Selective Survival: Consequences and Solutions in Demographic Research

Scott M. Lynch, Princeton University

Abstract

Despite its ubiquity, selective survival receives little attention in contemporary research outside of the areas of mortality and health demography. Yet, selective survival may affect substantive conclusions in a much broader array of substantive areas. Often, in analyses using panel data, missingness due to mortality is handled using missing data methods that simply do not resolve the biases introduced by mortality. In analyses using cross-sectional data, selective survival is generally not even considered, despite the fact that it affects the pool of potential respondents. In this research, I (1) discuss when selective survival may be a problem and when it can be safely ignored; (2) use simulated data to show the consequences of ignoring selective survival, in terms of biased estimates of means and regression coefficients; (3) show how popular missing data techniques fail to correct for selective survival; and (4) present methods that can partially correct for selective survival.

Extended Abstract

Selective survival affects the pool of potential respondents in both repeated cross-sectional and panel designs, and it has the potential to influence findings in a number of substantive areas. Despite its ubiquity, selective survival receives very little attention in contemporary research. To be sure, mortality demographers began discussing the issue in the 1970s (e.g., Vaupel, Manton, and Yashin 1979), in terms of its ability to influence aggregate mortality curves. Their research presaged a broader interest in the influence unobserved heterogeneity in demography, but in this contemporary literature, selective survival is often not discussed as a source of heterogeneity.

More recently, some health literature has begun to consider the influence of selective survival on substantive conclusions. Most notably, the literature on life course trajectories of socioeconomic status differences in health has considered the influence of selective survival (e.g., see Beckett 2000; Lynch 2003). In that literature, it is well known that the gap in health between high and low socioeconomic status groups narrows beyond midlife. Although the jury is still out regarding whether the entirety of the narrowing of the gap is due to selective survival, the research has shown that selective survival certainly downwardly biases estimates of the gap.

In other substantive areas of research, the effect of selective survival is almost never considered. Yet, it may lead to considerable bias if it is ignored, if (1) the research question concerns a life course process, and (2) the outcome of interest is related to mortality. For example, consider interest in stratification research in wealth accumulation across the life course. Wealth accumulation is, of course, a life course process, and is no doubt related to health and mortality. If wealth predicts mortality, with those with less wealth dying at a faster rate than those with greater wealth, then the slopes

of trajectories of wealth will be biased upward, and the estimated gap in wealth between those with higher vs. lower levels of wealth early in life will be downwardly biased.

This research addresses four issues: (1) I discuss when selective survival may be a problem and when it can be safely ignored in research. (2) I use simulated data to show the consequences of ignoring selective survival, in terms of biased estimates of means and regression coefficients. (3) I show how most popular missing data techniques fail to correct for selective survival appropriately. (4) Finally, I present a few methods that can partially correct for selective survival.