TEACHING TIPS

Tips for Teachers of Evidence-Based Medicine: Adjusting for Prognostic Imbalances (Confounding Variables) in Studies on Therapy or Harm

Cassie C. Kennedy1, Roman Jaeschke3, Sheri Keitz5,6, Thomas Newman7,8,9, Victor Montori1,2, Peter C. Wyer10, and Gordon Guyatt4 for the Evidence-Based Medicine Teaching Tips Working Group

1Department of Medicine, Mayo Clinic College of Medicine, Rochester, MN, USA; 2Knowledge and Encounter Research Unit, Mayo Clinic College of Medicine, Rochester, MN, USA; 3Department of Medicine, McMaster University, Hamilton, ON, Canada; 4Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, ON, Canada; 5Miami Veterans Affairs Medical Center, Miami, FL, USA; 6University of Miami Miller School of Medicine, Miami, FL, USA; 7Department of Epidemiology, University of California, San Francisco, CA, USA; 8Department of Biostatistics, University of California, San Francisco, CA, USA; 9Department of Pediatrics, University of California, San Francisco, CA, USA; 10Columbia University College of Physicians and Surgeons, New York, NY, USA.

KEY WORDS: evidence-based medicine; teaching tips; therapy.

The notion that prognostic differences between treatment groups might explain apparent differences in outcomes lies at the heart of clinical investigation. If clinicians are to understand the research studies that guide their clinical practice, they must grasp this concept. Further, clinicians should understand statistical adjustment for differences in prognostic variables. This tip describes approaches to conveying these concepts to learners.

As with other articles in this series, educators experienced in teaching evidence-based medicine developed these tips and used them extensively to clarify statistical concepts to learners solving clinical problems or critically appraising studies and reviews. An article from the Canadian Medical Association Journal described the development of this series and pertinent background information.1 For each of the tips we provide guidance on when to use them, a teaching script, a “bottom line,” and a summary card. For each tip we identify the appropriate target audience and provide time estimates for the exercise.

We present qualitative introductions to the concepts of confounding variables and adjusted analysis, followed by quantitative examples illustrating the potential impact of adjustment. These demonstrations are characteristically interpolated into clinically centered discussions with clinical learners. We do not in the course of using them routinely attempt to address performing or analyzing multivariable or multivariate analyses with such learners. We consider learner mastery of the concept of relative risk to be prerequisite for understanding the tips, particularly for tip 3. A previously published manuscript addressed teaching relative risk.2

TEACHING TIP 1: UNDERSTANDING COFIGUING VARIABLES

When to Use This Tip

This tip is suitable for those early in the process of learning critical appraisal. This exercise takes 10 to 15 minutes to teach and has the following specific goals:

- Explain the concept of a “prognostic variable”
- Teach how randomization can minimize confounding

Electronic supplementary material The version of this article (DOI:10.1007/s11606-007-0391-1) contains supplementary material, which is available to authorized users.
This tip is used when appraising the adequacy of randomization in a trial or for teaching the concepts of randomization and prognostic variables. Begin by asking the learners why clinical trials are randomized. Typical initial answers include “reduce bias” or “ensure validity.” Eventually, learners respond, “to ensure that the control and experimental groups are similar or balanced.” Focus on this answer and ask “similar (or balanced) with respect to what?” Typical answers include “age and sex.” Ask which patient characteristics, aside from age and sex, should ideally be balanced. Look for any answer relating to patient prognosis such as severity of disease, smoking history, or comorbid condition. Distribute or cite a specific therapy article to focus this discussion on a specific patient population. Depending on the context at hand, the learners may already have a suitable article in front of them. Ask the learners why the 2 groups should be balanced with respect to age, sex, disease severity, habits, and/or comorbidities. To illustrate the answer, contrast the above variables (likely to be prognostic) with variables unlikely to be prognostic—asking whether the experimental and control groups need to be balanced with respect to eye color or shoe size. Of course the answer is no. The key is to help learners identify why not. Ask why age (or sex, habits, etc.) is different from eye color. Learners might respond that “they are related to developing the disease” or “they affect the response to treatment.”

Eventually, someone will articulate that some patient attributes help predict future events. Reinforce the concept that age, sex, disease severity, habits, and comorbidities predict subsequent clinical events—they are prognostic factors. Now focus on the future event by asking “prognostic of what?” Learners may need coaching to identify that investigators are interested in prognostic factors for the relevant outcome(s) for their study. You then state that randomization is designed to balance groups for the relevant prognostic factors. Ideally, all treatment evaluations would compare 2 populations identical in all prognostic factors related to the outcome of interest. This would ensure that the only relevant baseline difference between the intervention group and the control group is whether they receive treatment. This enables clinicians to clearly evaluate the effect of an intervention on the outcome of interest (treatment effect). The reason we are more concerned about imbalance in age than in eye color or shoe size is that increasing age is likely associated with a higher risk for most important adverse outcomes, whereas differences in shoe size and eye color very likely are not so associated. A different distribution of age in the 2 groups can bias the results; a difference in distribution of eye color would likely not bias the results.

For example, in a recent study, critically ill patients with acute lung injury and acute respiratory distress syndrome (ARDS) were randomly assigned to either a lower positive end-expiratory pressure (PEEP) or a higher PEEP. By chance, the mean age of the lower PEEP group was significantly younger than the mean age of the higher PEEP group (49 vs 54 years). The results showed a rate of death from any cause of 24.9% in the lower PEEP group and of 27.5% in the higher PEEP group. The results raised the question of whether the mortality differences were attributable to the intervention or to the age differences. At the heart of this question is the concept of confounding variables. When the distribution of a known prognostic factor differs in the groups under comparison, it becomes a potentially confounding variable—which, in the absence of adjustment, can lead to overestimating or underestimating the true treatment effect.

Then ask learners whether the problem of confounding could be solved by matching groups with respect to all known prognostic variables. Typically, learners raise concerns about the feasibility of achieving an exact match between the 2 groups with respect to all such variables. The preceptor should acknowledge these barriers, but persist in asking whether, if feasibility barriers can be overcome, matching would be an appropriate alternative to randomization. Eventually, a learner volunteers that unmeasured or unknown confounders could still be unbalanced across the groups. Reinforce this concept with an example such as The Nurses’ Health Study, an observational cohort study of hormone replacement therapy (HRT) to prevent cardiovascular events. Because the nurses were not randomized to HRT or control, we cannot know why only some nurses chose HRT. There were many known, and likely many unknown, prognostic factors imbalanced between the 2 groups. The adjusted analysis, which controlled for the known confounding variables (such as smoking history, hypertension, etc.), demonstrated an apparent decrease in cardiovascular events with HRT. However, a large randomized controlled trial (producing 2 groups homogeneous for known and unknown prognostic factors) found an increased cardiovascular risk in women on HRT. The difference in results is almost certainly caused by unknown or unmeasured prognostic factors because women choosing HRT in the Nurses Health Study were destined to have a lower rate of cardiovascular events irrespective of using HRT. The benefit was no longer apparent in a randomized study designed to balance prognostic factors. This example demonstrates the limitations of adjustment or matching and reinforces the benefits of randomization.

The Bottom Line

- Randomization is the optimal approach for balancing the intervention and control groups with respect to patient characteristics associated with the outcome of interest.
- Such characteristics are called prognostic variables, and if unbalanced between groups, confounding variables.

See Appendix 1 for the summary card on this tip.

TIP 2: UNDERSTANDING ADJUSTMENT—QUALITATIVE DEMONSTRATION

When to Use This Tip

Adjusted analysis arises in 2 contexts: randomized trials, especially if prognostic variables are unbalanced, and observational studies. In either situation, users of the medical literature must consider the issue of adjustment when critically appraising an article. Tips 2 and 3 help facilitate learners’ understanding of adjusted analyses.

Tip 2 is suitable for beginners or intermediate learners who understand the concept of why randomized trials provide the best way of balancing prognostic factors, and thus understand the essential concept of confounding. This exercise takes about 15 minutes to teach.
This tip has the following goal:

- Demonstrate the fundamental logic underlying adjusted analyses.

The Script

Present the group with the following problem. A trial contains an intervention and a control group with dissimilar ages. You suspect that age is associated with the outcome of interest (and thus is a prognostic, and potentially confounding, variable) in this study. The distribution of age is as follows: in the treatment group, 80% of subjects are young and 20% are old. In the control group, 20% are young and 80% are old.

In this trial, treated patients appear to have superior outcomes to untreated patients. Ask the group for possible explanations. Students usually observe, “Treated patients are younger.” You agree, and reiterate that spurious treatment effects can result from a prognostic imbalance. Another possible explanation for the observed outcomes is that “treatment really works.” For purposes of clarity in this exercise, you should probably ignore the third possible answer—chance.

Next, ask the learners how to differentiate between the possibilities of a true benefit versus a spurious apparent benefit as a result of confounding. Learners often struggle with this challenge and answer “regression” or “logistic regression,” but are unable to further elaborate. You might need to provide the answer: calculate estimates of effect separately for the prognostic groups. In this case, the estimates of treatment effect for the young treatment, young control, old treatment, and old control groups are calculated separately. Next, combine effects across the young and old groups to calculate an overall “adjusted” effect. This process may be described as “creating a level playing field,” “creating 2 comparisons in which groups are homogenous for the prognostic variable,” “creating 2 comparisons in which patients in treatment and control group are prognostically similar,” or “creating an unconfounded analysis.”

The same process can be followed for several variables. Inform the learners that this study also has patients with and without diabetes that may influence the outcome of interest. A volunteer (or the preceptor) may list the 4 unconfounded comparisons that result: old diabetic treatment, old diabetic control; young diabetic treatment, young diabetic control; old non-diabetic treatment, old non-diabetic control; and young non-diabetic treatment, young non-diabetic control.

This process could continue for as many such prognostic groups (within the limits of sample size) as necessary. This principle can be extended to categorical variables with more than 2 categories (e.g., cancer stages) or to continuous variables (e.g., age). In each case, the principle is to create prognostically homogeneous groups, make comparisons between treatment and control within these groups, and then combine treatment effect estimates across the groups to obtain an adjusted estimate of the treatment effect. Computer programs for making such adjusted comparisons are widely available.

We return to the previous example of critically ill patients with acute lung injury and ARDS who were randomly assigned to either a lower PEEP or a higher PEEP. By chance, the mean age of the lower PEEP group was significantly lower than the mean age for the higher PEEP group (49 vs 54 years). The unadjusted results showed a rate of death from any cause of 24.9% in the lower PEEP group and of 27.5% in the higher PEEP group. However, the adjusted death rates are in the opposite direction: 27.5% in the lower PEEP group and 25.1% in the higher PEEP group.

The Bottom Line

- Studies that do not balance groups with respect to known prognostic factors may underestimate or overestimate treatment effects.
- The principle of an adjusted analysis involves creating groups that are homogeneous for known prognostic variables and then combining intervention effect estimates across groups.

See Appendix 1 for the summary card for this tip.

Tip 3: Understanding Adjustment—Quantitative Demonstration

When to Use This Tip

This tip is useful for any learner who has worked through tips 1 and 2. Typically, this follows immediately after tip 2. Use this tip for intermediate to advanced learners or anyone who is particularly interested in a deeper understanding and who already understands the concept of relative risk. This exercise takes 15 to 20 minutes.

This tip has the following goal:

- Reinforcing the concepts of confounding variables and adjustment.

Return to the potentially confounded comparison presented above in tip 2. Ask the group to make the following assumption: the event rate is 10% in the young and 20% in the old and there is no treatment effect. Ask the group to help construct a 2 × 2 table with 100 patients in the treatment and 100 patients in the control groups. How many events can we expect in treatment patients (80 young and 20 old)? The answer is 

\[
\text{Risk in young treatment} = (80 \times 0.1) + (20 \times 0.2) = 8 + 4 = 12.
\]

Enter this in a 2 × 2 table (Table 1). Lead the group through the same calculations for the control group (20 young and 80 old): 

\[
(20 \times 0.1) + (80 \times 0.2) = 2 + 16 = 18.
\]

Use these numbers to complete the 2 × 2 (Table 1).

Now ask the group to calculate the relative risk. Remind the group that the relative risk is the risk of an outcome in 1 group divided by the risk of the outcome in the other group. Facilitate the process by asking the group to state the risk in the treatment group (12%) and the risk in the control group (18%). Therefore, the relative risk, obtained by dividing the risk in the treatment group by the risk in the control group, is 0.67.

| Table 1. A 2×2 Table Used in Demonstration of Tip 3 |
|---------------------|---------------------|
|                    | Dead | Alive |
| Treated            | 12   | 88    |
| Control            | 18   | 82    |

Risk in treatment group: 12 of 100 or 12%; risk in control group: 18 of 100 or 18%; relative risk=12%/18%=0.67.
Next, direct the group back to the unconfounded comparison. Ask them to calculate the mortality risks for the treated (n=80) and control (n=20) young patients. Assuming no effect of treatment on mortality risk, 10% (n=8) of the 80 younger treated patients and 10% (n=2) of the younger control patients died. Because the mortality risks in the 2 groups are both 10% the corresponding relative risk is 1.0. Next, calculate the mortality risks among the old patients. Again, assuming no treatment effect on mortality risk, 20% (n=4) of the 20 old treated patients and 20% (n=16) of the 80 old control patients died; this also yields a relative risk of 1.0. This is demonstrated in Table 2.

This example points out that an adjusted analysis combines the relative risk for 1 prognostically homogeneous group (the young, 1.0) with the relative risk in another prognostically homogeneous group (the old, 1.0) to generate the relative risk for the entire group (1.0). You may mention that this is called a stratified or adjusted analysis.

Ask the group to reflect on the relative risk of 0.67 generated for the entire patient pool. The group will conclude that the apparent difference in treated patients was due completely to the confounding variable (age). Point out to the learners that the analysis could be adjusted for more than 1 factor as demonstrated conceptually in Tip 2. In fact, this is the principle that computers follow when performing multivariable analyses

## Bottom Line
- Imbalance of prognostic variables between groups can create spurious treatment effects. Adjusted analysis seeks to better estimate the true effect of treatment.

See Appendix 1 for the summary card for this tip.

## Report on Field Testing
We field tested the tips to verify the clarity and practicality of the descriptions. Field testing frequently generates examples of the kinds of variations in approach that occur when an experienced teacher of evidence-based medicine adapts the approaches to their own style, context, and learner level. S.K. conducted a field test of all 3 tips with 25 resident learners, half of whom were interns. Overall, these tips were rated highly by the learners and by S.K. The objectives were felt to be clear and the learners were able to articulate the major teaching points from the exercises. Preparatory time was approximately 2 hours. The field test utilized a board, data projector, and a handout.

Trainees from prior field tests have consistently requested that teaching tips were grounded in an article or specific case example to best appreciate the relevance of the content. Thus, S.K. opted to include a brief review of the NEJM article as an adjunct to Tip 1. Learners were asked to review the abstract, baseline characteristics for confounders, and then Table 5 of that article, which demonstrates the adjusted and unadjusted analysis. This required an additional 10 minutes. Tip 1 was recommended for beginners, and was appropriate based on field testing. S.K. recommended an abbreviated version of Tip 1 for more experienced learners. Tip 2 was recommended for beginner or intermediate learners and this also seemed appropriate based on field testing. Tip 3 was recommended for intermediate to advanced learners, and indeed the more novice learners required coaching through the mathematical formulas associated with Tip 3. Based on learner feedback, the concepts of relative risk (or risk ratio) and number needed to treat might be helpful prerequisites to this series of exercises.

The overall strength of the tips was the stepwise building on the same concept from an introductory to an advanced level. One point of clarification was required during the first tip: the difference between the authors’ intent for randomization and the actual achievement of balanced prognostic factors. In addition, S.K. felt it may be useful in the future to add a coin flip exercise to Tip 1, to randomize the learners. After the 2 groups are randomly allocated to either heads or tails, examine the distribution of males versus females, married versus single, etc., to illustrate that sometimes randomization falls short of its intention to have equal groups. S.K. also suggests a modified approach to Tip 3 by using a completed 2×2 table to calculate overall relative risk and number needed to treat to see the effect of a given treatment. She would then alert the group to the fact that the treatment and control populations were confounded. Finally, she would have the learners sort out the contribution of the confounding variable. S.K. believes that the impact of a confounding variable might be powerfully demonstrated this way.

## Table 2. Demonstration of Effect of Adjustment for Age on Observed Relative Risk

| Patients            | Risk in treatment group | Risk in control group | Relative Risk=Risk in treatment group/Risk in control group |
|---------------------|-------------------------|-----------------------|------------------------------------------------------------|
| Young patients      | 8 of 80=0.1 or 10%      | 2 of 20=0.1 or 10%    | 0.1/0.1=1.0                                               |
| Old patients        | 16 of 80=0.2 or 20%     | 16 of 80=0.2 or 20%   | 0.2/0.2=1.0                                               |

## CONCLUSION
Randomization is performed in clinical trials to achieve balance in prognostic factors. Confounding variables are imbalances in prognostic factors. These imbalances can lead to overestimating or underestimating the impact of an intervention under study. The known confounding variable can be adjusted for by performing the calculations of risk, risk difference, and relative risk for populations after stratifying by the additional prognostic factor.

## Conflict of Interest Statement
None disclosed.

**Corresponding Author:** Peter C. Wyer, 446 Pelhamdale Avenue, Pelham, NY, USA 10803 (e-mail: pw91@columbia.edu).
REFERENCES

1. Wyer PC, Keitz S, Hatala R, et al. Tips for learning and teaching evidence-based medicine: introduction to the series. Can Med Assoc J. 2004;171:347–8.

2. Barratt A, Wyer PC, Hatala R, et al. Tips for teachers of evidence-based medicine: 1. Relative risk reduction, absolute risk reduction and number needed to treat. Can Med Assoc J. 2004;171:353–8 (Online Appendix).

3. Guyatt G, Cook D, Devereaux PJ, Meade M, Straus S. Therapy. In: Guyatt G, Rennie D, eds. Users’ Guides to the Medical Literature: A Manual of Evidence-Based Clinical Practice. Chicago: AMA Press; 2002:55–79.

4. Brower RG, Lanken PN, MacIntyre N, et al. Higher versus lower positive end-expiratory pressures in patients with the acute respiratory distress syndrome. New Engl J Med. 2004;351:327–36.

5. Grodstein F, Manson J, Stampfer M. Postmenopausal hormone use and secondary prevention of coronary events in the nurses’ health study: a prospective, observational study. Ann Intern Med. 2001;135(1):1–8.

6. Rossouw JE, Anderson Gl, Prentice RL, et al. Risks and benefits of estrogen and progestin in healthy menopausal women: principal results from the Women’s Health Initiative randomized controlled trial. JAMA. 2002;288:321–3.
APPENDIX 1

Summary cards for 3 teaching tips on understanding, interpreting and quantifying confounding variables in studies of therapy or harm.

This Appendix has been designed so that it can be printed on a single sheet of 8 1/2 ×11 in. paper. The individual summary cards can then be cut out, if desired, for use during teaching sessions.

Teaching Tip 1: Understanding Confounding Variables

**Scenario:** Consider a randomized trial, in contrast to an observational study, on an issue of therapy or harm. Pose the issue: “Why do we randomize?” Lead the learners to understand the answer in successively more precise terms.

1. Randomization helps ensure that groups have similar characteristics.
2. Contrast characteristics such as eye color and shoe size to age and disease severity.
3. Lead the learners to understand the relevant characteristics to be those that affect outcome.
4. Lead the learners to distinguish between any outcomes and outcomes of interest to the question being studied.
5. Label such characteristics “prognostic variables” and clarify that they are appropriately called “confounding variables” when they are unequally distributed between study groups.
6. Lead the learners to understand that randomization minimizes the likelihood of both known and unknown confounding variables, whereas observational studies are only able to control for known confounding variables.

**Summary Points.**

- Randomization is the optimal approach to balancing the intervention and control groups with respect to patient characteristics associated with the outcome of interest.
- Such characteristics are called prognostic variables, and if unbalanced between groups, confounding variables.

Teaching Tip 2: Understanding Adjustment—Qualitative Demonstration

**Scenario:** Present the learners with a hypothetical study in which age affects the outcome of interest and potentially is a confounding variable.

1. Tell the learners that 80% of patients in the treatment group are young, in contrast to 20% of patients in the control group. Represent these ages on the blackboard.
2. Tell the learners that the study results suggest that treated patients have superior outcomes to untreated patients. Ask the learners for possible explanations for this difference.
3. After the learners determine that the observed difference could be attributable to either true effect of the therapy or to the difference in ages among the study groups, challenge the learners to propose a way of distinguishing between these 2 possibilities.
4. Guide the learners to the proposal to assess the effect of therapy in young and old groups separately. Represent the resulting 4 groups of patients, treatment, and control for both young and old patients on the blackboard.
5. Lead the learners to understand that the effects of additional confounders may be tested through extending the process, with each test serving to double the number of groups in pairs.

**Summary Points.**

- Studies that do not balance groups with respect to known prognostic factors may underestimate or overestimate treatment effects.
- The principle of an adjusted analysis involves creating groups that are homogeneous for known prognostic variables and then combining intervention effect estimates across groups.
Teaching Tip 3: Understanding Adjustment—Quantitative Demonstration

Scenario: Use the scenario described in Tip 2, adding the assumption that the treatment has no effect on outcome and that the event rate is 10% in younger patients and 20% in older patients.

1. Construct a 2×2 table with rows labeled “treated” and “controls” and the columns labeled “dead” and “alive”, respectively. Lead the groups in calculating the values to be assigned to the 2 cells corresponding to the treatment group. The first cell becomes (10% mortality × 80 young patients) + (20% mortality × 20) old patients = 8 + 4 = 12. The second cell is 88.

2. Repeat the process for the 2 cells in the control group, yielding 18 patients and 82 patients, respectively.

3. Have the learners determine that the resulting relative risk is 12% divided by 18% = 0.67.

4. Have the learners perform the same calculation for the younger and older patients separately, referring to the representation of the groups given previously at the beginning of Tip 2 and adhering to the above assumptions. They find that the relative risk for mortality is 1.0 in both groups when young and old patients are analyzed separately.

5. Explain to the group that the final step in the adjusted analysis is to combine the 2 relative risks, in this case 1.0 for both young and old, respectively, yielding a revised estimate of the relative risk pertaining to all patients, in the case 1.0

Summary Point.

- Imbalance of prognostic variables between groups can create spurious treatment effects. Adjusted analysis seeks to better estimate the true effect of treatment.