Mortality Selection and Sample Selection: A Comment on Beckett*

ANDREW NOYMER

University of California—Berkeley

Journal of Health and Social Behavior 2001, Vol 42 (Septmber): 326–327

In an interesting article, Megan Beckett (2000) examines the important question of converging health inequalities in later life. Many studies have shown that the differences in health across socioeconomic strata narrow at older ages. Using panel data from the National Health and Nutrition Examination Survey (NHANES), Beckett shows that the converging health inequality cannot be accounted for by mortality selection. The present comment reconsiders Beckett's approach to the selection problem, which, while creative, is open to multiple interpretations.

Consider the phenomenon that would cause converging health inequalities at later ages to be an "artifact," as Beckett puts it, of mortality selection. At younger ages, persons with higher socioeconomic status (SES) have lower levels of health problems than those with lower SES. At older ages, the prevalence of health problems in the two groups is closer to parity. If patterns in morbidity are mirrored in mortality, then at older ages a lower SES cohort (higher morbidity and mortality) will be smaller compared to its starting size than a higher SES cohort (lower morbidity and mortality). Since the seminal work of Vaupel, Manton, and Stallard (1979) and Keyfitz and Littman (1979), many demographers have assumed that there are different rates of "frailty" within a population, which determine an individual's deviation, at any age, from some baseline mortality risk. According to the frailty hypothesis, those who die at young ages tend to have high frailty, which skews the distribution of survivors to be more robust. If this condition of nonrandom mortality risks is met, then the aged low SES cohort will be more robust than when it started out. This reversal of fortune over the life course is called "cohort inversion" (Hobcraft, Menken, and Preston 1982). On the other hand, the low mortality, high SES cohort will have a much less-changed frailty distribution, and will experience less cohort inversion. The greater cohort inversion of the low SES cohort could be enough to overcome the mortality disadvantage of being low SES. This problem must be analyzed cautiously, however, as the entire framework for understanding mortality selection effects rests on a counterfactual foundation. If we hold that convergence is a result of mortality selection, we imply that an intrinsic differential persists into older ages and that we would observe it were it not for the selection effect. On the other hand, if we hold that the convergence is either intrinsic or the result of, for example, access to Medicare (Beckett 2000), we posit that even without selection we would see convergence. In both cases, there is the troubling verb "would." In reality, we can only observe vital rates that do occur, not those that would occur if some condition is met.

The general problem of sample selection is encountered frequently in the social sciences (cf. Stolzenberg and Relles 1997; Winship and Mare 1992), and as Beckett (2000) notes, differential mortality is just a special case of the more general problem. Although we cannot simply "control for" (i.e., condition on) selection bias the same way we would a confounding variable, statistical techniques do exist that try to counteract the bias. However, mortality selection is a *very* special case of sample selection, all the more so if the dependent variable in question is itself health-related. Because of cohort inversion, sample selection due to mortality has causal implications beyond nonrandom missing data in panel followups. This is what makes Beckett's approach problematic.

Consider the statistical technique used by Beckett to set up the hypothetical of no mor-

*Address correspondence to: Andrew Noymer, Departments of Sociology and Demography, University of California—Berkeley, 2232 Piedmont Avenue, Berkeley, CA 94720; email: andrew@demog.berkeley.edu.

tality selection. She estimates two successive ordered logistic regression equations first using the NHANES sample as it exists, then including those who are deceased, using "imputed 1992 health outcomes" to resurrect those lost from the sample through mortality. Comparing the results of the two regression tables, if the health convergence persists even after the selection has been corrected, then we may assert that selection does not cause the convergence in the first place. A similar approach is used in a third regression, additionally including those lost to follow-up and not known to be dead, but let us focus on the question of mortality selection.

Little information is given by Beckett on exactly how the 1992 outcomes are imputed. Presumably, NHANES data were used (along with post-regression prediction commands that have become standard in computing packages), combined with the modified pseudovariables approach. Since those already deceased don't have health problems that can be measured directly, imputation relies on data from those who survived, plus the notion that those who died would, had they survived, be less healthy than average. Imputing missing data about those who died using data from those who lived biases downward the imputed prevalence rates, and this bias makes the reconstituted sample look more like the survivors. This can lead to the spurious conclusion that in the hypothetical absence of mortality selection, we would still see the convergence. That is, we would conclude that the health convergence is not an "artifact" of mortality selection, since the convergence persists even after selection is (statistically) removed.

There is a well-known thought experiment in demography, whereby we assume that everyone is saved from death once and thereafter is subject to prevailing age-specific mortality rates (Keyfitz 1985). To continue the thought experiment, if we are to test the mortality selection hypothesis in its own terms, those who are saved from death would still have a high frailty, and would be subject to high mortality rates the "second time around." To consider them living on *ad infinitum,* assigning a low health status using pseudovariables, is not a prudent exercise because as social scientists we are, ultimately, interested in the true population at risk. Overaggressive correction of sample selection bias due to mortality leads to a distorted model population, the consequences of which are serious if the dependent variable is itself also health status. To sum up Beckett's result, convergence in health inequalities persist even when we pretend that those who died lived on. Given the way the counterfactual was set up, and the fact that any real population experiences deaths (which are, emphatically, not "artifacts" of study design), it is hardly a surprising finding.

Hazard rate models have been developed that can better address this issue (see Manton, Singer, and Woodbury 1992), but the NHANES data under consideration are not amenable to that method. Beckett's creative alternative approach is laudable, but relies on many strong assumptions, some of which are not made explicit. It may be that the magnitude of the cohort inversion caused by mortality selection is not great enough to explain the observed convergence in health inequality at older ages. However, Beckett's conclusion that mortality selection can be ruled out is premature.

REFERENCES

Beckett, Megan. 2000. "Converging Health Inequalities in Later Life—an Artifact of Mortality Selection?" *Journal of Health and Social Behavior* 41(1):106–19.

Hobcraft, John, Jane Menken, and Samuel Preston. 1982. "Age, Period, and Cohort Effects in Demography: A Review." *Population Index.* 48(1):4–43.

Keyfitz, Nathan. 1985. *Applied Mathematical Demography.* 2nd ed.. New York: Springer.

Keyfitz, N. and G. Littman. 1979. "Mortality in a Heterogeneous Population." *Population Studies.* 33(2):333–42.

Manton, Kenneth G., Burton Singer, and Max A. Woodbury. 1992. "Some Issues in the Quantitative Characterization of Heterogeneous Populations." Pp. 9–37 in *Demographic Applications of Event History Analysis,* edited by J. Trussell, R. Hankinson, and J. Tilton. Oxford: Clarendon Press/Oxford University Press.

Stolzenberg, Ross M. and Daniel Relles. 1997. "Tools for Intuition about Sample Selection Bias and its Correction." *American Sociological Review* 62(3):494–507.

Vaupel, James W., Kenneth G. Manton, and Eric Stallard. 1979. "The Impact of Heterogeneity in Individual Frailty on the Dynamics of Mortality." *Demography.* 16(3):439–54.

Winship, Christopher, and Robert D. Mare. 1992. "Models for Sample Selection Bias." *Annual Review of Sociology.* 18:327–350.