The Limitations of Statistical Adjustment

Kristin Sainani, PhD

It is common to use multivariate regression (such as linear or logistic regression) to statistically "control" or "adjust" for confounding variables in observational studies. However, many researchers and readers have only a sketchy understanding of what statistical adjustment is and what its pitfalls are. This article will attempt to demystify statistical adjustment and highlight its limitations. In particular, readers should be aware that some leftover (residual) confounding may remain even after adjustment for important confounders.

CASE STUDY

As an example, I will refer to an article that prospectively examined the relationship between meat intake and mortality using data from a large cohort of older people [1]. For simplicity, I will focus only on the results for red meat and men (n = 322,263), although findings were similar for processed meat and women.

The study provides a nice example of how risk factors tend to cluster. Those people who consumed the most red meat (highest quintile of intake) were heavier, smoked more, exercised less, were less educated, and ate more calories and fat and fewer fruits and vegetables than those who consumed less red meat (see Table 1). In other words, red meat eaters tended to be all-around more unhealthy; and the fact that they died faster is not surprising. The question is, how do we isolate red meat as a causative factor in their mortality versus any of their other risky characteristics? This is the aim of statistical adjustment: to tease out the independent effect of one risk factor out of many.

WHAT IS STATISTICAL ADJUSTMENT?

Consider a simple hypothetical study to estimate the effect of salt intake on blood pressure; suppose that age and smoking are the major confounders. To isolate the effect of salt, one could include only participants with the same age and smoking status, for example, only 50-year-old smokers. However, this is impractical and makes the results applicable to only a few people. Alternatively, one could divide the data into unique levels of age and smoking (50-year-old smokers, 60-year-old nonsmokers, etc), estimate the effect of salt for each subgroup separately, and then average these (weighted by subgroup size) into a single estimate. This approach also is impractical because some subgroups will be too small (eg, there may only be one 50-year-old smoker). However, this is essentially what statistical adjustment accomplishes using some elegant math; it finds the average effect of salt on blood pressure across fixed levels of age and smoking.

Practically, statistical adjustment is performed through the use of multivariate regression models, which can estimate the effects of multiple predictors (also called independent variables) simultaneously. Adjusting for a particular variable is accomplished by including that variable as a predictor in the model. For example, to control for age and smoking, researchers in our hypothetical example fit a linear regression model that included blood pressure as the outcome variable and salt intake, age, and smoking (ever/never) as the predictor variables. The resulting (hypothetical) equation is:

$$\text{Predicted systolic blood pressure (mm Hg)} = 60 + 5 * \text{salt consumption (tsp/d)} + 1 * \text{age (years)} + 10 * \text{ever smoker (yes} = 1, \text{no} = 0)$$

K.S. Division of Epidemiology, Department of Health Research and Policy, Stanford University, HRP Redwood Bldg, Stanford, CA 94305. Address correspondence to: K.S.; e-mail: kcobb@stanford.edu
Disclosure: nothing to disclose

Disclosure Key can be found on the Table of Contents and at www.pmrjournal.org

Submitted for publication June 21, 2011; accepted June 21, 2011.

PM&R
1934-1482/11/$36.00
Printed in U.S.A.

Vol. 3, 868-872, September 2011
DOI: 10.1016/j.pmrj.2011.06.006

Table 1. *Baseline characteristics of the men from the meat intake and mortality study**

Characteristics for Men (n = 322,263)†	Red Meat Intake Quintile, g/100 kcal				
	Q1	Q2	Q3	Q4	Q5
Meat intake					
Red meat, g/1000 kcal	9.3	21.4	31.5	43.1	68.1
White meat, g/1000 kcal	36.6	32.2	30.7	30.4	30.9
Processed meat, g/1000 kcal	5.1	7.8	10.3	13.3	19.4
Age, y	62.8	62.8	62.5	62.3	61.7
Race, %					
Non-Hispanic white	88.6	91.8	93.1	94.0	94.1
Non-Hispanic black	4.2	3.2	2.7	2.2	1.9
Hispanic/Asian/Pacific Islander/American Indian/Alaskan native/unknown	7.2	5.0	4.2	3.8	4.0
Positive family history of cancer, %	47.0	47.7	48.4	48.6	47.8
Currently married, %	80.8	84.4	86.1	86.7	85.6
Body mass index	25.9	26.7	27.1	27.6	28.3
Smoking history, %					
Never smoker	34.4	30.5	28.8	27.6	25.4
Former smoker	56.5	58.1	57.5	57.1	55.8
Current smoker or having quit <1 y prior	4.9	7.6	9.9	11.4	14.8
Education, college graduate or postgraduate, %	53.0	47.3	45.1	42.3	39.1
Vigorous physical activity ≥5 times/wk, %	30.7	23.6	20.5	18.6	16.3
Dietary intake					
Energy, kcal/d	1899	1955	1998	2038	2116
Fruit, servings/1000 kcal	2.3	1.0	1.6	1.4	1.1
Vegetables, servings/1000 kcal	2.4	2.1	2.0	2.0	1.9
Alcohol, g/d	20.2	20.4	17.6	15.3	12.5
Total fat, g/1000 kcal	25.8	30.5	33.5	35.9	39.4
Saturated fat, g/1000 kcal	7.6	9.4	10.5	11.3	12.7
Fiber, g/1000 kcal	13.2	11.0	10.2	9.6	8.8
Vitamin supplement use ≥1/mo	67.3	62.1	59.1	55.8	52.0

*Reproduced with permission from Sinha R, Cross AJ, Graubard BI, Leitzmann MF, Schatzkin A. Meat intake and mortality: a prospective study of over half a million people. Arch Intern Med 2009;169:562-571 [1].

†Body mass index, smoking, calories, saturated fat, and fat increased with increasing red meat consumption, whereas education, physical activity, fruits and vegetables, fiber, alcohol, and vitamin supplementation decreased.

The equation gives the expected blood pressure for any combination of salt intake, age, and smoking status. Thus, by holding age and smoking constant, one can isolate the effect of salt. For example, consider two 55-year-old smokers who eat different amounts of salt per day (3 versus 2 teaspoons). Their predicted blood pressures are:

$$\text{Predicted systolic blood pressure} = 60 + 5 * 3 \text{ tsp/d} + 1 * 55 \text{ years} + 10 * 1 = 140 \text{ mm Hg}$$

$$\text{Predicted systolic blood pressure} = 60 + 5 * 2 \text{ tsp/d} + 1 * 55 \text{ years} + 10 * 1 = 135 \text{ mm Hg}$$

By comparing these 2 equations, the independent effect of salt becomes apparent: on average, one additional teaspoon of daily salt translates to 5 additional mm Hg of blood pressure. The same result arises no matter what values are used for age and smoking, as long as age and smoking are held constant.

So, the coefficient for salt represents the effect of salt on blood pressure at any fixed level of age and smoking. Note that this model assumes that (1) salt has the same effect at all age and smoking levels; in other words, that there are no interactions between salt and age or between salt and smoking; and (2) the relationship between age (a continuous confounder) and blood pressure is linear, which means that it is the same across all ages. Violations of these assumptions can lead to imperfect control of confounding, as described in the next section (unless interaction or nonlinear terms are included).

The exact choice of regression model depends on the outcome variable (see Table 2 for details about the most commonly used models). Returning to our meat and mortality example, the outcome was time to death (mortality), so the researchers appropriately used Cox regression. Cox regression yields hazard ratios (a type of relative risk), whereas linear regression yields slopes. However, the principles of statistical adjustment are the same regardless of the type of regression. The outcome sides (right sides) of the regression

Table 2. *Common multivariate regression models*

Outcome (dependent variable)	Example Outcome Variable	Appropriate Multivariate Regression Model*	Example Equation†	What Do the Coefficients Give You?
Continuous	Blood pressure	Linear regression	Blood pressure (mm Hg) = α + β_{salt}*salt consumption (tsp/d) + β_{age}*age (y) + β_{smoker}*ever smoker (yes = 1/no = 0)	Slopes: how much the outcome variable increases for every 1-unit increase in each predictor
Binary	High blood pressure (yes/no)	Logistic regression	ln (odds of high blood pressure) = α + β_{salt}*salt consumption (tsp/d) + β_{age}*age (y) + β_{smoker}*ever smoker (yes = 1/no = 0)	Odds ratios: how much the odds of the outcome increase for every 1-unit increase in each predictor
Time to event	Time to death	Cox regression‡	ln (rate of death) = α + β_{salt}*salt consumption (tsp/d) + β_{age}*age (y) + β_{smoker}*ever smoker (yes = 1/no = 0)	Hazard ratios: how much the rate of the outcome increases for every 1-unit increase in each predictor

*The choice of appropriate model depends on the type of outcome (dependent) variable.
†Regression equations include an intercept (α) and coefficients (β); the coefficients summarize the relationship between each predictor and the outcome.
‡ln = natural log; note that no intercept is actually estimated in Cox regression.

equations differ, but the predictor sides (left sides) are the same (Table 2).

In the meat and mortality study, the researchers reported 2 models (Table 3): a basic model that included meat intake plus age, race, and total energy intake; and a "fully adjusted" model that, in addition, included total energy intake, education, marital status, family history of cancer, body mass index, smoking history, physical activity, alcohol intake, vitamin supplementation, fruit consumption, and vegetable consumption. With the basic model, the hazard ratio for total mortality for the top quintile of red meat eaters compared with the lowest quintile was 1.48 ($P < .001$), which indicates a statistically significant 48% increase in the mortality rate. This effect was attenuated but not eliminated by further adjustment: the hazard ratio was 1.31 ($P < .001$). The researchers concluded that red meat has an independent effect on mortality, and the study received widespread media coverage, with recommendations that people curtail their red meat consumption.

THE LIMITATIONS OF STATISTICAL ADJUSTMENT

A closer look at the meat and mortality study reveals some interesting findings. For example, heavy meat eaters had a significantly elevated risk of dying from injuries and sudden deaths (eg, car crashes, guns): the hazard ratio is 1.26, which is similar to the hazard ratios for cancer mortality (1.22) and heart disease mortality (1.27). This is troubling because, although there are plausible biological stories to connect red meat with cancer and heart disease, it seems unlikely that eating too much red meat could directly cause accidents and injuries. (Unless, as one of my students quipped, red meat eaters are swerving to avoid cows!) Red meat eaters also had an elevated risk (hazard ratio, 1.58) of death from "all other causes," including infections, liver disease, Alzheimer disease, and diabetes-related complications. Although some of these disorders can plausibly be linked to red meat consumption (eg, Alzheimer disease), others cannot (eg, infectious diseases). As noted by several letters to the editor [2,3], these "control" outcomes suggest the presence of residual confounding. Heavy red meat eaters are inherently different (and more risk prone) than low red meat eaters, and, no matter how hard one tries (indeed, the researchers made a valiant effort, including a 31-level smoking variable!), it is impossible to completely pin these differences down and account for them. Thus, some residual "riskiness" remains unaccounted for, which leads to spurious associations. Because the hazard ratios for cancer and heart disease mortality do not exceed those for injuries and "other deaths," it is likely that the former effects also are entirely spurious. So, although the study received much hype, it does not provide convincing evidence of a causal relationship between consuming red meat and death.

Residual Confounding

The concept of residual confounding takes a lot of people by surprise. Many people are under the impression that statistical adjustment is a panacea, that once a particular confounding variable has been included in the model, it is taken care of completely. Unfortunately, statistical adjustment can only work perfectly when all the variables in the model are measured perfectly. In reality, most variables are measured with considerable error. Take the hypothetical salt example. Smoking was measured as ever smoker versus never smoker. Obviously, this ignores substantial variability among smokers. If the heaviest smokers eat more salt and have higher

Table 3. *Multivariate analysis that examined the relationship between red meat and total and cause-specific mortality**

Mortality in men (n = 322 263)	Quintile†					*P* Value for Trend
	Q1	Q2	Q3	Q4	Q5	
Red meat intake‡						
All mortality						
Deaths	6437	7835	9366	10,988	13,350	
Basic model§	1 (Reference)	1.07 (1.03-1.10)	1.17 (1.13-1.21)	1.27 (1.23-1.31)	1.48 (1.43-1.52)	<.001
Adjusted model¶	1 (Reference)	1.06 (1.03-1.10)	1.14 (1.10-1.18)	1.21 (1.17-1.25)	1.31 (1.27-1.35)	<.001
Cancer mortality						
Deaths	2136	2701	3309	3839	4448	
Basic model§	1 (Reference)	1.10 (1.04-1.17)	1.23 (1.16-1.29)	1.31 (1.24-1.39)	1.44 (1.37-1.52)	<.001
Adjusted model¶	1 (Reference)	1.05 (0.99-1.11)	1.13 (1.07-1.20)	1.18 (1.12-1.25)	1.22 (1.16-1.29)	<.001
Cardiovascular disease mortality						
Deaths	1997	2304	2703	3256	3961	
Basic model§	1 (Reference)	1.02 (0.96-1.08)	1.10 (1.04-1.17)	1.24 (1.17-1.31)	1.44 (1.37-1.52)	<.001
Adjusted model¶	1 (Reference)	0.99 (0.96-1.09)	1.08 (1.02-1.15)	1.18 (1.12-1.26)	1.27 (1.20-1.35)	<.001
Mortality from injuries and sudden deaths						
Deaths	184	216	228	280	343	
Basic model§	1 (Reference)	1.02 (0.84-1.24)	0.97 (0.80-1.18)	1.09 (0.90-1.31)	1.24 (1.03-1.49)	.01
Adjusted model¶	1 (Reference)	1.06 (0.86-1.29)	1.01 (0.83-1.24)	1.14 (0.94-1.39)	1.26 (1.04-1.54)	.008
All other deaths						
Deaths	1268	1636	1971	2239	2962	
Basic model§	1 (Reference)	1.13 (1.05-1.22)	1.25 (1.17-1.35)	1.33 (1.24-1.42)	1.68 (1.57-1.80)	<.001
Adjusted model¶	1 (Reference)	1.17 (1.09-1.26)	1.28 (1.19-1.38)	1.34 (1.25-1.44)	1.58 (1.47-1.70)	<.001

*Reproduced with permission from Sinha R, Cross AJ, Graubard BI, Leitzmann MF, Schatzkin A. Meat intake and mortality: a prospective study of over half a million people. Arch Intern Med 2009;169:562-571 [1].

†The hazard ratios give the increase in mortality rate for each quintile of red meat intake compared with the lowest quintile (reference group). Data are given as hazard ratio (95% confidence interval) unless otherwise specified.

‡Median red meat intake based on men and women (g/1000 kcal): Q1, 9.8; Q2, 21.4; Q3, 31.3; Q4, 42.8; and Q5, 62.5.

§Basic model includes age (continuous), race (non-Hispanic white, non-Hispanic black, Hispanic/Asian/Pacific Islander/American Indian/Alaskan native, or unknown), and total energy intake (continuous).

¶Adjusted model: basic model plus education (<8 y or unknown, 8-11 y, 12 y [high school], some college, or college graduate), marital status (married: yes/no), family history of cancer (yes/no) (cancer mortality only), body mass index (18.5 to <25, 25 to <30, 30 to 35, ≥35 [calculated as weight in kg/m^2]), 31-level smoking history by using smoking status (never, former, current), time since quitting for former smokers and smoking dose, frequency of vigorous physical activity (never/rarely, 1-3 times/mo, 1-2 times/wk, 3-4 times/wk, ≥5 times/wk), alcohol intake (none, 0 to <5, 5 to <15, 15 to <30, ≥30 servings/1000 kcal), vitamin supplement user (≥1 supplement/mo), fruit consumption (1 to <0.7, 0.7 to <1.2, 1.2 to <1.7, 1.7 to <2.5, ≥2.5 servings/1000 kcal); and vegetable consumption (0 to <1.3, 1.3 to <1.8, 1.8 to <2.2, 2.2 to <3.0, ≥3.0 serving/1000 kcal).

blood pressures than the lightest smokers, then this will artificially inflate the relationship between salt and blood pressure. Adjusting for smoking as a binary variable (smoker-nonsmoker) will do nothing to remove this confounding. Smoking history also may be misreported, which further obscures the extent of confounding making it impossible to completely account for it.

There are many other potential sources of residual confounding as well. The model may ignore complex relationships; for example, it may not account for interactions or for nonlinear relationships between the confounder and the outcome. Plus, multivariate models can only accommodate a limited number of predictor variables (one predictor per 10 observations, according to one rule of thumb), so some confounders may have to be omitted. Also, modeling cannot account for confounders that are unknown or unmeasured.

Fortunately, there are some limits on how much damage residual confounding can actually do. Simulations and real examples [4-6] suggest that residual confounding can explain away small-to-moderate effects but not large effect sizes. For example, for a binary predictor, incomplete control of one or two strong confounders can plausibly generate spurious relative risks in the range of 1.0-1.6 (at most, a 60% increase in odds or risk) [4,5]. (For protective effects, this would roughly be equivalent to odds ratios or hazard ratios in the range 0.6-1.0). Although complete omission of strong confounders can create much larger spurious associations [4], it is uncommon that strong confounders (which are usually obvious) are missed. In addition to creating spurious associations, residual confounding also can obscure relationships, which leads researchers to miss associations (although these examples are harder to recognize).

Note that the effect sizes from the meat and mortality study are exactly in the range that is consistent with residual confounding (hazard ratios of 1.22-1.58, for the cause-specific mortalities). Thus, red meat eating may indeed be a marker of unhealthy lifestyle, but the study

provides little evidence that someone with otherwise healthy habits who eats a lot of red meat will increase his or her mortality risk.

CONCLUSIONS

Statistical adjustment is a powerful tool that can help isolate the effect of one risk factor out of a cluster of risk factors. However, it is a myth that it is possible to statistically adjust away all confounding. Readers and editors should be cautious in interpreting results from observational studies when the observed effects are small to moderate in size (odds ratios or hazard ratios between 0.6 and 1.6 for a binary predictor); if there are strong confounders at play, the observed effects could be entirely due to residual confounding. Inclusion of a control outcome, one plausibly related to the confounders but not to the risk factor of interest, can help readers evaluate the potential impact of residual confounding.

REFERENCES

1. Sinha R, Cross AJ, Graubard BI, Leitzmann MF, Schatzkin A. Meat intake and mortality: a prospective study of over half a million people. Arch Intern Med 2009;169:562-571.
2. de Abreu Silva EO, Marcadenti A. Higher red meat intake may be a marker of risk, not a risk factor itself. Arch Intern Med 2009;169:1538-1539.
3. Mozaffarian D. Meat intake and mortality: evidence for harm, no effect, or benefit? Arch Intern Med 2009;169:1537-1538.
4. Fewell Z, Smith GD, Sterne JAC. The impact of residual and unmeasured confounding in epidemiologic studies: a simulation study. Am J Epidemiol 2007;166:646-655.
5. Psaty BM, Koepsell TD, Lin D, et al. Assessment and control for confounding by indication in observational studies. J Am Geriatr Soc 1999; 47:749-754.
6. Chen C-L, Gilbert TJ, Daling JR. Maternal smoking and down syndrome: the confounding effect of maternal age. Am J Epidemiol 1999;149:442-446.