THE FIRST TWO AUTHORS REPLY

We are grateful for Dr. Cologne’s comments (1) on our paper (2). We agree that the implications of “baseline” adjustment depend on the timing of measurement of the baseline and its possible causal relations with other variables. In our paper on identifying determinants of change in a health outcome, we used the term “baseline” to refer to the first available measure of the health outcome of interest. In some studies, the earliest measure of the outcome construct preceded the treatment or exposure and, thus, provides a pretreatment baseline. Often, however, the earliest available measure occurred after the treatment; this was the primary case we considered in our paper. In typical randomized trials, baseline measures precede randomization, and these situations illustrate the key points. Below, we present two causal directed acyclic graphs (DAGs) to illustrate the harmlessness of baseline adjustment in a typical trial and the potential bias induced by adjustment for even a pretreatment baseline measure in the absence of randomization.

Consider a trial to test the effect of a memory-training intervention on cognitive change. The DAG in figure 1 shows no arrows pointing into “Randomization to memory training,” because randomization is not influenced by any other variables in the DAG. We are interested in identifying the effect of randomization on change in the cognitive score (represented as a dotted arrow in this DAG). Under the assumptions in this DAG, the statistical association between randomization and change score equals the causal association, regardless of whether the analysis is conditioned on “Cognitive score 1995.”

To illustrate how this differs from a typical observational analysis, suppose instead that we were interested in estimating the effect of education on the change score. Under the assumptions of the DAG in figure 1, the unadjusted statistical association between education and the change score is an unbiased estimate of the causal effect. However, conditioning on “Cognitive score 1995” will introduce spurious statistical associations among its causes. In this DAG, “Cognitive score 1995” is assumed to be affected by education, a gene that causes fast cognitive decline (both via “Cognitive function 1995”), and “Measurement error 1995.” In a regression of change score on education and score 1995, the coefficient for education will be biased away from the causal effect. In fact, there are two (potentially offsetting) sources of bias in the baseline-adjusted analysis, because education may be spuriously negatively associated with “Measurement error 1995” and spuriously positively associated with the gene for cognitive decline.

There are two crucial features of trials that generally render baseline adjustment harmless: 1) the treatment of interest is randomized, and 2) the analysis is not conditioned on postrandomization variables. Baseline adjustment may induce biases if either of these features is absent. Feature 2 is important because, if randomization precedes baseline, randomization may influence the baseline (i.e., there may be an arrow in the DAG above from randomization to “Cognitive function 1995”) and, if so, baseline adjustment could then induce a bias. Feature 1 is important because, without randomization, adjustment for even a pretreatment baseline measure may induce bias. For example, suppose we wish to study the effect of a “Memory training” on cognitive change in an observational study (i.e., when participation in memory training is not randomized). The DAG in figure 2 represents the assumption that the baseline “Cognitive function 1995” affects participation in “Memory training.” Under these assumptions, adjustment for the baseline “Cognitive score 1995” can induce a spurious statistical association between “Memory training” and change score.

If, contrary to the assumptions in figure 2, “Cognitive function 1995” directly influenced cognitive change, the bias introduced by baseline adjustment might be smaller than the bias from failing to adjust for a confounder. Simulation studies may be helpful for quantifying countervailing biases under a range of assumptions. They may also be useful for understanding the issue of statistical power raised in
the letter (1). For an example of STATA code (StataCorp LP, College Station, Texas) that can be used to generate data following the causal structure of the DAGs presented in our paper (2), please refer to the Web appendix. (This information is posted on the Journal’s website (http://aje.oxfordjournals.org/).

In nonexperimental research in which the outcome of interest is change, statistical adjustment for a baseline measure may bias effect estimates away from the causal parameters. When analyses of randomized trial data include only prerandomization covariates, baseline adjustment does not introduce a bias and often improves statistical power. Stratification on a prerandomization baseline value may also be of substantive interest, if heterogeneous treatment effects seem plausible.

ACKNOWLEDGMENTS
Conflict of interest: none declared.

REFERENCES

M. Maria Glymour1,2 and Jennifer Weuve3 (e-mail: mglymour@hsph.harvard.edu)
1 Department of Society, Human Development, and Health, Harvard School of Public Health, Boston, MA 02215
2 Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, NY 10032
3 Department of Environmental Health, Harvard School of Public Health, Boston, MA 02215

DOI: 10.1093/aje/kwj360; Advance Access publication October 16, 2006