Bayesian Analysis (2007) **2**, Number 1, pp. 31–36

Comment on Article by Dominici et al.

Samantha R. Cook* and Elizabeth A. Stuart†

1 Introduction

Causal inference, and specifically principal stratification (Frangakis and Rubin 2002), is an important area of statistics with the potential to answer fundamental questions in medicine, economics, and many other scientific disciplines. Dominici et al. have done a commendable job of applying and extending the ideas of principal stratification to address specific questions about the impact of vitamin supplementation on birth weight and infant mortality. We applaud and congratulate the authors on an insightful and important paper. After briefly reviewing the paper, we focus on three issues: the assumptions required for principal stratification applications such as this one, the particular causal quantities considered here, and possible model extensions to handle observational data or more complex outcomes.

This paper considers the effect of vitamin supplementation on infant mortality; analysis is aided by the fact that treatments were assigned randomly and most study participants complied with their assigned treatment. Inference was complicated, however, because the treatment effect was believed to be non-constant. Estimating causal effects conditional on covariates is straightforward; in this case, however, the causal effects of vitamin supplementation on mortality were believed to vary with birth weight, which is itself an outcome that may depend on the treatment received. We therefore have a primary outcome, infant mortality, that may be related to an intermediate outcome, birth weight. The authors focus on two quantities of interest: the percentile-specific effects of supplementation on birth weight and the effects of supplementation on infant mortality, principally stratified by birth weight.

2 Principal Stratification and Associated Assumptions

A causal effect is fundamentally a comparison of two *potential outcomes*: the outcome a single individual would experience if assigned to take the treatment and the outcome the individual would experience if assigned to control (Rubin 1977). Because we can observe at most one potential outcome on each unit, causal inference is inherently a missing data problem (Holland 1986). Inference tends to focus on average treatment effects or average treatment effects within subgroups rather than individual causal effects, which are never observed. When subgroups are defined by pre-treatment covariates and treatments are assigned randomly, causal inference remains relatively straightforward. Estimating treatment effects conditional on *post-treatment* variables is more complicated, however, because individuals with similar values of post-treatment variables are not necessarily

*Columbia University, New York, NY, mailto:cook@stat.columbia.edu
†Mathematica Policy Research, Inc, Princeton, NJ

 ba0005