Teaching and Learning Confounding in the Health Sciences

Felicity Enders, PhD[1], Miranda Kroehl[2], John McGready, PhD[3]
Rickey Carter, PhD[1]
[1]Mayo Clinic, Department of Health Sciences Research, Division of Biomedical Statistics and Informatics, 200 First Street, Rochester, MN 55905
[2]University of Anschutz Medical Campus, Department of Biostatistics and Informatics, 13001 E. 17th Place, Aurora, CO 80054
[3]Johns Hopkins Bloomberg School of Public Health, 615 N. Wolfe Street, Baltimore, MD 21205

Abstract
In order to read or publish in the medical research literature, students in the health sciences need a thorough understanding of confounding. However, research shows that confounding may be poorly understood by some students even after two courses in biostatistics at the graduate level. We introduce two problem-based guided examples to increase both the breadth and depth of students' understanding of this challenging subject. The first is a visual introduction to confounding and interaction in which students naturally lead the discussion through a set of increasingly complex models. The second example links the analysis to study design and compares results for a t test, a t test performed within the context of regression, and an adjusted analysis termed an "adjusted t test" to help link understanding. We also identify topics for improved teaching for such a challenging subject as confounding, with the goal of removing barriers to student understanding.

Key Words: statistics education, confounding, problem-based learning

1. Background

It is critical for students in the health sciences to gain a thorough understanding of confounding, as observational data is prevalent in this field. Indeed, the Strengthening the Reporting of OBservational studies in Epidemiology (STROBE) and Transparent Reporting of Evaluations with Non-randomized Designs (TREND) manuscript guidelines (Des Jarlais, Lyles et al. 2004; Vandenbroucke, von Elm et al. 2007; von Elm, Altman et al. 2007) both refer to statistical methods for assessing and accounting for confounding. Even the Consolidated Standards of Reporting Trials (CONSORT) guidelines (Moher, Hopewell et al. 2010), which targets randomized trials, mentions statistical methods used for adjustment. As such, understanding confounding is required to both read and publish in the medical literature. Confounding is also mentioned in the American Statistical Association's Guidelines for Assessment and Instruction in Statistics Education (GAISE) report for undergraduate statistics education (Aliaga, Cobb et al. 2005), though at that level the primary goal is for students to simply understand that a third variable may cause or hide the association of interest. However, students in the health sciences, who are typically at the graduate level, need to gain a rigorous understanding of confounding.

1.1 The Evidence: How Well is Confounding Understood?

Little information is available on how well students in the health sciences understand confounding, as shown in a recent review of assessment instruments for this population (Enders 2011). Berwick et al included two questions on confounding in their 1981 survey, but did not include the results for individual questions in the resulting paper (Berwick, Fineberg et al. 1981). Novack et al (Novack, Jotkowitz et al. 2006) queried respondents on what should be done after confounding was identified as a problem in the data, but little work has been done to show whether students at the graduate level can accurately identify confounding. However, the statistics education literature for undergraduates can help fill this gap.

delMas et al (delMas, Garfield et al. 2007) assessed 763 undergraduates before and after a first course in statistics, using their Comprehensive Assessment of Outcomes in Statistics (CAOS) test. Item 23 in the CAOS test addresses students' "understanding that no statistical significance does not guarantee that there is no effect." 63% of students achieved a correct response to this question at pretest, and 64% of students correctly answered the question at posttest. The CAOS test also includes a related question, on "understanding of the purpose of randomization in an experiment." 8.5% of students correctly responded to this question at baseline, and 12.3% correctly responded after taking an introductory statistics course.

Felicity Enders has developed an instrument, the REsearch on Global Regression Expectations in StatisticS (REGRESS) quiz, which is designed to assess conceptual understanding of linear regression in the graduate health sciences population. Preliminary evidence from the REGRESS quiz shows that among 49 students in the health sciences completing a course on regression, 75% are able to identify confounding as a potential problem prior to analysis using a scatterplot matrix, while 73% are able to identify whether confounding has occurred through assessing the change in the regression coefficient of interest following adjustment. These results are consistent with those from the CAOS test. Both sources suggest there are a substantial number of students exiting a course in statistics with room for improvement in their understanding of confounding.

2. Improving Teaching of Confounding

It seems likely that there are at least two contributing factors for students' lack of understanding on this topic. First, confounding is often not taught thoroughly. Below, we present two scenarios through which students may gain both a broader and deeper understanding of this topic. Second, as with any topic, there is the possibility that instruction may not be as thoughtful as we would like, perhaps due to pressures of time. We present a series of reminders on how to keep students engaged and learning generated by a group of students who completed a course on regression.

2.1 Learning by Vignette

Case vignettes are a good initial approach to teaching confounding through class discussion in a problem-based learning approach. We provide two examples to stimulate the reader.

2.1.1 A visual introduction to confounding and interaction

Traditional statistics and epidemiologic curricula seem to present confounding and effect modification briefly and simultaneously, often only in the context of 2 by 2 table analyses. This quick treatise not only facilitates the confusion, but gives the impression that confounding and/or effect modification can only occur in the context of a binary outcome and a binary predictor.

As linear regression lends itself best to helpful visualizations, in our first example the differences between confounding and interaction are introduced graphically in a scenario with a continuous outcome. Data from 64 Appalachian sites show elevation and percentage of dead or badly damaged trees in each of the assessed areas (reported by *Committee on Monitoring and Assessment of Trends in Acid Deposition, 1986*)(Committee on Monitoring Assessment of Trends in Acid Deposition Environmental Studies Board, National Research Council 1986). Information is also given about the region of each site (58 sites are "Northern", 8 are "Southern"). Although this example is outside the health sciences, it is so easily grasped by students that it may introduce understanding which can then be leveraged in complex health sciences datasets.

This example proceeds through each of the figures below, with class discussion as each figure is shown. As can observed in the scatterplot, the nature of the unadjusted relationship between damage and elevation is not easily discerned, and if a single regression line is fit the results for not adequately describe the association (Figure 1A). When students see the scatterplot with points labeled by region, they generally note the relationships between region and damage, and region and elevation (Figure 1B). This sparks discussion on how the crude relationship is being missed because of this third party variable's (region) associations with the outcome and predictor of interest. After showing the regression line estimates resulting from the region adjustment model (Figure 1C), students start to debate as to whether the same association between damage and elevation is appropriate given the data. This segues into a discussion about confounding versus interaction, and the difference between the two ideas can be shown explicitly with this data set (Figures 1C and 1D). This also sparks discussion about whether the design of the study adequately allows for the investigation of the role of region in the damage/elevation association, and how a study with this specific question could be conducted.

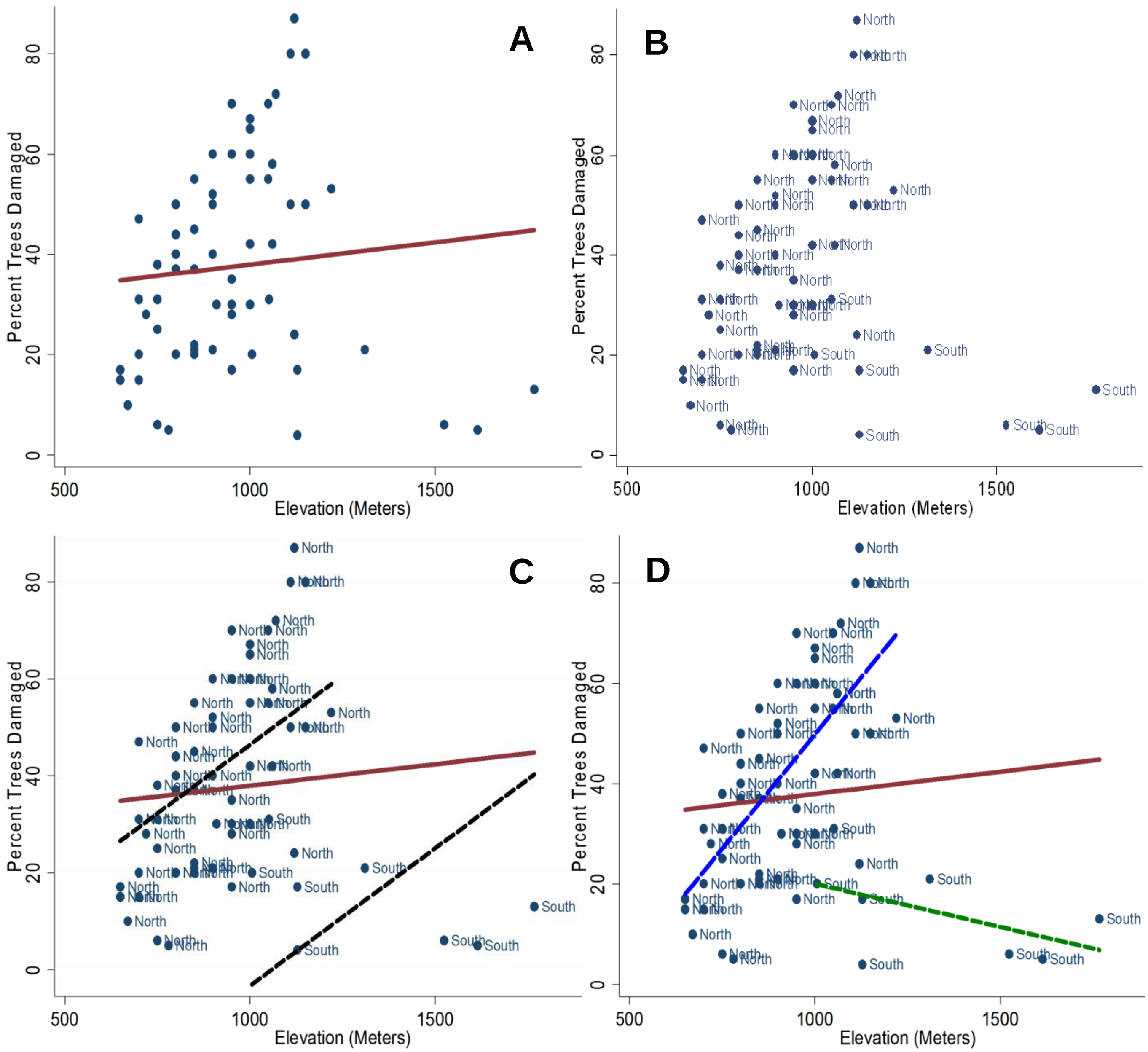

Figure 1: Graphs of percent trees damaged vs. elevation shown with increasingly complex analyses. A. students are told slope is 0.01 (-0.02, 0.03). B. region of each plot is revealed. C. students are told that the unadjusted (red, solid) slope is 0.01 (-0.02, 0.03) while the adjusted (black, dashed) slope is slope is 0.0 6 (0.03, 0.08). D. students are told that the unadjusted (red, solid) slope is 0.01 (-0.02, 0.03) while the Northern (blue, long dashed) sites' slope is 0.09 (0.06, 0.12) and the Southern (green, short dashed) sites' slope is -0.02 (-0.06, 0.02).

Further class discussion can be stimulated with the following series of questions:
We have shown the results from three models:

$$1: \textit{Estimated percent trees damaged} = \hat{\beta}_o + \hat{\beta}_1(\textit{elevation})$$

$$2: \textit{Estimated percent trees damaged} = \hat{\beta}_o + \hat{\beta}_1(\textit{elevation}) + \hat{\beta}_2(\textit{region})$$

$$3: \textit{Estimated percent trees damaged} = \hat{\beta}_o + \hat{\beta}_1(\textit{elevation}) + \hat{\beta}_2(\textit{region}) + \hat{\beta}_3(\textit{elevation} \times \textit{region})$$

- What model corresponds to each of the previous figures?
- What is the comparison being made by $\hat{\beta}_1$ in each of the three models?
- What model results do we compare to assess confounding?
- What model results do we compare to assess interaction?

2.1.2 An applied introduction to confounding in observational data

Another concern raised by the data from the CAOS test above is that students may not understand the association between study design and confounding. Our second example, shown in detail in Table 1, is a problem-based learning exercise focused on taking the broad view of an analysis together with study design, with the goal of helping students think beyond the introductory statistical methods such as the t-test. The example below is motivated by a recently published study (Saad, Man et al. epub 2012) and other research currently underway. The data have been altered by adding random noise and the study design greatly simplified for this summary, but the flowchart of ideas is meant to be representative of any observational study.

After the study information is introduced, students are encouraged to consider the study design. At this step, group discussion may be encouraged as to how best to analyze the data, with the discussion initiated by a proposed t test. The discussion should wind up with the ideas included in Preparation Step 1 (Table 1). After considering study design, the class should review figures comparing study variables between the two groups. The unadjusted analysis is then contrasted with the same unadjusted comparison assessed within a regression framework, with the result that students observe the striking similarity of the results between these analyses. This similarity serves as the basis of referring to regression as an “adjusted t test.” In this example, the coefficient for T1DM is largely unchanged in the adjusted model. The stability of these findings is encouraging in showing that accounting for differences in body mass index has not changed the estimated relationship between study group and Si. In this case, one can review the histogram for BMI from the preparation phase to link the lack of confounding to the lack of a distinction in BMI by study group. This discussion can serve as a reminder to students that a variable can only act as a confounder if is associated with both the predictor and the outcome. Students may then begin to wonder about the possibility of age acting as a confounder, illustrating the iterative nature of statistical analysis.

Table 1: Flowchart of information presented in our second example

Study Information

A study was recently conducted consisting of 20 type 1 diabetics (T1D) and 20 non-diabetic (non-DM) controls. The participants were given a ‘labeled’ meal and the Insulin Sensitivity index (Si) was measured. Insulin sensitivity is a measure of the overall effect of insulin to stimulate glucose disposal and inhibit glucose production. We seek to test if

Si is different between T1D and non-DM. Higher values of Si indicate better functioning.

Preparation Step 1: Consider the Study Design

Our example is a two group comparison, so all we need to use is a t test. . . . maybe. We have a non-randomized design. We may be concerned with potential confounding. Scientific goal to keep in mind: Prove beyond a reasonable doubt

How?

1. Checking assumptions and validity of approach (importance of study design)
2. Sensitivity analyses (adjustment for confounding)
3. What other literature supports
4. Peer review–anticipate questions, think like a skeptic

Preparation Step 2: Look at the Data

Histograms for Si, Age, and body mass index (BMI) are all shown. Si has a greater mean and greater variability in the Non-DM group. Average age is larger in the T1D group, while BMI appears similar in both groups.

Unadjusted Analysis: t test

Methods:	Results:
Calculated with pooled variances, as appropriate given graphical analysis of the two groups.	Estimated difference in means: -6.01 T test: t = -3.02, p = 0.005

Unadjusted analysis: "t test in a regression framework"

Segue: Another way to do compare a continuous outcome between two study groups is linear regression with a binary predictor.

Methods:	Results:
X is a binary indicator variable. t-test is for indicator variable	Estimated difference in means: -6.01 T test: t = -3.02, p = 0.005 F test: 9.11, p = 0.005

Adjusted Analysis: "adjusted t test"

Segue: Body Mass Index is a component of metabolic syndrome, high fasting blood glucose, and associated with T2D. While the groups look reasonably balanced on BMI (it was controlled through inclusion/exclusion criteria), a simple analysis is to adjust for BMI in our regression model.

Methods:	Results:
T1DM is a binary indicator variable. Continuous BMI also in model. Estimated mean difference and t-test are for T1DM.	Estimated difference in means: -6.04 T test: t = -3.03, p = 0.004

Potential Next Steps

Students could estimate:

1. Confidence interval for the mean Si for Non-DM and T1M
2. Confidence interval for the difference between the groups
3. Prediction interval for either group
4. Estimate a measure of effect (standardized group difference and coefficient of determination)
5. Students could learn when and how to interpret intercepts

2.2 Hints for thoughtful statistical teaching… from students

Introducing difficult statistical concepts to a novice is a challenge for even the most experienced teacher. Furthermore, professors and instructors teaching at the graduate level rarely receive any pedagogical training; they are almost certainly experts in the topics of statistics, but they might not be adequately prepared to aid others in understanding such complex matters. While there has been little research on effective teaching methods for graduate level statistics, some general pedagogical strategies can be applied to almost any classroom to aid in the students' abilities to learn and retain knowledge.

Perhaps one reason that complicated statistical topics, such as confounding, are not well-understood by health sciences students results from a lack of ability to successfully transfer classroom knowledge to outside areas. Developing competencies in any new subject involves not only initial learning, but the transfer of learning, which can be defined as the application of knowledge learned in one context to another. The evaluation of transfer by asking the student to solve a set of problems after engaging in the initial learning task, as is often done with homeworks covering the current course topics, can often seriously underestimate the amount of transfer reached by a student. It is important for instructors to instead view transfer as a dynamic process in which the students return to previous concepts as they are introduced to new ones, building on the different contexts to gain a deeper understanding (Committee on Developments in the Science of Learning, National Research Council 2000).

Both education research and personal experiences will tell us that teacher quality is an important aspect to the students' ability and desire to learn. It is possible that an instructor can improve the quality of their teaching simply by taking the time to prepare, in advance, a well-organized course. While this requires substantial upfront work for a semester-long course, it forces the instructor to identify the important topics and not spend unnecessary time on less critical information. Consequently, the time spent on complex topics appearing later in the semester will not be sacrificed to too much time spent on simpler content. In addition, a well-structured course will allow for more opportunities to repeat previous topics in different contexts, allowing for the dynamic nature of knowledge transfer.

There are many ways that a well-organized course might be developed. Here, we will describe one format that has been successful within a traditional lecture-based setting. In this introductory linear regression course, students were taught general regression topics in a lecture-based format. After the course concluded, some of the students were interviewed as part of an informal focus group to assess which aspects of the teaching process were most helpful. The suggestions below are the result of that review.

Prior to the start of the semester, the professor had determined what topics would be covered on what days, and prepared the lectures to ensure that she would have time to cover the intended information within the pre-determined timeframe. Lectures were presented in PowerPoint, and handouts of the presentation were provided to students as the basis for course notes.

There were two small but important aspects to the lecture notes which helped to remove potential learning barriers during lecture. The first was that all information on the slides was discussed during the lecture, and anything that would not be covered during lecture would be provided in a supplemental handout. This seemingly small practice is significant because it helps students to remain focused on the discussion, and not become distracted by trying to read and understand the material that was skipped over. The other noteworthy item was to ensure that all mathematical steps in an example, no matter how small or “intuitive,” were provided in the notes. Again, this helped the students to remain focused on the important statistical concept at hand, and not become distracted with working through the mathematics.

Throughout the semester, important topics would be revisited both during lecture and within the homeworks. For example, the assumptions of linear regression were discussed during each lecture for the first several weeks of class, and revisited several times throughout the remaining semester, emphasizing that statistical analysis should not be treated as a black box. Finally, students were required to write-up everything as if they were providing a report for an investigator, including a description of their approach and interpretation of results. This method helped the students to move out of the context of solving a homework problem to synthesize their understanding into a real-world application of statistical techniques.

3. Conclusion

We believe it is possible to use brief problem-based learning exercises to add considerable depth to students’ understanding of confounding. We have introduced two such examples, both of which are flexible in their presentation yet powerful in terms of learning value.

Neither of the examples we introduce is intended to be used in a vacuum. We believe confounding is best taught through repetition, when it is included throughout the curricular topics of study design, bivariate analysis, linear, logistic, and Cox regression, and discussions on how to read a research manuscript. This serves not only to continually reinforce the issue of confounding, but also emphasizes one of the main utilities of regression methods in public health and medical research. A regression framework helps students think outside the box. A simple illustration such as these opens the door to more advanced analyses. Students like the idea of an “adjusted T-Test”, especially those that have had an epidemiology course. However, ideas such as these do take time for students to digest. Truly ‘thinking outside the box’ is more difficult, requires more depth in training, and requires critical evaluation of competing options.

Most of the research in statistics education to date has focused on the K-12 and undergraduate level, for whom confounding need not be taught in the depth at which it is needed at the graduate level in the health sciences. As a result, little work has yet done on how confounding might be better taught. While we have introduced two vignettes to teach confounding thoroughly, we believe research is needed in this area. Are some teaching vignettes more successful than others? Should confounding be taught as a single topic within a course or spread throughout the course within the context of other related topics? Do students understand confounding well when it is introduced separately in courses on epidemiology and statistics? How well are students able to articulate or apply concerns regarding confounding? Above all, are these topics retained after course completion?

Working knowledge of regression and confounding is absolutely essential to successful practice as a health sciences researcher. As research funding becomes more and more limited in relation to the number of medical researchers, we anticipate an increase in observational studies. Students in the health sciences will need a thorough understanding of confounding in order to have sufficient biostatistical literacy to read the medical literature. An even greater level of understanding is required for those reporting their work in the literature. Finding ways to teach confounding well will help us provide a basis of strong research methods and real biostatistical literacy for our students.

References

Aliaga, M., G. Cobb, et al. (2005). Guidelines for Assessment and Instruction in Statistics Education (GAISE): College Report. Alexandria, VA, The American Statistical Association.

Berwick, D. M., H. V. Fineberg, et al. (1981). "When doctors meet numbers." AM J Med 71(6): 991-998.

delMas, R., J. Garfield, et al. (2007). "Assessing Students' Conceptual Understanding After A First Course in Statistics." Statistics Education Research Journal 6(2): 28-58.

Des Jarlais, D., C. Lyles, et al. (2004). "Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement." Am J Public Health 94(3): 361-366.

Enders, F. (2011). "Evaluating Mastery of Biostatistics for Medical Researchers: Need for a New Assessment Tool." Cts-Clinical and Translational Science 4(6): 448-454.

Committee on Developments in the Science of Learning, National Research Council (2000). How People Learn: Brain, Mind, Experience, and School: Expanded Edition, The National Academies Press.

Moher, D., S. Hopewell, et al. (2010). "CONSORT 2010 Explanation and Elaboration: updated guidelines for reporting parallel group randomised trials." BMJ 340:c869: 1-28.

Committee on Monitoring Assessment of Trends in Acid Deposition Environmental Studies Board, National Research Council (1986). Acid Deposition:Long-Term Trends, The National Academies Press.

Novack, L., A. Jotkowitz, et al. (2006). "Evidence-based medicine: assessment of knowledge of basic epidemiological and research methods among medical doctors." Postgrad Med J 82: 817-822.

Saad, A., C. Man, et al. "Diurnal Pattern to Insulin Secretion and Insulin Action in Healthy Individuals." Diabetes.

Vandenbroucke, J. P., E. von Elm, et al. (2007). "Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration." PLoS medicine 4(10): e297.

von Elm, E., D. G. Altman, et al. (2007). "The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies." Bulletin of the World Health Organization 85(11): 867-872.