who did not (95%CI: 0.79, 1.51).
2. **EMM (Nativity = 0)** Among those not born in a country where brucellosis is endemic, the prevalence of severe infection in those who consumed unpasteurized dairy is 1.13 times the prevalence of severe infection in those not exposed (95%CI: 1.00, 1.27).
3. **Confounding:** The prevalence of severe brucellosis infection in those who consumed unpasteurized dairy was 1.11 times the prevalence of severe brucellosis in those who did not, regardless of where patients were born (95%CI: 1.01, 1.23).

Each method and associated interpretation, no matter how similar the values, are specific to a specific set of circumstances—either those who were born a brucellosis-endemic country, those who were not, or both. The values from one group cannot be generalized to the other, as the populations represented by that association are not the same. We can see this when we set the numbers aside:

4. **EMM (Nativity = 1):** There is not a clear association between consuming unpasteurized dairy and severe brucellosis in patients born in a brucellosis-endemic country (1.08, 95%CI: 0.79, 1.51).
5. **EMM (Nativity = 0):** There is a harmful association between consuming unpasteurized dairy and severe brucellosis in patients who were not born in a brucellosis-endemic country (1.13, 95%CI: 1.00, 1.27).
6. **Confounding:** There is a harmful association between consuming unpasteurized dairy and severe brucellosis in the patient population, regardless of where individuals were born (1.11, 95%CI: 1.01, 1.23).

By changing *who* we are looking at, each of these statements answers very different research questions. Mismatching research questions and methodological approaches cannot lead you to your answer. In a way, epidemiology is more than a quantitative science—it's an art of asking the right question and charting the right path to get the right answer. No matter the data set or topic, it always comes down to the research question you are pursuing.

A. What Does This Mean for Beginning Researchers?

Good news! You have options!

Bad news! You have options!

If you're interested in the influence of a third variable on an exposure-outcome relationship, there are different ways to incorporate it. There may not necessarily be a wrong way to account for it in any sort of regression, but it is only helpful if it fits the context of your research question. You must ask yourself questions like, "Why am I interested in the influence of this third variable on this relationship?", "How do I expect the third variable to fit into the relationship I am studying?", and "Why might understanding the effect of this third variable be meaningful?"

As a beginning researcher, you may not have a sense of these answers by yourself. That's okay! You're learning! If you are still familiarizing yourself with your topic, it might not be immediately clear to you what factors may influence the relationship between your exposure and outcome of interest. To start to build your expertise, and to create some direction for your analysis, a great place to start is the literature. What studies exist on your topic? What do they credit to be influential in the relationship?

It may also help to work backwards.

If I am interested in answering a question related to Social Determinants of Health, I may want to think about using EMM, as I would then have two measures of association to compare and then determine if there is a higher likelihood of disease in one group relative to another. In this example, we would want to use EMM if we were interested in how the prevalence of severe brucellosis due to certain exposures differs between those born in an endemic country compared to those who weren't. We are inherently asking a question of social determinants about how place of birth influences health outcomes.

On the other hand, if I am interested in the overall risk of a population, I may want to think about controlling for confounders to have one summary statistic representative of the whole cohort. Here, we are interested in adjusting for differences rather than looking how the conditions change the association. In our example, we would control for nativity as a confounder if we were interested in the risk for all patients with a CRF "regardless of where they were born."

As a researcher, you must choose what questions you want to answer, therefore selecting a method. You may face cases, like this, where it may make sense to treat your third variable of interest as either an EMM or confounder, but, once again, *it depends. It depends* on the question you're asking.

B. What This Doesn't Mean for Beginning Researchers

Now that you (hopefully!) have a solid grasp confounding and EMM, you may start more spotting them everywhere. That's great! But be careful—you don't want to turn your analysis into a wild goose chase. After all, confounding and EMM aren't imaginary friends—they either exist in your research question and data collection or they don't.

It's easy to get carried away. Excited young epidemiologists—eager to flex their new analytical muscles—may start wondering if *every* variable they encounter is a confounder or an EMM. Sometimes, they're absolutely right, and those extra covariates clearly belong in one of those categories. In other cases, like the example here, it falls into a gray area, open to debate. Of course, *it depends* on the research question and related literature.

And then, there are the variables that, despite our best instincts, simply don't belong in the model at all. Sometimes, we must kill our darlings.

As painful it is to let them go some covariates need to be dropped, especially if they don't actually influence the exposure-outcome relationship or introduce problems like collinearity. Sure, a noninfluential variable won't *harm* your model or association, but it won't *help* either. Instead, it clutters your analysis and makes it harder to see the real story your data is trying to tell.

So, stay vigilant—but don't overcomplicate things! Not every variable is a hidden EMM or confounder waiting to be discovered. Sometimes, less really *is* more.

C. When It Isn't Confounding

Figuring out when a third variable *isn't* confounding is surprisingly straightforward, especially when we have a crude measure of association to guide us. In fact, with the right

preparation, we might be able to sidestep the whole process for controlling for a variable altogether.

Before diving into an analysis, we need to visualize and justify the relationships we care about. Enter directed acyclic graphs, or DAGs, our trusty tool for mapping out connections between exposures, outcomes, and any other variables we're considering. If we can't establish a "causal triangle", or causal relationships between our third variable and both our exposure and outcome, then we probably don't need to control for it.

But let's say we have a strong causal triangle and decide to include that third covariate. Even then, it's possible that that variable isn't pulling its weight in the analysis and should be dropped. That's where the 10% Rule comes in, a data-based confounding assessment to check if a variable is actually influential in our association. If the difference between the crude measure of association and adjusted measure is greater than 10%, then we should control for that variable. Otherwise, it's just taking up space.

Let's put this to the test! Looking at our example, with unpasteurized dairy as the exposure, we suspected confounding, thanks to our DAG. Now that we have both the crude prevalence ratio and the adjusted prevalence ratio, we can evaluate if confounding is *actually* present:

$$C = ([\text{crude} - \text{adjusted}]/\text{adjusted}) \times 100\%$$

$$C = ([1.13 - 1.11]/1.11) \times 100\%$$

$$C = 1.8\%$$

Since 1.8% is well below our 10% threshold, nativity doesn't appear to influence the relationship between unpasteurized dairy consumption and severe brucellosis inn reported U.S. cases from 2010 to 2024. Simply put, here, being born in an endemic or non-endemic country does not confound this relationship. And, if you were wondering, no, nativity doesn't confound

the relationship between animal fluid exposure and severe brucellosis either, but now you can see for yourself! A little visualization, some strategic thinking, and a quick calculation can save you from unnecessary covariates and a whole lot of overthinking.

D. When It Isn't EMM

Determining whether EMM is present is a bit trickier than our confounding assessments. In this case, we lean on statistical tests, like the Wald Test and the gold-standard Likelihood Ratio Test, to see if there is a statistically significant difference between subgroups of interest.

But here's an important reality check: statistically meaningful and clinically meaningful aren't always the same thing! Just because a test spits out a significant p-value, or doesn't, it doesn't mean that the difference is impactful in public health practice.

We can apply the Likelihood Ratio Test to our unpasteurized dairy example. We want to see if there's a statistically significant difference in the association between our exposure, unpasteurized dairy, and severe brucellosis among those born in endemic and non-endemic countries. Running the Likelihood Ratio Test and comparing the full and reduced models, we have a p-value of 0.87 (code in Appendix II).

Translation: there is not enough evidence to suggest a statistically significant difference between those born in an endemic country compared to those not born in an endemic country. This likely explains why one of our prevalence ratios from our model above straddled the null. And if we perform the same Likelihood Ratio test for our model with the animal fluid exposure, we land on the same conclusion.

E. When It Isn't Confounding OR EMM

Let's be clear about what we investigated here—the possibility of an **inherited** biological difference in brucellosis risk in those born in endemic areas and those born. We hypothesized

that there could be an immunological mechanism that either protects those born in endemic countries from severe disease (again, like sickle cell and malaria) or exacerbates a brucellosis infection they may acquire (again, like West Nile Virus). This is distinctly different from an environmental hypothesis—that routinely engaging in the cultural norms of the environment imbues such a difference in individuals through repeated exposure, not through heredity. From our analysis, it seems like, at least in this sample, there likely is not a heredity-based biological mechanism that influences risk differently in those born in endemic and non-endemic countries.

It's completely natural for a budding epidemiologist to be disappointed when they've poured time into gathering evidence, synthesizing research, and building a case, only to find that the relationships they were investigating don't actually exist in their dataset. But don't be discouraged! Just because your third variable of interest may not appear to be a confounder or EMM, doesn't mean it isn't important.

Here, in our example, it turns out the nativity isn't a confounder or an EMM! That's okay! In our practice of these core epidemiology methods, we found similar associations between the crude relationships of the exposures, like unpasteurized dairy consumption, and severe brucellosis and the relationships that consider nativity. We still learned how to think about confounding and EMM like a pro and how to applies these methods to our research. That's certainly worth *something*!

Still, maybe nativity does matter, but not in the way we expected. We may suggest that nativity – was simply proxy variable for culture-associated risk – meaning proximity to cultures that have behaviors, such as routine consumption of unpasteurized dairy, place individuals born in that culture or county at differential, dose-dependent risk. After all, we learned that unpasteurized dairy increases risk and that there are brucellosis-endemic countries. What we do

know, however, is that nativity does not directly influence the relationship between different exposures and severe brucellosis in this sample of reported brucellosis cases.

Remember, your dataset just a snapshot, a small representation of the larger population. And, like any snapshot, it comes with limitations and biases. In any disease surveillance program, we only see the tip of the iceberg—we don't see anyone who was misdiagnosed, undiagnosed, or did not seek care. And, in this case, where we are looking at nativity— a sensitive topic in the United States—we may have further underreporting or misrepresentation due to concerns about immigration status. On top of that, we also only see what states *chose* to report to the CDC. As such, we don't know what factors influenced whether a case even made it into the CDC's surveillance system in the first place.

Even with several hundred patients in our sample, there are still limitations. Effect size is constrained by sample size, as well as possibly dose-dependence of the exposure, underlying health conditions, or other factors weren't captured in the CRF. Whatever the case, it just means that nativity isn't influential here in this sample.

Does that mean the same holds true in the larger population? Maybe! Maybe not! If we meet the thresholds for internal and external validity, we might feel comfortable making that assumption. The only way to truly know, however, is to survey every brucellosis patient in the United States from 2010 to 2024, which is virtually impossible!

So, don't fret! There is no such thing as a "non-finding"! After all, knowing what *isn't* influencing an outcome is just as valuable as knowing what *is.*

V. Conclusions

Now that we've walked through the entire analysis process and each complexity of confounding and EMM, you now have the skills to follow your own research interests. Together,

we went stage by stage, considering what confounding and effect measure modification (EMM) are and the evidence needed to consider each one in real research question. We learned how to both implement these methods, as well as how to evaluate when they are actually necessary. Finally, we reflected on the very real trial and error, the surprises and disappointments, of application.

I hope that, as a fellow student, my own struggles with this concept with a messy real-world surveillance dataset, and how I worked through it, can guide you forward with confidence. Maybe, after finishing this, you immediately understand. Maybe you don't and you need extra practice. Either is fine! Keep asking for help and be patient with yourself as you put your understanding to the test. Practice makes perfect!

In the end, whether a variable is a confounder, an effect modifier, or simply an unnecessary covariate isn't always straightforward. Like so much in epidemiology, *it depends*.

VI. References

1. Confounding vs effect modification - Reddit Search! [Internet]. [cited 2025 Feb 19]. Available from: https://www.reddit.com/search/?q=confounding+vs+effect+modification&cId=09e5a7c0-a333-4572-b86e-96bec43fe10e&iId=3c3ff379-c5ee-4743-a4e1-ddc1e17f3b1d

2. Bovbjerg ML. Effect Modification. 2020 Oct 1 [cited 2025 Feb 19]; Available from: https://open.oregonstate.education/epidemiology/chapter/effect-modification/

3. Vander Weele TJ. Confounding and effect modification: distribution and measure. Epidemiol Method. 2012 Aug 1;1(1):55–82.

4. Vetter TR, Mascha EJ. Bias, Confounding, and Interaction: Lions and Tigers, and Bears, Oh My! Anesthesia & Analgesia. 2017 Sep;125(3):1042.

5. Christiansen-Lindquist L, Wall KM, Lash TL. Fundamentals of Epidemiology [Internet]. 1st ed. New York: Springer Publishing Company; Available from: https://connect.springerpub.com/content/book/978-0-8261-6694-4

6. Kavousi M. Confounding and effect measure modification: analysing sex in cardiovascular research [Internet]. [cited 2025 Feb 19]. Available from:

https://eurointervention.pcronline.com/article/confounding-and-effect-measure-modification-analysing-sex-in-cardiovascular-research

7. Pinn-Woodcock T, Frye E, Guarino C, Franklin-Guild R, Newman AP, Bennett J, et al. A one-health review on brucellosis in the United States. J Am Vet Med Assoc. 2023 Apr 1;261(4):451–62.

8. Brucellosis | CDC Yellow Book 2024 [Internet]. [cited 2024 Sep 26]. Available from: https://wwwnc.cdc.gov/travel/yellowbook/2024/infections-diseases/brucellosis

9. Negrón ME. Notes from the Field: Human Brucella abortus RB51 Infections Caused by Consumption of Unpasteurized Domestic Dairy Products — United States, 2017–2019. MMWR Morb Mortal Wkly Rep [Internet]. 2019 [cited 2024 Sep 6];68. Available from: https://www.cdc.gov/mmwr/volumes/68/wr/mm6807a6.htm

10. Zoonotic pathogens from feral swine that pose a significant threat to public health - PubMed [Internet]. [cited 2024 Sep 6]. Available from: https://pubmed.ncbi.nlm.nih.gov/29388363/

11. Laine CG, Johnson VE, Scott HM, Arenas-Gamboa AM. Global Estimate of Human Brucellosis Incidence - Volume 29, Number 9—September 2023 - Emerging Infectious Diseases journal - CDC. [cited 2024 Sep 20]; Available from: https://wwwnc.cdc.gov/eid/article/29/9/23-0052_article

12. Sfeir MM. Raw milk intake: beware of emerging brucellosis. J Med Microbiol. 2018 May;67(5):681–2.

13. Norman FF, Monge-Maillo B, Chamorro-Tojeiro S, Pérez-Molina JA, López-Vélez R. Imported brucellosis: A case series and literature review. Travel Med Infect Dis. 2016;14(3):182–99.

14. Rubach MP, Halliday JEB, Cleaveland S, Crump JA. Brucellosis in low-income and middle-income countries. Curr Opin Infect Dis. 2013 Oct;26(5):404–12.

15. Serpa JA, Knights S, Farmakiotis D, Campbell J. Brucellosis in Adults and Children: A 10-Year Case Series at Two Large Academic Hospitals in Houston, Texas. South Med J. 2018 Jun;111(6):324–7.

16. Social determinants of health [Internet]. [cited 2025 Feb 19]. Available from: https://www.who.int/health-topics/social-determinants-of-health

17. Race and nativity are major determinants of tuberculosis in the U.S.: evidence of health disparities in tuberculosis incidence in Michigan, 2004–2012 | BMC Public Health [Internet]. [cited 2025 Feb 19]. Available from: https://link.springer.com/article/10.1186/s12889-017-4461-y

18. Ögredici Ö, Erb S, Langer I, Pilo P, Kerner A, Haack HG, et al. Brucellosis Reactivation after 28 Years. Emerg Infect Dis. 2010 Dec;16(12):2021–2.

19. Secondary Serological Response of Patients with Chronic Hepatosplenic Suppurative Brucellosis | Clinical and Vaccine Immunology [Internet]. [cited 2025 Feb 19]. Available from: https://journals.asm.org/doi/full/10.1128/cvi.00086-06

20. Ariza J, Pigrau C, Cañas C, Marrón A, Martínez F, Almirante B, et al. Current Understanding and Management of Chronic Hepatosplenic Suppurative Brucellosis. Clinical Infectious Diseases. 2001 Apr 1;32(7):1024–33.

21. Nowrasteh A, Forrester AC. No Relationship Between Notifiable Diseases and Immigrant Populations. 2020 Apr 13 [cited 2025 Feb 19]; Available from: https://policycommons.net/artifacts/1320880/no-relationship-between-notifiable-diseases-and-immigrant-populations/1924168/

22. Antibody-dependent enhancement of severe dengue disease in humans | Science [Internet]. [cited 2025 Feb 19]. Available from: https://www.science.org/doi/10.1126/science.aan6836

23. Protective effects of the sickle cell gene against malaria morbidity and mortality - The Lancet [Internet]. [cited 2025 Feb 19]. Available from: https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(02)08273-9/abstract

24. Bacterial and Special Pathogens Branch [Internet]. [cited 2024 Dec 8]. Available from: http://medbox.iiab.me/modules/en-cdc/www.cdc.gov/ncezid/dhcpp/bacterial_special/index.html