THE EFFECT OF MISSING DATA IN THE ANALYSIS OF A BARIATRIC SURGERY PROGRAM

Katharine Fischer Berry

As obesity rates in the United States continue to rise, numerous programs across the nation have been created in order to educate the population about, and reduce the prevalence of, obesity and cardiovascular disease. The purpose of the first part of my research is to investigate the effects of a particular bariatric program at North Shore Medical Center (Salem, MA). The bariatric program involves bariatric surgery coupled with a 13-week long program on cardiovascular risk reduction. I will look at two-year weight loss, reduction in cholesterol, and reduction in blood pressure as well as various categorical variables associated with obesity.

Although 397 patients underwent the procedure before September of 2004, only 191 patients have returned for their two year follow-up. Are patients who returned systematically different in some way than patients who did not return? Conclusions made based on available-case analysis are only valid if the missing data are missing at random—that is, observed values are a random subsample of the complete dataset.\(^1\) When data are missing at random, the observed distributions match the underlying distributions, and the missing data mechanism is ignorable.\(^2\) When the data are not missing at random, available-case analysis underestimates and/or overestimates effects, and thus the missing data mechanism must be taken into account in order to reach valid conclusions.

Unfortunately, we cannot know the missing data mechanism for the bariatric surgery dataset and so we cannot assume that the data are missing at random. The second part of my investigation is therefore dedicated to exploring the missing data mechanism, imputing values based on hierarchical loglinear models, and generating new, improved estimates.

---


\(^2\) Ibid, 15.