The Osteological Paradox

Problems of Inferring Prehistoric Health from Skeletal Samples


Paleodemography and paleopathology presuppose that direct relationships exist between statistics calculated from archaeological skeletal series (e.g., skeletal lesion frequencies and mean age at death) and the health status of the past populations that gave rise to the series. However, three fundamental conceptual problems confound the interpretation of such statistics: demographic non-stationarity, selective mortality, and unmeasured, individual-level heterogeneity in the risks of disease and death. Using simple sample models of the relationship between individual “ frailty” and the hazard of death at each age, this paper explores the implications of these problems for archaeological interpretation. One conclusion is that the skeletal evidence pertaining to the transition from hunting-and-gathering to settled agriculture is equally consistent with an improvement in health and a deterioration in health resulting from the transition.


GEORGE R. MILNER is Associate Professor and Museum Curator in the Department of Anthropology at Pennsylvania State University. He was born in 1953 and received his B.A. from Beloit College in 1975 and his Ph.D. from Northwestern University in 1982. He has been curator of the Museum of Anthropology of the University of Kentucky and has conducted archaeological fieldwork in Illinois. He has published “Social and Temporal Implications of Variation among American Bottom Mississippian Cemeteries” [American Antiquity 49:468–88], “The Late Prehistoric Cahokia Cultural System of the Mississippi River Valley: Foundations, Florecence, and Fragmentation” [Journal of World Prehistory 4:1–43], and (with D. H. Humph and H. C. Harpending) “Pattern Matching of Age-At-Death Distributions in Paleodemographic Analysis” [American Journal of Physical Anthropology 80:49–58].

HENRY C. HARPENDING is Professor in the Department of Anthropology and the Graduate Program in Genetics at Pennsylvania State University. He was born in 1944 and educated at Hamilton College (B.A., 1964) and Harvard University (Ph.D., 1972). He taught at the University of New Mexico from 1973 to 1985. Among his publications are (with L. Sattenspiel) “Stable Populations and Skeletal Age” [American Antiquity 48:489–98], (with A. Rogers and P. Draper) “Human Sociobiology” [Yearbook of Physical Anthropology 30:127–50], (with R. Pennington) “Age Structure and Sex-biased Mortality among Herero Pastoralists” [Human Biology 63:337–52], and (with A. Rogers) “Population Growth Makes Waves in the Distribution of Pairwise Nucleotide Differences” [Molecular Biology and Evolution, in press].


The present paper was submitted in final form 28 XI 91.

Measurement and interpretation of differences in the level of health in prehistoric populations are two fundamental goals of paleodemography and paleopathology. Studies of such differences are vital to our understanding of how the adaptive success of human populations has varied through time and space. For the purposes of these studies, several standard indices of morbidity and mortality derived from skeletal series (e.g., skeletal lesion frequencies, life expectancies, and average age at death) would seem readily interpretable. Common sense suggests that there should be some reasonably direct association between these aggregate-level measures and the risks of illness and death experienced by the individual members of past populations. For example, an increase in the frequency of a particular type of skeletal lesion might be interpreted as indicating an elevation in the individual risks of experiencing the conditions that induce the lesion and hence an increase in the population prevalence of those conditions. Similarly, a decrease in the average age at death might be taken as signaling an elevation in the risks of death experienced by the individuals in the once-living population. Most interpretations of paleodemographic and paleopathological data presuppose that such straightforward relationships exist.

1. A version of this paper was presented at the 1989 meetings of the Society for American Archaeology in Atlanta, Ga. We thank Eve Anderson, Gillian Bentley, Anne Buchanan, Timothy Gage, Darryl Holman, Patricia Johnson, Paul Leslie, Claire McHale Milner, and Richard Paine for comments and suggestions. We also thank the Illinois State Museum for permission to examine the Oneota skeletal collection.
We suggest that interpretation of these standard measures may be more difficult than it first appears, especially when dealing with skeletons from archaeological contexts. Failure to address three important conceptual issues—demographic nonstationarity, selective mortality, and hidden heterogeneity in risks—can render inferences based on various demographic and epidemiological measures meaningless. These problems are, we believe, among the most difficult currently facing reconstructions of the demography and health of past populations, far more difficult than the issues raised by Bocquet-Appel and Masset [1982] in their "Farewell to Paleodemography" [indeed, their points have largely been addressed by Van Gerven and Armelagos [1983] and Buikstra and Konigsberg [1985], among others].

This paper is not, however, a counsel of despair. Continual reexamination of underlying premises is essential to any vigorous field of research. In the case of paleopathology, studies of ancient bone have undergone significant changes in orientation and sophistication since the mid-1960s [Armelagos, Goodman, and Jacobs 1978, Buikstra and Cook 1980, Huss-Ashmore, Goodman, and Armelagos 1982, Ubelaker 1982, Martin, Goodman, and Armelagos 1985, Larsen 1987, Ortner and Aufferheide 1991]. During this period, the field has moved from a particularistic concern with individual lesions or skeletons to a population-based perspective on disease processes, the evolution of humans and their pathogens, and the evaluation of past community health, especially as it has changed with the adoption of fundamentally new ways of life. Similarly, paleodemography has moved from simple and often idiosyncratic tabulations of age and sex data to more rigorous analytical investigations of demographic processes in prehistoric populations [Angel 1969, Acsádi and Neméskeri 1970, Weiss 1973, Ubelaker 1974, Moore, Swedlund, and Armelagos 1975, Buikstra 1976, Lovejoy et al. 1977, Lalco, Rose, and Armelagos 1980, Buikstra and Konigsberg 1985, Buikstra and Mielke 1985, Buikstra, Konigsberg, and Bullington 1986, Johannson and Horowitz 1986, Horowitz, Armelagos, and Wachter 1988]. The concerns raised in this paper would not be relevant had paleopathology and paleodemography not become established fields of active research.

Conceptual Problems

Demographic nonstationarity refers to the departure of a population from the stationary state—a state characterized by closure to migration, constant age-specific fertility and mortality, zero growth rate, and an equilibrium age distribution. In nonstationary populations, age-at-death distributions are extremely sensitive to changes in fertility but not to changes in mortality [Coale 1957, Keyfitz 1985]. Thus, if a population is not stationary—and changing populations never are—small variations in fertility have large effects on its age-at-death distribution, while even quite large modifications of mortality have virtually none. Paradoxically, then, summary statistics such as life expectancies or the mean age at death are often effectively measures of fertility rather than mortality.

The implications of this basic fact of mathematical demography for prehistoric skeletal samples have been discussed elsewhere [Sattenspiel and Harpending 1983, Johannson and Horowitz 1986, Milner, Humph, and Harpending 1989, Paine 1989]. Paleodemographers have started to pay some attention to this problem, although they still tend to underrate its importance [Cohen 1989]. Because the issue of demographic nonstationarity is not new, we concentrate in this paper on the other two problems, selective mortality and hidden heterogeneity in risks. This treatment is convenient because those two problems are fundamentally related: in the absence of heterogeneity, selectivity does not occur.

The selective mortality problem is easy to state but difficult to solve. There is one, and perhaps only one, irrefutable fact about the cases making up a skeletal series: they are dead. We never have a sample of all the individuals who were at risk of disease or death at a given age, but only of those who did in fact die at that age. For example, the only 20-year-olds we observe in a skeletal sample are those who died at age 20. While we may find the skeletons of many of the other individuals who had been at risk of death at age 20 but who died later, say, at age 60, we observe their characteristics as 60-year-olds, not 20-year-olds. As a result, the sample of 20-year-olds [or any other age-group] is highly selective for lesions that increase the risk of death at that age. Estimates of the population prevalences of such lesions from skeletal series are therefore subject to precisely the same sort of selectivity bias as the derivation of population prevalences from clinical data, because neither type of data constitutes a representative sample of the entire population at risk. In both cases, the observed frequency of pathological conditions should overestimate the true prevalence of the conditions in the general population. This bias cannot be avoided simply by obtaining larger, more representative skeletal samples; it is built into the very structure of the data.

The problem of hidden heterogeneity in risks is perhaps more counterintuitive. In the present context, hidden heterogeneity means that the population from

3. This selectivity differs from that attributable to having an unrepresentative sample of the dead, for example, because of poor preservation of infant skeletons. Our point is that all samples of the dead are inherently unrepresentative of the original living population at risk of death, even a skeletal collection that is a perfect random sample of all those who died.

3. This generalization leaves aside the question of how sensitive a skeletal lesion may be as an indicator of disease. Pathological changes in bone tend to be markers of chronic conditions, and often only a small fraction of cases develops skeletal lesions. Thus, as paleopathologists frequently point out, skeletal lesions may be expected to underestimate the population prevalences of their associated conditions. At present, it is difficult to weigh the countervailing effects of the underestimation caused by low sensitivity and the overestimation caused by selectivity bias. There is no reason to expect, however, that these errors will perfectly balance each other. Therefore, we doubt that it will ever be possible to estimate population prevalences reliably from skeletal lesion frequencies.
which the skeletal series is drawn was made up of an unknown mixture of individuals who varied in their underlying frailty or susceptibility to disease and death. Such heterogeneity may arise from genetic causes, from socioeconomic differentials, from microenvironmental variation, or even from temporal trends in health, since most skeletal series, especially large ones, represent accumulations over more or less prolonged periods of time. Hidden heterogeneity in risks makes it virtually impossible to interpret aggregate-level age-specific mortality rates in terms of individual risks of death. This problem is far from restricted to archaeological samples but is now widely acknowledged throughout demography (Vaupel and Yashin 1985a, b; Trussell and Rodriguez 1990).

All these problems reflect two unavoidable facts. First, it is impossible to obtain direct estimates of demographic or epidemiological rates from archaeological samples. Such estimates require that we know the amount of exposure to the risk of illness or death, that is, the number of individuals exposed to the risk and the length of their exposures, neither of which is ever known with sufficient precision in archaeological research to permit formal estimation by standard methods. Assumptions such as stationarity are ruses that allow us to reconstruct exposure indirectly, if imperfectly. Second, although “health” (however defined) is a biological characteristic of the individual, inferences about it must be based on aggregate- or population-level statistics. Just as it is a truism in epidemiology that the single-case study is of limited value, it is widely recognized in paleodemography that reports on single specimens tell us little about the disease experience of ancient populations. However, when the population of interest is heterogeneous for factors that affect health, the relationship between aggregate measures and the experience of the individuals making up the aggregate can be remarkably tenuous.

For example, imagine a living population comprising three subpopulations, all of which are potentially exposed to an acquired condition reflecting some form of stress. This condition, when it occurs, increases the risk of death and may eventually produce a distinctive skeletal lesion in survivors. Members of the first subpopulation never experience the stress; therefore, none of them develops skeletal lesions. The second subpopulation experiences moderate stress, sufficient to cause widespread sickness lasting long enough to produce bony lesions but causing few deaths; most individuals with bony lesions survive the stress but join the mortality sample later when they succumb to some other cause of death. The third subpopulation suffers heavy stress, resulting in numerous deaths soon after the onset of the disease and typically before a bony response is elicited; in this group, rapid death results in few, if any, skeletal lesions in the mortality sample. Considering the differential exposure to stress experienced by these three groups, it seems appropriate to rank them from high to low according to their level of health, at least with respect to the condition of interest. Judging solely from the skeletal lesions, however, there appear to be not three groups but only two: one healthy (unstressed) subpopulation with no lesions and one stressed subpopulation with many. The groups experiencing the least and the most stress are indistinguishable with respect to the distribution of this particular kind of lesion.

This fictitious (but not implausible) example introduces several of the themes explored in this paper: the possible presence within the population of interest of multiple, undetected subgroups that experience varying risks of disease and death, the difficulty of reconstructing population prevalences of pathological conditions from skeletal lesion frequencies, the complex relationship between the degree of stress and the likelihood of developing a lesion, and the possibility that individuals displaying lesions may actually be healthier than at least some individuals without lesions. In what follows, we show how each of these factors can confound the interpretation of osteological evidence about the health of past populations.

Hidden Heterogeneity and the Deviant Dynamics of Death

Life-table estimates of age-specific mortality are often used in paleodemography to draw inferences about the levels and age patterns of risks experienced by individual members of the population being studied. However, when individuals vary in frailty and thus their risk of death and this variation is “hidden” (i.e., not captured by observed covariates), death rates computed from aggregate-level mortality data may provide a misleading picture of individual-level risks. This can be shown with a very simple model, one that is intended not to be particularly realistic but simply to make a basic point about hidden heterogeneity and what Vaupel and Yashin (1985b) call “the deviant dynamics of death.”

Imagine that we observe a population of newborns, each of whom experiences an absolutely constant risk of death but who vary among themselves in their risks. The hazard of death (the continuous-time analog of the central mortality rate in the life table) experienced by each child might be modeled as \( h(t) = z_i \cdot h_c \), where \( t \) is the child’s age, \( z_i \) is a measure of its individual-level frailty, and \( h_c \) is a component of the hazard common to all children in the population. While we write \( h_i(t) \) as a function of age, both \( z_i \) and \( h_c \) are constants, therefore, the overall hazard for the \( i \)th child does not in fact change with age.

Now suppose that \( z_i \) is distributed in some regular fashion among the newborns; for example, the frailty of newborns might be normally distributed as in figure 1,

---

4. The term frailty, which has entered standard demographic usage, was originally defined by Vaupel, Manton, and Stallard (1979) as an individual’s relative risk of death compared to a standardized cohort risk. More generally, it refers to individual biological characteristics associated with persistent differences among individuals in susceptibility, propensity, or relative risk with respect to disease or death (Vaupel 1990).
FIG. 1. A model of heterogeneous frailty and selective mortality. Each child's hazard of death is assumed to be constant and proportional to its individual-level frailty \(z\). Frailty in turn is assumed to be normally distributed among newborns [top]. Deaths during the first week of life [dots] are selective with respect to the frailty distribution; that is, children of high frailty make up a disproportionately large fraction of all deaths. As a consequence, the frailty distribution shifts downward by the second week of life [bottom], and mean frailty decreases [arrow]. Since the aggregate-level hazard of death at each age is proportional to the mean frailty of survivors at that age, the aggregate hazard declines even though the hazards of the individual children remain constant.

In other words, the mean hazard at any given age is simply the product of the common hazard and the mean of the frailty distribution at that age.

The crucial point here is that both \(g_0(z)\) and \(z_1\) change over time, because deaths of children at any given age are selective with respect to the frailty distribution; the greater a child's frailty, the more likely it is to die. Thus, the frailty distribution shifts with age as high-frailty individuals are eliminated by death [fig. 1, bottom]. As a result, the mean frailty of surviving children declines with age, causing the mean hazard of death to decline as well. In other words, the aggregate-level risk of death, given by the mean hazard, drops with age even though the individual-level risks remain absolutely constant. Thus, where there is heterogeneity in frailty, changes in the aggregate-level hazard do not reflect the risks experienced by any of the individuals who make up the population.

The confounding effect that heterogeneous frailty can have on the relationship between aggregate-level mortality and individual health can be illustrated with reference to a real archaeological sample. Figure 2, top, shows the age-at-death distribution up to age five years for an unusually well-preserved sample of late prehistoric skeletons from a completely excavated Oneota cemetery in west-central Illinois (Milner and Smith 1990) and [below] age-specific hazards of death estimated from the age-at-death distribution under the assumption that the Oneota population was stationary. The smooth curve is a two-parameter negative Gompertz function fitted to the hazards by ordinary least squares.

\[ r(t) = 0.244\exp(-0.751(t)) \]
\[ r^2 = 0.901 \text{ (on log scale)} \]
west-central Illinois [Milner and Smith 1990]. A total of 264 skeletons is included in this sample, with 125.4 of them estimated to be below age five.\(^5\) The age-at-death distribution closely approximates the mortality patterns expected of a traditional, preindustrial population with high fertility [Milner, Humph, and Harpending 1989, Paine 1989]. Most important, there is no obvious under-enumeration of the very young in this sample, a common problem in archaeological collections. The age-specific hazards of death estimated from this sample by standard life-table techniques\(^6\) are entirely plausible for this sort of population [fig. 2, bottom], with mortality during the first six months of life just above 200 per 1,000 live births and a rapid and consistent decline in mortality beyond early childhood. The estimated rates are fit quite satisfactorily by a two-parameter Gompertz hazard function [fig. 2, bottom], a model that is widely used to describe the decline in human mortality at early ages [Gage 1990].\(^7\) From this point on, we will take the fitted Gompertz curve as a reasonable representation of early childhood mortality in the Oneota population, as estimated at the aggregate level.

What do these aggregate mortality rates reveal about the risks of morbidity and mortality experienced by individual children in the Oneota population? A conventional interpretation, common in both the health sciences and paleodemography, is that the “average” newborn infant is at a high risk of death but its risk declines rapidly as it grows older. This interpretation, which would arouse no controversy, is not necessarily consistent with recent individual-level studies of disease processes and specific causes of death. For example, the risks of morbidity and mortality from diarrheal disease, a major killer of children in the contemporary Third World, tend to increase over the first several months of life as exposure to pathogens increases and as maternal antibodies from breast milk decline [Long et al. 1992]. And yet the hazard rates in figure 2 seem to speak for themselves: the risk of death is initially high but then declines steadily.

\(^5\) Fractional individuals result from the way in which skeletons assigned to various ages are distributed over adjacent six-month intervals, a reflection of uncertainty in estimating the age of individuals.

\(^6\) These hazard rates were estimated on the assumption that the population was stationary [see Milner, Humph, and Harpending 1989, Paine 1989]. We make this assumption here strictly for convenience, since our intention is not to highlight the nonstationarity problem but rather to focus on the issues of heterogeneity and selectivity. Paine [1989] applied maximum-likelihood methods to fit model stable populations to the Oneota sample and estimated that the population growth rate was approximately –0.002. This estimate suggests that, while the Oneota population may not have been absolutely stationary, it was probably not far from it. However, because the likelihood surface was quite flat near the maximum, it is impossible to rule out either positive or negative growth rates of small absolute value. Consequently, the hazards in figure 2 should not be taken as final estimates of mortality at these ages.

\(^7\) Many applications of the Gompertz function to preadult mortality are based on a three-parameter form that includes a constant, age-independent component [we are grateful to Timothy Gage for this observation]. For the age-range considered here, this more complicated specification adds little to the analysis.

Unfortunately, the aggregate-level hazards of death may tell us disturbingly little about individual risks of disease and death if the population is heterogeneous with respect to frailty, as any real population almost certainly must be. For example, assume that the Oneota population that gave rise to the skeletal sample was actually made up of two distinct and nonoverlapping subpopulations, denoted 1 and 2, with differing levels and age patterns of mortality. Of course, there is no prior reason to believe that exactly two subpopulations exist, but this is the simplest form of heterogeneity imaginable. For further simplicity, assume that each subpopulation was homogeneous with respect to death (i.e., every individual within a given subpopulation experienced the same hazard of death at each age). How might the two subpopulations be combined to yield the Gompertz mortality curve characteristic of the Oneota skeletal sample as a whole?

The model presented in the appendix provides a way to address this question. All that has to be done is to specify a mathematical form for the age-specific hazard function in one of the subpopulations in order to find the corresponding function in the other. Suppose, for example, that the population contains a group of relatively disadvantaged [i.e., frail] children who represent approximately half of all newborns. If the hazard of death for these children is constant and equals 0.25 at all ages, then we have the mixture of subpopulations shown in figure 3, A. The hazard experienced by the second, less frail subpopulation now declines at early ages, but it does so much more rapidly than the aggregate mortality curve, reaching zero by about one year of age. In this instance, the aggregate hazard curve is not an accurate reflection of the hazards experienced by individuals in either subpopulation. Instead, the shape of the aggregate hazard function is dominated by a selection process: because of their high risk of death, children in subpopulation 1 are progressively selected out of the initial pool of newborns, and subpopulation 2 comes to represent an ever-increasing fraction of all surviving children. The aggregate hazard thus converges on the hazard of subpopulation 2 as time goes by.

Next, suppose that the hazard in subpopulation 1 remains constant with age but drops to 0.16 and that the fraction of newborns belonging to subpopulation 1 is reduced to 0.4. These changes produce the curves shown in figure 3, B. Now the hazard function in subpopulation 2 starts out higher than the aggregate curve but drops below it at about nine months of age. Again, the aggregate hazards fail to capture the pattern of risks of any individuals in the population.

An immediate reaction to these subpopulation hazard curves is liable to be that, while mathematically possible, they make little sense from either a demographic or a biological point of view. We suggest that this reaction may be incorrect, precisely because commonsense notions about what age-specific mortality rates ought to look like are based entirely on aggregate-level data. By and large, we have little idea how hazards should behave at the individual level. As we have already noted, there
Fig. 3. Different mixtures of subpopulation hazards, $h_1(t)$ and $h_2(t)$, all of which are equally consistent with the same Gompertz function for the average age-specific hazard, $\bar{h}(t)$ (based on the two-point frailty model of Vaupel and Yashin [1983a]). In each case, the Gompertz curve is the same as in figure 2 (bottom). A, the mixture assuming that $h_1(t) = 0.25$ and $p$ (the proportion of newborns belonging to subpopulation 1) = 0.5; B, the mixture assuming that $h_1(t) = 0.16$ and $p = 0.4$; C, the mixture assuming that $h_1(t) = 0.02 + 0.04t$ and $p = 0.4$; D, the mixture assuming that $h_1(t) = 0.3 - 0.1t$ and $p = 0.2$.

is reason to suspect that the hazard for diarrhea may actually go up at early ages as the duration of exposure increases, and the same is likely to be true of other infectious diseases. Figure 3, C, allows for such an increasing hazard in subpopulation 1. Although radically different from the aggregate pattern, an increasing hazard in one subpopulation is perfectly compatible with a decreasing aggregate hazard so long as the hazard in subpopulation 2 starts higher than the aggregate curve but drops below it at about one and a half years of age.

Finally, with a linear decline in the hazard for subpopulation 1 (fig. 3, D), the curve in subpopulation 2 behaves very oddly indeed. It roughly parallels the aggregate hazard up to age two but then skyrockets as the aggregate curve approaches zero.

These examples could be multiplied endlessly. However, the general point of the exercise should be clear. If there is hidden heterogeneity in the risk of death, then a given aggregate mortality pattern can result from an infinite number of possible combinations of subpopulation mortalities. Moreover, we rarely if ever have any idea how many subpopulations actually exist. It could in fact be two, but it could also be three, four, or more, and indeed at the limit every individual could have his or her own unique level of frailty, giving rise to the sorts of continuous distributions of frailty shown in figure 1. This fact only adds to the range of possible models that are perfectly consistent with a given aggregate mortality curve. The problem is not simply a reflection of the fact that we have arbitrarily assumed two subpopulations.

This is a specific instance of what is known in mathematical statistics as the identification problem [Hsiao 1990, Lancaster 1990]. A model is said to be uniquely identifiable if, when fitted to a particular set of data by some formal method such as maximum likelihood, there is one and only one value of each of its parameters.
that is consistent with the data. The models we have just reviewed are unidentifiable in that an infinite number of values of the subpopulation hazards are equally compatible with the observed aggregate hazards. For more disturbing, every other model of heterogeneous frailty, whether discrete or continuous, is equally unidentifiable. What this means in practice is that there is no way to infer individual risks of death from aggregate mortality data in the absence of prior theory that specifies the way in which frailty varies among individuals.

The implications of this problem for the demographic treatment of mortality are profound. Indeed, hidden heterogeneity in risks and the identification problem associated with it have become dominant concerns in demography during the past decade [Keyfitz and Littman 1980; Heckman and Singer 1984b; Vaupel and Yashin 1985a, b; Heckman and Walker 1987; Manton and Stallard 1988; Lancaster 1990; Trussell and Rodriguez 1990]. Two general strategies have been developed for dealing with hidden heterogeneity. One is to model the distribution of frailty directly in a way that makes biological sense [Weiss 1990]. Often, however, there is too little prior theory about how the relevant processes work to permit this sort of etiologic modeling. The second approach involves the development of more general models of heterogeneous frailty [e.g., Heckman and Singer 1984a, Trussell and Richards 1985, Heckman and Walker 1987, Vaupel 1990]. Here one specifies an arbitrary probability density function for the unobserved heterogeneity and then checks to see if the final numerical results are robust for changes in the specification. While it is not yet certain how successful these strategies will be, recent applications in the analysis of data on morbidity and mortality offer some hope [Manton and Stallard 1988, Weiss 1990].

For the present, it is important for paleodemographers and paleopathologists to be aware of the general problem. When hidden heterogeneity in risks exists, as we suspect it must in all populations, then population-level mortality patterns may tell us little if anything about individual risks of death. As a consequence, inferences from paleodemographic life tables to the health status of individuals in prehistoric populations are, to say the least, problematic. Even if the osteologist wishes merely to rank prehistoric populations by their general level of health, the problem is still important: populations cannot be compared meaningfully if their distributions of frailty differ in unknown ways. A solution to this problem will require serious thought about the best ways to model the variation in frailty that almost certainly exists in all archaeological contexts.

Selective Mortality and the Mysterious Meaning of Morbidity

Heterogeneity and selectivity complicate the interpretation of morbidity as well as mortality. As has been pointed out [Cook and Buikstra 1979; Cook 1981, 1984; Palkovich 1985; Milner and Smith 1990], skeletal samples are by definition composed of a very special subset of individuals who were exposed to the risk of death at each age, namely, the nonsurvivors. By their nature, the dead are a biased sample of all the individuals in the population who were alive at a given age. Because of this selectivity, it is surprisingly difficult to draw inferences about the prevalence of pathological conditions marked by hard-tissue abnormalities in a once-living population from the frequency of lesions in a skeletal sample. The reason for this difficulty is simple: if the disease state of interest affected individuals' risks of death [and presumably that is precisely why it is interesting], then the frequency of the disease must be greater among nonsurvivors than among the living. Stated differently, proportional mortality [the fraction of deaths attributable to a particular cause] overestimates population prevalence because mortality itself is selective.

Selective mortality is a special concern when bone lesions that were active at the time of death indicate an elevated risk of death. Enhanced risk may arise directly from the disease process that induced the bony response—that is, that particular disease may be the primary cause of death. Alternatively, the disease may be a contributory cause, or it may simply be correlated with some other factor that affects mortality. In all these instances, the presence of an active skeletal lesion marks a person at elevated risk of death. Therefore, the frequency of active lesions in a skeletal sample is greater than the fraction of affected individuals in the living population from which the sample was drawn. This is true not only of the skeletal sample as a whole but also of its constituent age- and sex-groups considered individually. Without knowledge of the risk of death associated with a particular condition, it is impossible to predict from modern population prevalences [e.g., from epidemiological studies] the proportion of a skeletal sample that would be expected to show a hard-tissue response. Despite this fact, direct extrapolations from clinical and epidemiological findings to mortality samples are sometimes encountered in the literature on archaeological skeletons [Morse 1978, Blakely and Detweiler-Blakely 1989].

The effects of selective mortality can be illustrated with reference to the Oneota skeletal series. Numerous crania in this collection exhibit the distinctive orbital and cranial-vein lesions usually referred to, respectively, as cribra orbitalia and porotic hyperostosis. These lesions are known to accompany various anemias and, when encountered in pre-Columbian skeletons from the Americas, are usually attributed to iron-deficiency anemia [El-Najjar et al. 1976; Steinbock 1976; Mensforth et al. 1978; Ortner and Putschar 1983; Stuart-Macadam 1985, 1989; Larsen 1987; Palkovich 1987]. The presence of specific nutrients are frequently part of a broader pattern of nutritional inadequacy, often complicated by infection. It has long been argued that the elevated mortality rates among young children common in the contemporary Third World reflect the interaction of infection and malnutrition [Gordon, Chitkara, and Wyon 1963, Gordon, Wyon, and Ascoli 1967, Scrimechaw, Taylor, and Gordon 1968, Pufer and Serrano 1973, Scrimechaw 1977,
of these cranial lesions is thought to reflect a bony response in early childhood to pathological expansion of hematopoietic tissue in restricted diploic spaces (Stuart-Macadam 1985). Although iron-deficiency anemia certainly occurs in later life as well, it is unlikely to induce comparable bony responses at those ages. Once developed, however, these cranial lesions can be retained throughout life, as is indicated by their frequent occurrence in the skeletons of adults from many parts of the world.

In the Oneota skeletal collection, both cribra orbitalia and porotic hyperostosis are present on most of the juvenile crania but few adult crania (fig. 4). Lesions indicative of an active bony response are restricted to children, principally those less than five years old. The discrepancy in the numbers of affected juvenile and adult crania presumably reflects ongoing selectivity that preferentially eliminates the most frail individuals within each age interval. The remodeled lesions, in this view, identify individuals who managed to survive a period of early illness only to succumb somewhat later to stress of unknown origin—stress that may or may not have had any relation to the conditions that produced the lesion.

Since we are viewing the end result of selective entry into the mortality sample, the frequency of pathological hard-tissue lesions, especially those that appear to be active at the time of death, should be higher than their prevalence in a comparable age-group in the living population. This is shown in figure 5, which presents estimates of the proportion of the once-living population at each age that may have exhibited bony lesions. The reconstructed population at risk at each age was derived by summing all observed crania and then decrementing the sum by the number of individuals that died in the previous age-group; the number of individuals with lesions at each age in the reconstructed population was similarly estimated. This reconstruction is based on three assumptions: (1) the population was stationary, (2) the bony lesions indicate a condition that developed during the first five years of life, and (3) the lesions are not entirely obliterated by subsequent bone remodeling. Undoubtedly, all these assumptions oversimplify reality. Nonetheless, they permit a tentative reconstruct—

Chandra and Chandra 1986), and several osteologists have stressed this synergism in the etiology of cribra orbitalia and porotic hyperostosis (Mensforth et al. 1978, Walker 1985, Palkovich 1987, Stuart-Macadam 1989).

Fig. 4. Numbers of Oneota crania in five age-groups with active cribra orbitalia or porotic hyperostosis, inactive (remodeled) lesions, and no lesions. While the differentiation of active and inactive lesions is inexact, it is the overall age pattern, not individual crania, that is important here.

Fig. 5. Age-specific proportions of Oneota crania with cribra orbitalia or porotic hyperostosis (solid curve) and a reconstruction of the age-specific prevalences of those lesions in the living population (broken curve). Because of the selective effects of mortality, lesion frequencies are higher among those who died (the skeletal sample) than in the reconstructed population at risk.
tion of the population at risk and the frequency of lesions in that population and thereby reveal trends during the juvenile years that reflect selective mortality. Under this pattern of selectivity acting upon heterogeneous susceptibility to stress, the lesion frequencies among the dead are consistently greater than the prevalence of lesions in the living population.

It is important to emphasize that this particular reconstruction is based on a series of assumptions (e.g., stationarity) that may not be correct. When such assumptions cannot be made, there is no way to reconstruct the numbers of individuals at risk and thus no way to infer population prevalences from proportional mortality.

Selective mortality is also a concern in dealing with more subtle markers of poor health. For example, because of its known multifactorial etiology, short stature is often taken as an indicator of general stress. A number of researchers have used the length of long bones in juveniles to identify differences in linear growth among groups of skeletons drawn from separate time periods and have linked the bone length data to other skeletal evidence of ill health [Hummert and Van Gerven 1983, Cook 1984, Goodman et al. 1984, Mensforth 1985]. In these studies, shorter bones are consistently interpreted as reflecting increased stress. This seemingly straightforward interpretation does not take into account that the differences in stature are observed strictly in individuals who died at each age. Alternative interpretations are possible when the skeletons are viewed as nonsurvivors who entered the sample as a result of selective mortality acting on heterogeneous frailty. In figure 6, for example, we assume that the distribution of stature among living children at a given age is absolutely invariant between two populations or time spans, that is, that there is no differential, stress-related stunting between these two groups. We further assume that mortality is selective with respect to the distribution of stature—in particular, that shorter children are at higher risk of death, a relationship commonly observed in the contemporary developing world [Kiellmann and McCord 1978, Chen, Chowdhury, and Huffman 1980, Heywood 1982]. Under these assumptions, variation in the stature of dead individuals may indicate different levels of mortality [and perhaps stress] but not in the direction commonly supposed. When mortality is high, a larger fraction of the entire distribution of stature is represented among dead individuals; as a result, dead individuals are comparatively tall on average [fig. 6]. If mortality slackens, only the most frail (i.e., short-statured) individuals die. Periods of low mortality are therefore characterized by comparatively low mean stature among the dead. In general, the frequency of apparent “stunting” among the dead is uninformative about the distribution of stature or relative health among the living unless the level of mortality and the relationship between stature and frailty are known. Once again, proportional mortality is a poor guide to population prevalence.

Proportional Mortality and Competing Causes of Death

One solution to the difficulties posed above is to abandon any attempt to estimate population prevalences from frequencies of skeletal lesions, interpreting such frequencies strictly as estimates of proportional mortality or the fraction of deaths attributable to a given cause.

10. Differential susceptibility to life-threatening stress has also been invoked to explain the smaller size of permanent teeth in juveniles when compared with their adult counterparts [Guagliardo 1982].

11. It is worth noting that the interpretive problems posed by selective mortality are not restricted to comparisons between ancient samples. In a recent review of studies on hard-tissue growth and development, differences in long bone length between prehistoric and contemporary samples are attributed to the “poorer environments” of earlier peoples, thus ignoring the fact that the prehistoric samples are made up of dead people and the contemporary samples of the living [Johnston and Zimmer 1989:18].
This, in effect, is the solution advocated by Cohen [1989:106], who suggests that “estimates of the incidence [sic] of disease in skeletal populations are more useful for comparison to one another than for direct comparison to rates reported in living groups.” Unfortunately, estimates of proportional mortality “can give rise to misleading conclusions if used to compare [the] mortality experience of populations with different distributions of causes of death” [Last 1988:106; see also Miettinen 1985:260]. The reason for this is that, by definition, proportional mortality for a single cause of death reflects the importance not only of that particular cause but also of all the other competing causes of death experienced by the population.

This fact is easiest to demonstrate for completely independent competing causes, although it is equally true of interacting causes. If a population of individuals is exposed to k independent causes of death, then the total hazard of death for an individual at age t is simply the sum of the hazards associated with each independent cause (Elandt-Johnson and Johnson 1980:273):

$$h_{total}(t) = h_1(t) + h_2(t) + \cdots + h_k(t).$$

This is the total hazard experienced by an individual at age t, not the total population hazard combining individuals of differing frailty at that age. The number of deaths to individuals exposed to the total hazard given by equation 2 depends not only on the total hazard itself but also on the number of individuals exposed to that hazard (N) and the duration of exposure (δt). For δt small, the total number of deaths at age t is approximately

$$N h_{total}(t) \delta t = N h_1(t) \delta t + N h_2(t) \delta t + \cdots + N h_k(t) \delta t.$$

Now suppose that we are interested specifically in cause 1 because it leaves an unambiguous skeletal signature. The proportional mortality attributable to cause 1 is simply

$$\frac{\text{deaths from cause 1}}{\text{total deaths}} = \frac{N h_1(t) \delta t}{\left[ \sum_{j=1}^{k} N h_j(t) \delta t \right]}$$

$$= \frac{h_1(t)}{\left[ \sum_{j=1}^{k} h_j(t) \right]} = \frac{h_1(t)}{\left[ \sum_{j=1}^{k-1} h_j(t) + h_k(t) \right].}$$

Writing the expression this way highlights the dependence of the proportional mortality associated with cause 1 on the hazard of death from some other cause, k.

If conditions in our population change in such a way that the hazard of death from cause k decreases or is eliminated altogether while the hazard from cause 1 is unaffected, the proportional mortality from cause 1 will then increase even though its hazard remains unchanged. This effect is illustrated in figure 7, which also shows that proportional mortality from any one cause is determined in part by the total number of causes operating in the population. Paradoxically, what might be considered an improvement in health [either a reduction in risk from cause k or a decrease in the total number of causes] will produce an increase in the proportional mortality from the remaining causes. A familiar example of this phenomenon is the recent increase in cancer deaths as a fraction of all deaths in industrial societies, reflecting the virtual elimination of deaths from such infectious diseases as tuberculosis, smallpox, diphtheria, and typhoid. In this instance, the shift in causes of death has resulted in greater longevity but also a higher prevalence of cancers and more years of life spent disabled, indicating that the very concept of “health” may be more complex than is often assumed.

The dependence of proportional mortality on the total distribution of causes would occasion no alarm if we knew every single competing cause present in the population. This is difficult enough, and perhaps impossible, to achieve for living populations. For prehistoric populations, we know only a tiny subset of all competing causes, namely, those that leave skeletal traces. As a result, it becomes extremely difficult to interpret differences in proportional mortality among skeletal samples. An increase in the fraction of deaths associated with a particular cause may indicate an increased hazard of death from that cause, but it may also indicate a decrease in the hazard of death from some other, totally unrelated cause. Thus, the same change in proportional mortality may signal either an improvement or a deterioration in the population’s health. Once again, we are faced with a fundamental identification problem when dealing with paleopathological and paleodemographic data.

### Pathological Processes and the Healthiness of the Healed

Paleopathologists routinely note the importance of distinguishing active from inactive or healed lesions, even while recognizing that this simple dichotomization may not reflect all the complexities of the pathological process [Mensforth et al. 1978, Martin, Goodman, and Armelagos 1985, Walker 1985, Palkovich 1987]. As crucial as this distinction may be from a biological perspective, however, it further complicates the interpretation of paleopathological evidence. The active-inactive distinction represents an attempt to separate disease processes that were ongoing at the time of death, including chronic conditions, from those that had occurred earlier. If this distinction is correct, the presence of an inactive lesion indicates survival of a disease process earlier in life and thus may signify an individual whose frailty is low compared with those who died at earlier ages.

From this perspective, the differing age patterns displayed by various kinds of lesions take on a new and largely unexplored significance. In figure 8, we show the age-specific frequencies of cribra orbitalia, and porotic
hyperostosis and of periostitis during the first four years of life in the Oneota sample, distinguishing between active and healed lesions. These lesions are of interest because of their distinct etiologies: crudely speaking, periostitis is a lesion of infectious origin, while cribra orbitalia and porotic hyperostosis reflect synergistic interactions of infection and nutrition. Bearing in mind that the estimates in figure 8 are lesion frequencies among the dead at each age, one might suggest that cribra orbitalia and porotic hyperostosis represent the common view of nutritional stress: children are differentially susceptible early in life but once stressed remain frail throughout early childhood. Therefore, even skeletons with healed lesions are selectively added to the mortality sample. Periostitis, in contrast, shows a quite different pattern. Early life is again a time of elevated stress, but now there appears to be no differential tendency for individuals with healed lesions to die and thus enter the skeletal sample. In fact, individuals with healed lesions appear to be much less likely to die than those with active lesions. In this case, the presence of a healed lesion may actually be evidence that the individual was of lower frailty and therefore at a lower risk of death; because of his or her superior condition, such an individual was able to survive long enough to manifest a healed lesion. Frailer individuals, in contrast, would die soon after the onset of infection, either before a bony response had time to develop or during the time when bone lesions were active. Thus, the presence of a healed lesion may sometimes indicate a state of comparatively good overall health.

An alternative and radically different interpretation of the data on cribra orbitalia and porotic hyperostosis is possible. In this interpretation, virtually all individuals develop the lesions early in life, but the underlying condition places these individuals at no elevated risk of death whatsoever. Thus, all children survive long enough to recover and for their lesions to heal, unless they happen to die from some unrelated cause. In this scenario, but only in this scenario, the frequencies of
active and healed lesions in the skeletal sample would provide unbiased estimates of the population prevalence of such lesions. Of course, if this proved to be true, cribra orbitalia and porotic hyperostosis would be remarkably uninformative about stress and the risk of death. Clearly, this interpretation makes little sense given what is known about the etiology of these lesions. However, the general point stands: unless we have a strong prior model of the way in which a given lesion, whether active or inactive, relates to frailty and of the way in which frailty varies among individuals, skeletal lesion frequencies cannot be interpreted in any straightforward fashion.

The Tangled Tales That Dead Folk Tell: An Extended Example

Skeletal samples from archaeological contexts can be expected to present all the problems associated with non-stationarity, hidden heterogeneity in risks, and selective mortality simultaneously. In such situations, obvious or intuitive interpretations are likely to be badly misleading. In this section, we illustrate some of the difficulties that can arise in interpreting the distribution of hard-tissue lesions that are relics of stress followed by survival. For concreteness, we restrict attention to the developmental defect of the teeth known as enamel hypoplasia, which provides evidence of stress experienced during childhood [Goodman and Rose 1990].

Goodman and Armelagos (1988) have reported data on enamel hypoplasia in the permanent dentition for three samples of late prehistoric skeletons from Dickson Mounds, Illinois. These data indicate that the overall fraction of individuals with such enamel defects increases consistently from the earliest to the most recent sample, that the overall mean age at death is lower in the most recent sample than in the two earlier ones, and that, except for the earliest period, skeletons with enamel hypoplasia within each sample had a lower mean age at death than their contemporaries without enamel hypoplasia. They suggest three hypotheses to account for the lower age at death in individuals with enamel hypoplasia: [1] that the enamel defects directly reflect frailty, lesions and early death being correlated outcomes of an innate constitutional susceptibility to stress; [2] that childhood stress resulting in the formation of an enamel defect permanently damaged the individual’s capacity to mount an effective biological response to later stress, and this acquired susceptibility

12. This particular study is singled out for four reasons. First, these researchers have been instrumental in developing the data-recording protocol now commonly used for hypoplastic enamel defects. Second, they have produced some of the most innovative osteological analyses to date. Third, their use of carefully collected dental data, in conjunction with skeletal-lesion and age-at-death information, contributed in the early 1980s to a reevaluation of the biological effects of the adoption of agriculture [Goodman et al. 1984]. Finally, we believe that they deserve special credit for explicitly considering multiple hypotheses in interpreting their results.
led to premature death at a later time; and [3] that social conditions produced disproportionate exposure of selected segments of the population to stress of diverse origins, members of social groups with higher risks of exposure exhibiting more stress-related enamel defects as well as premature mortality. Although Goodman and Armelagos [1988] do not rule out any of these three hypotheses and indeed acknowledge that others may be possible, they lean toward the third.

A fourth hypothesis might also account for their findings. This hypothesis is based on the fact that enamel hypoplasia does not occur unless the child survives the period of stress and resumes normal enamel formation (Suckling 1989). Under this hypothesis, the population from which each skeletal sample was drawn was made up of two subgroups with differing levels of fertility and mortality—one relatively advantaged [for whatever reason] and the other disadvantaged. Members of the disadvantaged group did not entirely escape childhood insult but were able to survive it long enough to develop enamel hypoplasia and then go on to enjoy higher fertility than their disadvantaged counterparts. Thus, individuals with observable lesions were principally from the less frail [i.e., advantaged] group, and they had a lower mean age at death than the disadvantaged group because of their higher fertility, not because of poor survival. If the differences in survival and fertility between the two groups persisted across generations, then the advantaged group would come to represent an increasing fraction of the population as a whole. This hypothesis would therefore explain the temporal trends reported in both lesion frequency and mean age at death. This hypothesis and the one preferred by Goodman and Armelagos [1988] are mutually exclusive, since under our hypothesis individuals with detectable lesions and early death are preferentially drawn from the advantaged group.

The implications of this fourth hypothesis for the distribution of skeletons with hard-tissue lesions can be explored using a simple demographic model that generates various statistics about hypothetical populations given two input parameters, the total fertility rate (TFR) and the level of mortality. Expected proportions of dead individuals displaying hypoplasia at each age were derived from this model under the following assumptions: [1] The accumulation of skeletons represents a random sample of deaths from a mixture of two stable populations, [2] the advantaged and disadvantaged groups of our hypothesis, the advantaged group initially makes up 10% of the pooled population, but the mix of the two populations changes over time according to their respective fertility and mortality patterns. Both populations experience fairly high mortality, corresponding to an expectation of life of about 30 years for the disadvantaged group and 37 years for the advantaged group. [3] The disadvantaged group has low fertility, in particular a total fertility rate of four, approximating levels reported for the !Kung San (Howell 1979, Harpending and Wandsnider 1982) and the Gainj of highland New Guinea (Wood, Johnson, and Campbell 1985). The advantaged population, in contrast, has high fertility, corresponding to a total fertility rate of eight (four and eight are about equally distant from the mean total fertility rate observed in preindustrial societies (Wood 1990). [4] Because of their comparatively poor condition, children in the disadvantaged group who become seriously ill die very quickly, thus, they do not survive episodes of stress that would otherwise have induced an enamel defect. For example, the children in this group might not have sufficient nutritional reserves to sustain them through a bout of diarrheal disease. Consequently, there are no dental signs of stress among the survivors of childhood in the disadvantaged population simply because those who experience stress do not survive long enough to develop hypoplastic enamel defects. In contrast, childhood illness is often survived in the advantaged group, resulting in numerous older children and adults who exhibit enamel hypoplasia reflecting episodes of stress earlier in life.

Table 1 lists the characteristics of the two model populations constructed on the basis of the above assumptions, while figure 9 shows how the expected proportions of dead individuals with enamel hypoplasia change over time. The better health of the advantaged population is reflected in its higher survival and fertility. As a consequence of its higher fertility, the mean age of the skeletal remains from the advantaged group is 15 years, while the mean age of those from the disadvantaged group is 32. We emphasize that this difference is entirely attributable to the difference in fertility: it is uninformative about any difference in mortality. Finally, because the presence of a skeletal lesion is indicative of low frailty (since frail individuals die before developing a

---

13. Under this model, age-specific fertility schedules were generated using the rates given by Coale and Demeny [1983] for a mean age of women at childbirth equal to 29 years; these rates were multiplied by one-half the input value of the TFR, roughly equal to the gross reproduction rate. Age-specific mortality rates were produced by the two-parameter logit method developed by Brass [1971, 1979]. Using the so-called African standard life table as a starting point, a wide variety of mortality schedules can be generated by varying $\gamma_0$, the expectation of life at birth [set by the parameter $\alpha$ in the Brass model], and the ratio of childhood to adult mortality [set by the parameter $\beta$]. In order to accommodate the varying ways in which ages are pooled in published paleopathological sources, we have generated a smooth year-by-year survivorship curve by spline interpolation through the African standard values and have used this as the standard against which variant survivorship schedules are generated.

14. A stable population is one with constant age-specific fertility and mortality rates and an equilibrium age distribution but not necessarily a zero growth rate. The stationary population is a special case of the stable population.

15. The mortality schedules for the two populations are derived from the Brass logit model using the African standard with $\alpha = 0.4$ and $0.2$ in the disadvantaged and advantaged groups, respectively, and $\beta = 1$ for both groups. Sensitivity tests of the model indicate that the qualitative results do not depend upon either the specific levels of mortality or the standard mortality schedule chosen.
TABLE I

Characteristics of Two Model Populations

<table>
<thead>
<tr>
<th>Input into model</th>
<th>Advantaged Population</th>
<th>Disadvantaged Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mortality</td>
<td>moderate ((e_0 = 37) yrs.)</td>
<td>high ((e_0 = 30) yrs.)</td>
</tr>
<tr>
<td></td>
<td>((TFR = 8))</td>
<td>((TFR = 4))</td>
</tr>
<tr>
<td>Fertility</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean age at death</td>
<td>15</td>
<td>32</td>
</tr>
<tr>
<td>Mean age at death, individuals &gt; 15 yrs.</td>
<td>44</td>
<td>53</td>
</tr>
<tr>
<td>Crude birth rate [per 1,000]</td>
<td>56</td>
<td>30</td>
</tr>
<tr>
<td>Crude death rate [per 1,000]</td>
<td>37</td>
<td>33</td>
</tr>
<tr>
<td>Intrinsic rate of increase</td>
<td>+0.029</td>
<td>-0.003</td>
</tr>
</tbody>
</table>

As the intrinsic growth rates in Table I indicate, our model does not produce a stable mixture of populations: the advantaged population is growing rapidly while the disadvantaged group is almost stationary and indeed is declining slightly. Therefore, given enough time, the disadvantaged population will disappear from the mix. As a result of this intergenerational selection process, the mean age at death will go down and the frequency of skeletal lesions will go up as time goes by (Fig. 9), even though the characteristics of the separate populations remain fixed. The model thus accounts for the temporal trends observed at Dickson Mounds (Goodman and Armelagos 1988).

Reinterpreting the Transition to Agriculture

The issues raised here have serious implications for what is arguably the most important potential contribution of paleopathology and paleodemography to health science: elucidation of the biomedical consequences of the transition from hunting-and-gathering to agriculture. In 1982 a decade of osteological findings pertaining to this transition was summarized and evaluated at the influential conference “Paleopathology at the Origins of Agriculture” (Cohen and Armelagos 1984a). In summaries of this conference, Cohen and Armelagos (1984b) and Roosevelt (1984) point to a common but not universal pattern in which the adoption of agriculture and sedentary settlement is associated with a reduced mean age at death and comparatively high frequencies of skeletal lesions, including those attributable to undernutrition and infectious disease. These data are interpreted as indicating increased stress and reduced survival at various ages, both signs of an apparent deterioration in general health among the early agriculturalists. Cohen (1989) summarizes further evidence related to this issue and interprets it as showing that agriculture and sedentism resulted in increased local environmental contamination with pathogens, more frequent person-to-person transmission of infections, and a decline in dietary quality and diversity.

While these interpretations may be correct, the data do not force them upon us. As is emphasized by Sattenspiel and Harpending (1983) and by Milner, Humph, and Harpending (1989), it is at least as plausible that the shift in mean age at death reflects an increase in fertility rather than an increase in mortality associated with a deterioration in general health. Such an elevation in fertility, if it occurred, might have reflected greater availability of digestible weaning foods and, hence, shorter periods of lactational infecundability among the early agriculturalists (see Buikstra, Konigsberg, and Bullington 1986). Further, the logic developed in this paper suggests that the higher frequencies of skeletal lesions...
observed in early agricultural samples could reflect an enhanced ability to survive episodes of illness and stress or an amelioration of other, competing and perhaps unobservable causes of death. Thus, the trends summarized by Cohen and Armelagos (1984b), Roosevelt (1984), and Cohen (1989) are equally consistent with an interpretation that is the opposite of the one they endorse.

It is important to emphasize that our reinterpretation of the health consequences of early agriculture is not necessarily more correct than previous interpretations. And, indeed, we suspect that both interpretations may be correct for different periods and locations. The point, however, is not that we are right and other authors are wrong but that the data support both interpretations equally well. The correct model for linking data and interpretation is in this instance, as in so many others, unidentifiable.

Is There Hope?

We remain convinced that osteological research will continue to play a central role in the study of human disease interactions. This is especially true of attempts to understand variation in the health of peoples with contrasting subsistence and settlement systems, natural environments, levels of cultural complexity, and population sizes. Present characterizations of the health of preindustrial populations are based upon a complicated amalgam of information from modern medical surveys, the ethnographic and historical records, and archaeological excavations of ancient sites [e.g., Cohen 1989]. Archaeological evidence is crucial to this effort since contemporary preindustrial populations, especially modern hunter-gatherers, are undoubtedly a highly select sample of all such populations that have ever existed. We will continue to need the perspective and time depth that archaeology can provide.

Nevertheless, using the archaeological record to infer the health characteristics of a once-living population is far more difficult than is commonly acknowledged. Most of the widely recognized problems associated with osteological materials from archaeological sites—including determining the age and sex of skeletons [Buikstra and Mielke 1985; Işcan and Loth 1989, Işcan 1989, St. Hoyme and Işcan 1989, Meindl, Russell, and Lovejoy 1990], the inadequate size and unrepresentative composition of many skeletal samples [Howell 1982, Milner, Humpf, and Harpending 1989, Paine 1989], variable and selective preservation of bones [Gordon and Buikstra 1981, Waldron 1987, Walker, Johnson, and Lambert 1988], and the differential diagnosis of lesion-producing diseases [Steinbock 1976, Buikstra and Cook 1980, Zimmerman and Kelley 1982, Ortner and Putschar 1985, Larsen 1987]—can probably be resolved by additional excavation, the development of better chronological and contextual controls, reference to comparative materials drawn from well-documented medical collections, and the continued refinement of field, laboratory, and analytical methods. In contrast, the problems discussed in this paper require a complete rethinking of the relationships among pathological processes, the risk of death, and the formation of mortality samples. In the face of demographic nonstationarity, hidden heterogeneity in frailty, and selective mortality, intuition can be a poor guide for evaluating how demographic and epidemiological changes should be reflected in a sample of skeletons. Unfortunately, existing theory may be no more useful as a guide because most of it is concerned with the characteristics of the living, not the dead. As a consequence, models for death distributions are poorly developed.

What must be done to elucidate the complex relationship between the health of a community and the characteristics of the skeletons it leaves behind? Four tasks immediately suggest themselves.

First, we need to develop deeper insight into the likely sources of heterogeneity in frailty and the shape of the frailty distribution in real populations, work currently being pursued by epidemiologists, physiologists, and geneticists [Weiss 1990]. This research will need to pay due attention to the fact that frailty is multidimensional: different disease processes interact with each other and also with an individual's constitutional susceptibility to stress in determining frailty.

Second, we need to understand how a given frailty distribution is related to the distribution of risks of death among individuals. A considerable part of recent demographic theory has been devoted to this issue [Manton and Stallard 1984, 1988; Vaupel 1988, 1990], and we suspect that it may be resolved during the coming decade.

Third, we need a better understanding of the details of various pathological processes at the cell, tissue, and organ levels. Specific aspects of pathology that need to be addressed include how long it takes to develop particular hard-tissue lesions, how overall health affects the probability of lesion development, and how frailty and the risk of death vary by stage of lesion development, including the inactive or healed stage. This is a line of research best pursued by pathologists working with recent, well-documented biological materials and by epidemiologists.

These first three tasks are ones to which archaeologi- cal osteologists are unlikely to make fundamental contributions: they must remain consumers rather than producers of the relevant theory and observations. There is, however, a fourth task to which all anthropologists can contribute, namely, development of a better understanding of the role played by cultural context in de-

16. See, for example, the current controversy over the status of the !Kung [Solway and Lee 1990, Wilmsen and Denbow 1990].

17. Another problem, currently of major concern to epidemiologists and demographers but receiving only passing attention from osteologists, is that of distinguishing primary from contributory causes of death [Manton and Stallard 1984, Nam 1990].

18. Information on the course of disease and its relationship to mortality in the preantibiotic era would clearly be of great interest in this connection, but pertinent data are difficult to locate and interpret [Alter and Riley 1989].
terminating heterogeneous frailty and the level of selective mortality. For example, interpretation of the Oseota data is facilitated by the fact that these skeletons were drawn from a relatively simple tribal society. The cemetery, which was completely excavated, is comparatively homogeneous in composition and appears to have been used for a relatively short period of time, spanning decades rather than generations. There is no archaeological evidence suggesting that the sample represents a complex accumulation from distinct social groups with differential access to basic resources or exposure to pathogens, such as might occur in samples drawn from complex societies or accumulated over multiple generations when either fertility or mortality was changing. While these conditions almost certainly do not eliminate the problem of heterogeneous frailty entirely, they undoubtedly minimize its effects. The Dickson Mounds skeletal series, in contrast, is more difficult to interpret because of its cultural context. The cemetery is large and its internal organization is complex, consisting of multiple, superimposed mounds (Harn 1980). Much of the burial area remains unexcavated, increasing the likelihood that the available sample is selective. The skeletons from Dickson Mounds were accumulated over many generations spanning hundreds of years during a period of considerable cultural change. Thus, sample heterogeneity and selectivity are likely to be of extreme importance in their interpretation.

In sum, choosing among competing interpretations of the osteological evidence requires tight control over cultural context as well as a deeper understanding of the biology of frailty and death. These problems deserve far more attention than they have received to date if we are to make sense of the biomedical consequences of the major social and environmental changes that have occurred during the course of cultural evolution.

**Appendix**

Here we present a hazard model of mortality with a two-point distribution of frailty, based on a model originally developed by Vaupel and Yashin (1985a). We assume that the population of interest comprises two non-overlapping fractions, subpopulations 1 and 2. These subpopulations differ from each other in the level and age pattern of risk of death, but the individuals within each subpopulation have identical risks at each age. If \( h_1(t) \) and \( h_2(t) \) are the hazards of death at age \( t \) in subpopulations 1 and 2, respectively, then the mean hazard as a population as a whole is simply the weighted average,

\[
\bar{h}(t) = p(t)h_1(t) + (1 - p(t))h_2(t),
\]

where \( p(t) \) is the fraction of surviving individuals at age \( t \) who belong to subpopulation 1. The value of \( p(t) \) is determined by the initial mixture of subpopulations (i.e., by \( p(0) \)) and by \( S(t) \), the probability that an individual in the \( i \)th subpopulation \( i = 1, 2 \) survives from birth to age \( t \):

\[
p(t) = p(0)S_1(t)/(p(0)S_1(t) + (1 - p(0))S_2(t)) \tag{A.2}
\]

\[
p(t)S_1(t)/\bar{S}(t),
\]

where

\[
\bar{S}(t) = p(0)S_1(t) + (1 - p(0))S_2(t) \tag{A.3}
\]

is the mean survival from birth to age \( t \) in the population as a whole. Being a constant, \( p(0) \) can be written as \( p \) with no loss of generality.

Combining equations A.1 to A.3 and rearranging, we can write \( h_2(t) \) as a function of \( p \), \( \bar{h}(t) \), \( \bar{S}(t) \), \( h_1(t) \), and \( S_1(t) \):

\[
h_2(t) = [\bar{h}(t)\bar{S}(t) - ph_1(t)S_1(t)]/\bar{S}(t) - pS_1(t). \tag{A.4}
\]

Assuming that \( \bar{h}(t) \) is a two-parameter negative Gompertz function, then

\[
\bar{h}(t) = ae^{-bt} \quad \text{and} \quad \bar{S}(t) = \exp[-\alpha(t - e^{-bt})/\beta]. \tag{A.5}
\]

If we substitute these expressions in equation A.4 and provide empirical estimates of \( \alpha \) and \( \beta \), then \( h_2(t) \) is uniquely determined by \( p \) and \( h_1(t) \) since \( S_1(t) \) can be found from \( h_1(t) \) as

\[
S_1(t) = \exp\left[-\int_0^t h_1(y)dy\right]. \tag{A.6}
\]

Thus, if we specify a functional form for \( h_1(t) \) and a value for \( p \), we can immediately solve for \( h_2(t) \).

**Comments**

MARK N. COHEN

*Department of Anthropology, State University of New York, Plattsburgh, N.Y. 12901, U.S.A.*

Wood and colleagues raise a number of significant questions about whether cemetery populations can be presumed accurately to reflect the health status of the once-living populations from which they are derived. Specifically they challenge several of my conclusions (Cohen and Armelagos 1984, Cohen 1989) about the meaning of apparent trends in skeletal data related to the origins of agriculture and the rise of civilization. They suggest that my conclusions are possibly, but not necessarily, accurate because the skeletal data are amenable to more than one reading.

But those conclusions come from several sources of data, not just from skeletons. Skeletal data display an apparent increase in infection associated with sedentism and large population aggregates in many parts of the
world. This trend might be spurious, as Wood et al. suggest, but it is also predicted by epidemiological theory and demonstrated, repeatedly, in ethnographic comparisons and transitions. Similarly, tuberculosis not only blossoms with incipient urbanization in the archaeological record but also is clearly a disease of cities in the modern world and one which is considered unlikely to have been capable of propagation in a sparsely populated preagricultural ancient world.

The apparent increase in anemia in skeletal populations associated with farming, sedentism, and large population aggregates might also be spurious, but anemia is also a well-known consequence of the increasing parasitism and disease which predictably accompany larger and more sedentary groups in the modern world. Moreover, contemporary hunter-gatherers really do have low rates of anemia as well as low rates of protein, vitamin, and mineral deficiency. A similar argument can be made about the apparent increase in frequency of enamel hypoplasia of teeth associated with the Neolithic revolution. Hypoplasia is a general stress marker mostly for the infant and toddler years. One prominent cause of hypoplasia, weaning diarrhea, is, in fact, conspicuously rare or absent among contemporary hunter-gatherers even when common in neighboring farmers. Most diarrhea in fact is density- and sedentism-dependent and is associated with levels and kinds of malnutrition rarely seen among hunter-gatherers. Moreover, the major epidemic diseases, such as measles, which are also considered possible causes of hypoplasia arguably did not exist in the ancient world. It seems to me that, in the name of parsimony, the obvious interpretations of the skeletal trends in these cases are also the right ones.

On a more general level, I am concerned with the clear pro-state or pro-civilization bias which dominates both popular and scientific interpretations of history [and of current events]. Much of the work that is at stake in this discussion (Cohen 1989) has challenged this bias, and it disturbs me that the questions raised by Wood et al. emerge now, just as uncivilized life-styles and the meaning of our history are being reevaluated. In giving the questions raised by Wood et al. the attention they merit, I hope that we will remember to be equally skeptical of all conclusions from skeletal data [including those already embedded in our history books] and not just discard some of the recent revisions to standard history.

LES LIE E. EI S E N B E R G  
85 Curtis Place, Lynbrook, N.Y. 11563, U.S.A.
24 III 92

In this cautionary yet optimistic article, the authors address a number of issues fundamental to the reconstruction and interpretation of prehistoric health using skeletal series derived from archaeological contexts. While my comments focus primarily on the impact of this article on paleopathological studies, they should not be viewed as separate from related demographic considerations.

Paleopathologists and bioarchaeologists excavate and analyze human remains which are often incomplete and poorly preserved and may fall short of being representative of the population[s] from which they are derived. Even worse, skeletal lesions can be expected to develop only during periods of chronic illness, and of those not all are known to induce quantifiable or even recognizable bony changes. This thought-provoking article attempts to make explicit, and therefore more objective, the multiple factors which play into what the authors term “selective mortality” and which color all subsequent efforts to interpret lesion frequencies among survivors and nonsurvivors in a particular age-group. The notion of “frailty,” or susceptibility to stress, presented here is critical to a deeper understanding of both active and healed lesion frequencies and distributions and raises the larger and more complex question of just what observed skeletal lesion frequencies really mean.

The authors are in good company, because even as early as 1985, D. J. Ortner [personal communication] was convinced that the presence of healed skeletal lesions suggested that the individual manifesting such bony changes was immunologically capable of “fighting back.” As Wood et al. note, “the presence of an inactive lesion indicates survival of a disease process earlier in life and thus may signify an individual whose frailty is low compared with those who died at earlier ages.” Rather than being viewed as the “sick ones,” Wood et al. believe that “individuals with detectable lesions and early death are preferentially drawn from the advantaged group.” That is, the individuals with the healed bony lesions and the hypoplastic enamel defects are actually the healthier ones. I must confess that it has taken me a number of years to accept this argument, but I finally have; fact, it makes good inherent sense. What we are looking at when we analyze prehistoric skeletal populations and observe healed bony pathology is those who, in a sense, have “made it,” to succumb, perhaps, to a different stressor or group of stressors later in life. This is all the more reason to insist that an explicit analytical distinction be made between healed and active lesions; as intuitive as this may seem, some paleopathological studies still do not do it, leaving us with the impossible task of blindly teasing out the meaning hidden in the numbers.

While I am convinced that the notion of heterogeneous frailty is an extremely important one, I am considerably less optimistic than the authors that we will ever be able to develop an index of frailty for either individuals or populations. From their discussion, however, it is clear that we should not assume that hidden heterogeneity is normally distributed in all populations. It is by applying these concepts that we can begin to make sense of, or at least offer cogent competing hypotheses for explaining, observed bioarchaeological phenomena such as the economic, technological, and social changes inherent in the transition from a more mobile hunter-gatherer life-style to a more sedentary agricultural one.
That the transition may not have been accompanied by a deterioration in general health status is well worth examining in light of the issues raised here.

DALE L. HUTCHINSON
Department of Anthropology, University of Illinois, Urbana, Ill. 61801, U.S.A. 19 III 92

How well a skeletal sample represents the society from which it is drawn is a long-standing issue. Much of the controversy involves methods for estimating age and sex from skeletal remains [Bocquet-Appel and Masset 1982, 1985; Buikstra and Konigsberg 1985] and biases due to preservation differences [Walker, Johnson, and Lambert 1988]. There is general agreement among researchers that inadequate techniques for estimating age and sex could result in aberrant demographic reconstructions. Wood and his colleagues focus on another issue—inferring the health of a past population from frequencies of pathology preserved in its skeletal remains. As they point out, many of the problems encountered in estimating the prevalence of particular diseases are the same for skeletal and living populations; estimates of disease prevalence in both are dependent on the samples from which they are drawn. For example, it is not surprising that the prevalence of tuberculosis in contemporary U.S. populations is likely to be greater in a sample from a county health clinic subsidized by state funding than in a sample from a community health-management organization that is not.

The emphasis on populations rather than individuals in osteological studies during the past few decades has been an attempt to model studies of the dead on studies of the living. One question that Wood et al. raise is whether an aggregate is a population and how many subpopulations it contains. This question was a major focus in the Bocquet-Appel and Masset [1982, 1985] exchanges, and their concern with the effect of migration on demographic reconstructions remains a problem.

Estimates of disease prevalence in skeletal populations are further hindered by the unavailability of clinical data. Do skeletal lesions accurately reflect the prevalence of a disease in a living population, or are they representative of the survivors of a disease who lived long enough to manifest them? The reconstruction of disease prevalence in a past population from skeletal lesions is certainly problematic. Many of the lesions preserved in skeletal remains are nonspecific. Wood et al. discuss one such, enamel hypoplasia. Another, periosteal inflammation, may result from infectious disease or from minor injury. One group of related diseases, the treponematoses, may produce periosteal inflammation, but only some of the individuals who test positive serologically exhibit skeletal involvement [Grin 1956, Murray et al. 1956]. Furthermore, only a small percentage of the individuals with skeletal lesions develop lesions characteristic of treponemal infection (stellate cranial lesions, saber tibiae), usually in the secondary and tertiary stages of the disease. Consequently, the less severe and more general lesions, such as periostitis, could be attributed to other etiologies. In areas where the treponemal diseases are endemic, such as Africa, morbidity is high but mortality is low. Consequently, the individuals who exhibit skeletal lesions characteristic of treponematosis probably represent not all survivors of the disease but those who have had the disease longest. In this case, periosteal inflammations may actually underestimate the prevalence of treponematosis in a population. It is equally as plausible that nonspecific periosteal inflammations have other etiologies. This is aside from other health factors which might exacerbate the infections, such as nutrition.

That Wood et al. have left me thinking about other examples of hidden heterogeneity, selective mortality, and demographic nonstationarity demonstrates the utility of their discussion. Consideration of a suite of possible explanations for skeletal lesions is important for the development of more reliable interpretations of disease in past populations. As they point out, however, many of the tasks here are likely to fall to researchers other than anthropologists. There is still much to be learned about the interaction of disease, environment, and cultural factors in the overall epidemiology of particular diseases. Finally, while anthropologists can contribute more toward understanding of cultural context in the interpretation of past disease, factors such as skeletal preservation and basic demographic methods for skeletal materials will still play a role in accurate estimates of the health of past populations.

RIMANTAS JANKAUSKAS AND GINTAUTAS ČESNYS
Department of Anatomy, Histology, and Anthropology, Faculty of Medicine, Vilnius University, Ciurlionis 21/27, 2009 Vilnius, Lithuania. 6 I 92

This thought-provoking paper is of great interest for paleoanthropologists. While the amount of skeletal material is rapidly increasing, most studies of it remain on a rather simple empirical level, limiting themselves to the mere statement of the facts observed. There is a growing need for theoretical evaluation of these observations, and Wood et al.’s paper is a nice attempt to fill this demand. It is axiomatic although often forgotten by paleo-osteologists that skeletal series first of all represent dead people and this means that direct extrapolation from their data to a living population is problematic [higher frequencies of active lesions is one example]. The idea of hidden heterogeneity and selective mortality in the dynamics of death is another interesting and important suggestion. For hunter-gatherer and early agricultural societies it is perhaps more a theoretical possibility, but for populations with social stratification it is of increasing importance. The problem can be solved only through close collaboration with archaeologists. Another question of particular interest is the validity of skeletal markers of stress. We completely agree with the authors that there can be no simple way of interpreting
these markers. For example, short stature and periostitis incidence can be understood better when the levels of mortality are known. In our late-medieval materials, short life-span is correlated with shorter stature, but these samples of shorter stature (and lower level of sexual dimorphism) and shorter life-span have a higher incidence of healed periosteal lesions. According to the hypothesis suggested by Wood et al., this might mean that short stature, short life-span populations with high frequencies of periosteal lesions (in our case, medieval town populations) were healthier. This conclusion differs dramatically from traditional interpretations and calls for explanation.

This paper is an excellent presentation of the possibilities of interpretation of osteological materials that will serve as a basis for future discussion and a call for better understanding of cultural context as well as of the biology of pathological processes. It should encourage research collaboration between physical and cultural anthropologists.

M. ANNE KATZENBERG
Department of Archaeology, University of Calgary, Calgary, Alberta, Canada T2N 1N4, 23 III 92

Wood, Milner, Harpending, and Weiss have very eloquently pointed out that we still have a long way to go before we can speak with confidence about the health of prehistoric people. They have also delimited some problems that distort interpretations of prehistoric health from skeletal remains. While some of these problems have been brought forward in earlier works (e.g., McHenry 1968 on differing interpretations of Harris lines and Howell 1982 on paleodemographic reconstruction), this paper is unique in highlighting and questioning specific assumptions in paleodemography and paleopathology which influence larger-scale interpretations of health among earlier peoples such as those presented in the volume by Cohen and Armelagos (1984a). My comments focus on the problem of evaluating morbidity.

The emphasis on problem-oriented research in paleopathology, paralleling research in paleodemography, has resulted in a shift from describing individual lesions in detail to grouping lesions with similar appearances together and focusing on their frequencies. This population approach, starting most recently with the work of Armelagos (1969) and becoming widespread throughout the 1970s and 1980s, has resulted in a wealth of literature linking diet, settlement patterns, and health (e.g., Cohen and Armelagos 1984a, Rose and Rathbun 1987). Ortner (1991) has warned researchers of the dangers of making diagnoses without providing careful descriptions of bony lesions. He calls for a return to emphasis on detailed descriptions so that the findings of different investigators will be comparable. While this seems regressive, a more careful consideration of the causes of lesions would aid in interpreting what recovery with some scar on bone really means in terms of adaptation. Recent in-depth studies of conditions such as enamel hypoplasia (reviewed by Goodman and Rose 1990 and Skinner and Goodman 1992) and porotic hyperostosis (Stuart-Macadam 1985, 1987) bring us closer to understanding the implications of finding these lesions in prehistoric peoples.

A problem not mentioned in this work is that of differential preservation of skeletons based on age at death (see Walker, Johnson, and Lambert 1988). This compounds the problems pointed out by Wood and colleagues in that it is usually the very young and very old who are omitted from skeletal samples because of differential preservation.

It seems that the best approach we can take at this point is the one these authors suggest: to provide alternative explanations for the set of data at hand. Along the same lines as Ortner’s plea for better descriptions of skeletal lesions, this approach will allow others to work with the data and the possible scenarios when future studies provide additional insights into the interpretation of past populations.

JOHN R. LUKACS
Department of Anthropology, University of Oregon, Eugene, Ore. 97403, U.S.A. 9 III 92

In many ways this article addresses theoretical and practical concerns for interpreting the human skeletal record parallel to those expressed in the early 1980s by students of vertebrate taphonomy (Behrensmeyer and Hill 1980, Shipman 1980). The latter field of enquiry is essentially concerned with understanding the demographic characteristics of living populations the better to interpret the factors influencing the creation of vertebrate fossil assemblages. One wonders why it has taken ten long years for physical anthropologists to address similar complex taphonomic processes in human mortuary samples. Though many physical anthropologists are familiar with the works just cited, paleopathology apparently has borrowed these concerns not directly from paleontology but from demography. Nevertheless, the problematic issues of sampling bias and the complicating effects of distinct demographic patterns on the genesis of vertebrate faunal assemblages directly parallel concerns expressed here regarding the measurement and interpretation of pathology and vital statistics in prehistoric skeletal series.

Human osteologists and dental anthropologists should benefit from the cautionary tone of this paper, which serves the valuable purpose of questioning accepted methods for measurement and interpretation of health status in prehistoric populations. Since the authors rely primarily upon examples from prehistoric Native Americans and many readers may be unfamiliar with these and similar issues among South Asian populations, my comment will be devoted to additional examples from India and Pakistan that generally support their contentions.

The problems of hidden heterogeneity in frailty and sample selectivity are especially well illustrated in In-
dia, where society is clearly divided into distinct groups: the very wealthy and the very poor. Adequate nutrition, education, and access to good health care enable the wealthy to live a healthier and less stressed life in a culturally buffered environment. Their impoverished neighbors are routinely subjected to malnutrition, lack of education, more vectors of disease, and minimal or inadequate health care. This visible heterogeneity in affluence among the living translates into a culturally induced variability in frailty that must be considered simultaneously with the genetic variation in frailty that exists within both wealthy and poor segments of Indian society. An individual of “medium frailty” might survive in a wealthy environment but perish in a poor one. Variations in affluence and frailty are not always clearly discernible in prehistoric mortuary collections.

Among the Bronze Age Harappans the interpretive difficulties raised by hidden heterogeneity and selectivity are numerous and not easily resolved. Aggregate prevalence rates for dental diseases for the entire Harappan skeletal sample undoubtedly combine the prevalences for more than one subpopulation at this 3rd-millennium urban site [Lukacs 1992]. The series is not large enough to permit subdivision into socioeconomic groups on the basis of funerary goods or ornamentation of body ornaments, which would be one approach to identifying the prevalence of pathological lesions in subgroups based on affluence. Alternatively, the human skeletal samples from Mesolithic sites of the Gangetic Plains are much more likely to be culturally homogeneous than either modern Indian society or Harappan society and therefore less susceptible to these interpretive dilemmas [Lukacs and Hemphill 1992, Lukacs and Pal 1992].

The interpretation of linear enamel hypoplasia data on living Indians and prehistoric Harappans is an excellent example of the complexities that Wood et al. discuss. At the most elementary level, the genders could be considered as subgroups which occur in all human populations. In the Harappan skeletal sample as whole, linear enamel hypoplasia prevalence is 72.2%, but significant differences are evident between the sexes both in this feature [males 52%, females 93%] and in dental caries [Lukacs 1992, Hemphill, Lukacs, and Kennedy 1991]. These results have been interpreted to suggest that females were subjected to greater stress than males during childhood and consumed a more cariogenic diet than males as adults. While the mean age of linear enamel hypoplasia formation is not significantly different for males and females, females are affected at younger and older ages and have more events, or “growth disruptions,” than males. These observations go beyond mere subgroup prevalence rates to suggest that females were ill or malnourished more often than males. Among the select sample of Harappan skeletons available for analysis, females appear to have been more often affected by factors causing linear enamel hypoplasia than males and may be considered frail, or less well-buffered against those factors, in this respect. Alternatively, males may be envisioned as less frail with regard to those factors but, in that they are part of a mortuary sample, frail in other respects. These sex differences in linear enamel hypoplasia among Bronze Age Harappans have been interpreted to reflect the culturally sanctioned differential treatment of children commonly known as “son preference” or “daughter neglect” [Miller 1985; see also Lukacs and Joshi 1992].

Using natural rather than clinical samples, our investigation of three living ethnic groups of northwestern India was expected to show that female Rajputs, members of a caste group known to practice son preference and daughter neglect, displayed a higher incidence of linear enamel hypoplasia than male Rajputs, female tribals, or low-caste females [Lukacs and Joshi 1992]. Gender differences proved not significant, however, and Rajput women had the lowest prevalence of the three groups of females studied. These findings suggest that the osteological paradox has its corollary in studies of the living. In the absence of other reasonable explanations, the hypothesis was advanced that if the mortuary sample of Rajput females could have been analyzed, a higher prevalence of linear enamel hypoplasia would have been found.

Human osteologists and dental anthropologists whose research encompasses both living and prehistoric populations are routinely confronted by the issues raised in this paper, and because of the dramatic differences in the composition of their study samples (living vs. dead) many of these investigators are especially cognizant of the problems of hidden heterogeneity in frailty and selectivity in skeletal samples. While being aware of these problems does not guarantee that workable solutions will be immediately forthcoming, it does demand that conclusions regarding the health status of prehistoric human skeletal series be carefully thought out and advanced with considerable caution. The authors of “The Osteological Paradox” should be commended for bringing these problems to the attention of a wider audience. May their discussion generate a higher level of achievement in paleodemography and paleopathology!

JANET W. MCGRATH
Department of Anthropology, Case Western Reserve University, Cleveland, Ohio 44106-7125, U.S.A.
21 III 92

“Better health makes for worse skeletons” is a statement with complex and interesting implications for the study of ancient disease. Wood et al. present a well-reasoned article that points to what now seems obvious: since skeletal lesions develop as the result of a long-term disease process, individuals with skeletal lesions are likely to have been rather healthy, having survived long enough to develop them. The potential reinterpretation of paleopathological data in light of this discussion is exciting to anticipate.

I confess that I have always had trouble interpreting the data suggesting that the transition to agriculture reduced health. There is compelling evidence that the sedentism associated with agriculture increased exposure
to pathogens, thereby potentially increasing the risk of death. Many of these pathogens, however, cause acute, infectious disease which would not result in skeletal lesions, and therefore this increase in mortality would not be reflected in the skeleton. Discussion has focused instead on change in diet, and it has always been difficult to explain to students why humans turned to agriculture under such conditions. As Wood et al. point out, the work of Armelagos and Goodman has been exemplary here in offering multiple hypotheses to explain the archeological record, and they clearly demonstrate yet another set of explanations.

The next important step, as Wood et al. indicate, will be the clarification of patterns of frailty. If one aspect of frailty is the immunologic status of the individual, which is influenced by both diet and infection, then frailty might have been affected by the transition to agriculture and it could be argued that the transition increased mortality through increased frailty. This, however, leaves unanswered the question why specific skeletal lesions and skeletal markers of stress (e.g., hypoplasia and Harris lines) are reported in higher frequencies in populations in transition to agriculture. Perhaps frailer individuals produce bone lesions more quickly? We will not know, as Wood et al. point out, until we understand more about the way in which frailty, risk of death, and lesion development interact.

Although they address disease in ancient populations only, the question of “hidden heterogeneity of risks” is relevant for studies of living populations as well. For example, if the case fatality rate for a particular disease is equal to the number dying of the disease/the number with the disease, then differences in frailty in the sick population will mean that not everyone with the disease is at equal risk of dying. Differences in frailty are not considered in the calculation of case fatality rates, mostly because of the difficulty of identifying the distribution of frailty. If we examine behavioral risks for people who die of this disease, we may erroneously conclude that a particular behavior or set of behaviors carries a higher risk of mortality when the differences are really due to differences in frailty.

Wood et al. deserve praise for the scientific rigor of their multiple-hypothesis approach. In interpreting data on prehistoric populations it is critical to walk a fine line between pat answers that simply reflect the biases of the investigators and complete dismissal of the field of research. These authors do this well by reminding us of the complexity of the matter while retaining enthusiasm for the effort of learning about health and disease in human populations.

In what is inevitably to become the “Farewell to Paleodemography” [Bocquet-Appel and Masset 1982] of the 1990s, this paper raises a number of important, far-reaching points concerning morbidity and mortality analyses of prehistoric skeletal populations. In contrast to the aforementioned work, it adopts a positive approach to the problems it addresses. Responses to Bocquet-Appel and Masset’s attack on paleodemography (Buikstra and Konigsberg 1985, Van Gerven and Armelagos 1983) constituted critical advances in a field which had been largely spared self-criticism if not external critiques [cf. Petersen 1975, Howell 1982]. I hope that research responding to the issues raised in this study will further solidify paleodemography’s status as a valid, albeit specialized subfield of human demography.

I concur with the authors, however, that the points raised here are far more difficult than those cited by Bocquet-Appel and Masset. While selectivity and demographic nonstationarity are certainly nontrivial matters, recent research has addressed them [cf. Sattenspiel and Harpending 1983, Saunders and Katzenberg n.d.]. In my opinion, the most bedeviling of the basic conceptual problems raised by the authors is that of hidden heterogeneity. While they effectively use hazards analysis to define and describe it, they also point to the *identification problem* associated with testing the goodness-of-fit of these models to actual data. As they note, one way to test for goodness-of-fit is to model frailty distributions on known biological patterns, but at the same time they detail examples, both contemporary (e.g., diarrheal disease) and prehistoric (e.g., cribra orbitalia, enamel hypoplasia), in which counterintuitive mortality distributions are either apparent or plausible. Problems of interpretation of the resulting ill fit between such data and models based on biological patterns can only be exacerbated in prehistoric skeletal analyses, given the compounding problems of demographic nonstationarity and/or selectivity bias. To my mind, this is the crucial point where all three problems converge.

Highlighting these problems does not mean that paleodemography should be abandoned. Instead, Wood et al. are to be congratulated for bringing them to the attention of researchers. Consideration of them, while certainly difficult, can only improve paleodemography.

ERI C A B E L L A R O TH
Department of Anthropology, P.O. Box 3050,
University of Victoria, Victoria, B.C., Canada
V8W 3P5. 6 III 92

In what is inevitably to become the “Farewell to Paleodemography” [Bocquet-Appel and Masset 1982] of the 1990s, this paper raises a number of important, far-reaching points concerning morbidity and mortality analyses of prehistoric skeletal populations. In contrast to the aforementioned work, it adopts a positive approach to the problems it addresses. Responses to Bocquet-Appel and Masset’s attack on paleodemography (Buikstra and Konigsberg 1985, Van Gerven and Armelagos 1983) constituted critical advances in a field which had been largely spared self-criticism if not external critiques [cf. Petersen 1975, Howell 1982]. I hope that research responding to the issues raised in this study will further solidify paleodemography’s status as a valid, albeit specialized subfield of human demography.

DOUGLAS H. UBE LA KER
Department of Anthropology, National Museum of Natural History, Smithsonian Institution,
Washington, D.C. 20560, U.S.A. 11 III 92

Wood et al. raise key issues in the interpretation of demographic and paleopathological data gleaned from the analysis of human skeletal remains. One of the issues is related to sampling. We have long known that sampling problems plague our interpretations of past human biology. If we have no way of knowing from archeological or other sources what inadequacies or complexities exist within the samples used for our studies, then we have little opportunity to overcome them or to incorporate them into a more sophisticated research design. I would agree that selective mortality and hidden heterogeneity
are potential sources of error, but it is important to remember that sampling problems do not preclude the formulation of general statements about the disease experience, mortality, and/or fertility of the population represented.

If interpreted properly, skeletal lesions usually indicate that the individual survived a disease experience or stress episode, and that individual is obviously healthier than someone who died from the same disease experience without its having had an opportunity to leave its mark on the skeleton. As the authors point out, frequencies of skeletal lesions cannot be equated with disease frequencies among the living. Similarly, we have learned that life tables developed from mortality data can be influenced by changes in fertility, migration, and other factors that remain largely unknown to the investigator. Elucidation of the dynamic relationship among skeletal evidence of disease in mortuary samples, the distribution of ages at death, disease frequency among the living represented by the skeletal sample, the demographic structure, and the health of the living populations remains a worthwhile but elusive goal of our research. By raising issues and clarifying assumptions of the research process, this article may enable us to advance.

RICHARD G. WILKINSON
Department of Anthropology, University at Albany, Albany, N.Y. 12222, U.S.A. 10 III 92

Wood and colleagues offer practitioners of paleopopulation [re]construction the provocative proposition that “better health makes for worse skeletons.” We are asked to accept the idea that people with lesions—who are, as the authors note, indisputably dead—were healthier than those without lesions (who, by virtue of their inclusion in our samples, are equally dead). To the extent that skeletal lesions may represent survived stress and the absence of lesions may denote acute, lethal stress, their point is well taken. As they note, it has serious repercussions on our interpretations of the health of prehistoric populations.

The concept of differential morbidity (and mortality), expressed in terms of heterogeneity of frailty, may not be as broadly applicable as Wood et al. imply, and the concept[s] of health could have been clarified. A lesion-free skeleton of 40 years at death should not suggest to us the same thing as a lesion-free infant skeleton in terms of relative health, frailty, or immunological competence. Aggregated age-at-death data may be interpretable as Wood et al. suggest, but the interpretation seems unlikely to be applicable to all individuals; some cautionary statement about age-specific application is in order, especially since the examples given center on 0–5-year age-groups. Pathological conditions also need to be specified, as “better health makes for worse skeletons” hardly applies to treponemal diseases, leprosy, and the like.

The implications of this study for the analysis of stature and the use of stature as a measure of adaptability (or as a response to stress) are interesting and important. The well-known stature reduction in prehistoric Mesoamerica, or even the unexpected differences in stature between status groups at Monte Albán, where the seemingly higher-status individuals have shorter stature (Wilkinson and Norelli 1981), may be nicely explained by differential mortality, but there is an opposite relationship among the prehistoric Maya (Haviland 1967) and at Illinois Hopewell (Buikstra 1976). The relationship between stature and mortality could be investigated with relatively bountiful historical data. Stature data on military conscripts could be compared with male mortality during the past 150 years in the United States and an even longer period in Europe.

The bottom line of the pessimistic prognosticators from Pennsylvania is a serious one. They note that “populations cannot be compared meaningfully if their distributions of frailty differ in unknown ways.” Since heterogeneity of frailty is very likely and the distribution of frailty in prehistoric populations may be unknowable, we are in a pickle. The concept of “identification problems” is a critically important one, and this is, as the authors note, an area much in need of refinement. We too often accept a skeletal series as a unified entity rather than as a collection of subgroups. Many of our large, complex skeletal series need to be re-investigated for the existence of such subdivisions. Subdivisions with genetic bases should be identifiable through gross anatomical comparisons and/or molecular methods. Subdivisions of a more genetically homogeneous population which differ in their relative frailty, however, would seem exceedingly difficult to identify. The existence of a group of weaning-age skeletons may suggest increased frailty relative to weaning diarrhea, but this same group also provides evidence—by its very existence—of low frailty relative to neonatal and infant mortality. A low frequency of lesions (from, say, porotic hyperostosis or enamel hypoplasia) in this group informs us only that members avoided anemia and growth arrests earlier.

In the section “Is There Hope?” we are presented with a litany of confounding problems and some excellent suggestions for advancing our science, specifically in the area of research into questions of the heterogeneity of frailty and its relationship with distributions of deaths. The call for increased pathological research on hard-tissue lesions is also well reasoned, but it remains to be seen whether specialists outside of anthropology and paleopathology can be coerced into dealing with dry bone. Finally, we need to understand the “role played by cultural context in determining heterogeneous frailty and the level of selective mortality.” There are “good” skeletal series from short-term cemeteries, utilized by a single cultural unit itself characterized by relative equality and, ideally, no migration to muddy the analytical waters. Then there are sites such as Dickson Mounds, representing relatively long temporal spans and a good deal of cultural complexity, both synchronically and diachronically. We do need to concentrate on the “simple” sites before we can approach the issue of the heterogeneity of frailty in the complex ones, but we cannot aban-
don the latter, which are the rule and not the exception. And at some level, these complex sites are more interesting in that they offer tantalizing evidence of biocultural change and of the interrelationships among populations. It now seems unlikely that “additional excavation” can be pointed to as a potential solution to data-related problems, and we must continue to make advances with the material at hand.

Reply

Department of Anthropology, Pennsylvania State University, University Park, Pa. 16802, U.S.A.
27 IV 92

We thank all the respondents for their thoughtful comments. While we cannot pretend that the problems we have flagged are easy to solve—much less that we have solved them—it is encouraging to see how many of our readers are willing to give them the serious consideration they deserve. Most commentators recognize that we are trying to be constructive rather than simply bidding farewell to paleodemography and paleopathology. And most realize that we are urging consideration of multiple hypotheses rather than taking sides on difficult and complex issues. Because the comments are generally positive, we confine our response to a few points that deserve amplification or clarification.

Eisenberg and Hutchinson point out that, in addition to the issues we raise, sensitivity is a special concern for paleopathology. The sensitivity of a diagnostic criterion [e.g., presence of a particular kind of skeletal lesion] is the proportion of truly diseased individuals in a sample who are identified as diseased according to that criterion. As Eisenberg and Hutchinson emphasize, most skeletal lesions are insensitive indicators of their associated disease processes in that they typically develop in only a small fraction of cases [e.g., advanced, chronic cases of tuberculosis and treponemal infections]. We agree that sensitivity is a serious problem for paleopathology but did not focus upon it simply because the issue has already been widely discussed in the field. It is important, however, to be cautious in responding to the sensitivity problem. For example, if we were to find a few cases of a particular lesion known to have low sensitivity, we might be tempted to infer that the disease causing the lesion must have been very widespread among the living. This inference is problematic for three reasons. First, as Hutchinson notes, most skeletal lesions are not only insensitive but also of low specificity. That is, many distinct pathological conditions induce lesions that are morphologically similar, making accurate diagnosis difficult or impossible. Second, absence of a particular type of lesion in a skeletal sample cannot be taken as evidence that the associated disease was not widespread. Since skeletal lesions are typically rare outcomes and skeletal samples are always small [at least by epidemiological standards], the probability of including a lesion in the sample is slim even if the associated disease was common in the living population. Third, selective mortality may add cases with lesions to the skeletal sample at a rate that overrepresents their prevalence in the living population. This last point is the one stressed in our paper, in large part because it has not been emphasized enough in the existing paleopathological literature. We are left in a quandary. Does the frequency of skeletal lesions underestimate disease prevalence among the living because of the sensitivity problem or overestimate it because of the selectivity and specificity problems? At this stage, it is anyone’s guess. But we can say this with confidence: disease prevalence and skeletal-lesion frequency are two very different things and, as Ubelaker and Hutchinson note, must never be equated.

One point should be reiterated for clarity’s sake. The presence of a particular kind of skeletal lesion is evidence for either persistence until death of the pathological process responsible for it or survival long after a disease episode ended. It is more difficult to interpret the absence of such lesions in ancient skeletons. Their absence may indicate either that the causal condition was not present or that early death occurred before a distinctive skeletal response had time to develop. The first interpretation is the traditional explanation, while the second is the alternative explored in our paper. We do not mean to suggest that the traditional explanation must always be wrong and the alternative correct; the point is that it is generally impossible to tell.

Eisenberg suggests that we may be overly optimistic about the possibility of developing an “index of frailty for either individuals or populations.” On the contrary, we share her pessimism. Still, we think it is worth trying to measure frailty, as long as it is acknowledged just how difficult a task it is. We also think an alternative approach is more likely to yield results in the short run. This approach involves modeling the distribution of frailty in a population, preferably on the basis of sound biological principles but at the very least using flexible statistical distributions that will fit a wide variety of cases [see Weiss 1990]. As we detail elsewhere [Wood et al. 1992], such models can then be used to adjust statistically for differing frailty distributions when comparing populations, even in the absence of individual-level measures of frailty. However, in view of the identification problem that Roth and Wilkinson so rightly underscore, it will still be necessary to test as many different specifications of the frailty distribution as possible.

Lukacs and McGrath make a point of fundamental importance; hidden heterogeneity and selectivity are not just problems for osteological research but occur in studies of living populations as well. McGrath’s example of the case fatality rate is especially apt in this regard. We would only add that at least a substantial minority of researchers in demography and epidemiology are already struggling with these same problems [e.g., Keyfitz and
Littman 1980, Heckman and Singer 1984a, Trussell and Richards 1985, Vaupel and Yashin 1985b, Manton and Stallard 1988, Gage 1989). There is one important respect in which studies of the living have an advantage: when dealing with living subjects, it may be possible to “uncover” some of the heterogeneity by actually measuring variables associated with frailty. McGrath points to immunocompetence as a major component of frailty. While immunocompetence cannot be measured on the dead, there are new procedures that allow it to be measured on the living, even in remote field settings (Shell 1992). In general, the most satisfying solution to the hidden heterogeneity problem is to convert as much of the variation as possible from the “hidden” to the “measured” category. Unfortunately, it is precisely this approach that is so difficult, perhaps impossible, to implement in archaeological studies. Still, the problem cannot be solved by ignoring it.

Wilkinson suggests that we place too little emphasis on the information provided by the age distribution of skeletal lesions. He is entirely correct that consideration of the age pattern of lesion frequencies is an essential part of any osteological interpretation. We would only point out that an examination of age specificity was the very foundation of our interpretation of cribra orbitalia, porotic hyperostosis, and periostitis in the Oneota sample. It is equally essential, when trying to make sense of the age distribution of various lesions, to distinguish insofar as possible between hard-tissue responses that were active at the time of death and those that were healed. While recognizing that it is often difficult to differentiate active and inactive lesions when dealing with dry bone, we concur with Eisenberg that it is important to try. We further suggest that well-remodeled skeletal lesions—evidence of events occurring long before death—are more informative about the disease experience of past peoples than has previously been realized. Using the Oneota skeletons, we show that particular kinds of healed skeletal lesions may indicate groups of individuals who experienced an elevated risk of death relative to their contemporaries, while others may mark individuals of comparatively low frailty. By examining the age patterns of healed and active lesions and by comparing those patterns across lesions of differing etiologies, we can begin to link pathological changes in bone to frailty in a meaningful way. For this reason, osteologists should never attempt to draw inferences from skeletal lesions considered individually.

Jankauskas and Cesnys make the interesting observation that heterogeneity and selectivity are likely to become more important as social complexity increases. Doubtless they are right, as suggested in our brief remarks on the archaeological contexts of the Norris Farms (Oneota) and Dickson Mounds sites. We must not take it for granted, however, that simple, small-scale societies never contain significant variation in frailty. One important lesson of the last 20 years of research in population genetics is that even small, localized, inbred human populations retain a considerable amount of genetic variation, some portion of which may have important phenotypic effects. Thus, biological variation may be present even when social variation is minimal. Nonetheless, Jankauskas and Cesnys are correct in stating that social complexity only makes the heterogeneity-selectivity problem worse. In light of this fact, Wilkinson’s advice that we concentrate on the simple sites before tackling the complicated ones is well worth considering.

Several of the commentators remind us that it is important, when considering the issues we raise, not to lose sight of other problems that have plagued osteologists for many years. Katzenberg and Hutchinson point to the differential preservation of skeletons by age at death, while Jankauskas, Cesnys, and Lukacs highlight the need for detailed attention to the cultural contexts from which skeletal samples are drawn. Roth and Ubelaker emphasize the untoward effects of demographic nonstationarity on paleodemographic analyses, and Katzenberg stresses the need to learn more about the pathological processes that result in hard-tissue lesions. We agree that all these concerns are fundamentally important and anticipate that osteologists will continue to wrestle with these and other issues that confound the interpretation of ancient bones. Such wrestling is an absolute prerequisite for continued scientific progress.

Cohen apparently regards our comments about the origins of agriculture as a “challenge” to his own position. This reaction is disappointing, because we had no wish to become embroiled in the debate—however popular and politically charged—over whether agriculture and the state were good or bad developments in human history. As we tried to make clear, the point “is not that we are right and other authors are wrong” about the effects of agriculture but that multiple hypotheses need to be entertained and means of distinguishing among them found. To a large degree, Cohen does precisely what we advocate. He turns to other kinds of evidence—demographic, biomedical, and epidemiological—to supplement the skeletal remains. It is of great importance, however, that such material be used critically and with extreme care, since it is subject to many of the same biases and difficulties of interpretation discussed in our paper, as well as additional confounding problems of its own.

Cohen points out, for example, that there are many reports, both archaeological and ethnographic, of deteriorating health following the settling of formerly nomadic groups. Those reports cannot be dismissed out of hand, although we note that the effects of sedentism on morbidity and mortality in living populations are more variable than Cohen acknowledges [see, e.g., Roth 1985]. More fundamental, however, we suggest that two entirely different meanings of the word “health” are being confused. One meaning of health is something that adversely affects an individual’s relative risk of death. In conformity with what has become standard demographic usage, we call this form of health “frailty.” A newborn who experiences an unusually high probability of dying during the first year of life is thus “frail.” Another meaning of health is general physical condition, which if poor might be termed “decrepitude.” A thin and anemic child with chronic skin lesions and respira-
tory disease is "decrepit." We assume, with Cohen, that physical decrepitude can leave signatures in the skeleton and dentition, especially if the illness is survived [e.g., enamel hypoplasia] or develops into a chronic condition [e.g., skeletal tuberculosis].

One fundamental point that we have tried to make in this paper is that these two meanings of health need have no simple relationship to each other. We suggest that in many groups with poor nutritional status—perhaps including many foragers—children are so frail that they do not have sufficient reserves to support chronic decrepitude, which takes time to develop. Infant and childhood death, when it occurs, happens quickly without leaving any hard-tissue signature. With more food, in contrast, children are capable of sustaining themselves in a decrepit state—a statement that seems paradoxical only if decrepitude and frailty are confused. If decrepitude lasts long enough before death, then markers of this state could show up in the archaeological record. But actual hazards of death, which are proportional to frailty, are lower in such a situation.

For example, Harpending and Wandsnider [1982] have described !Kung Bushmen living in settlements around Ghanzi in Botswana. The children in these settled groups were noticeably decrepit compared with their nomadic counterparts in the bush (Truswell and Hansen 1976): many were thin and inactive, and respiratory disease and skin and eye infections were common. Yet the estimated infant mortality rate in that undeniably decrepit population was half that observed in the bush, and mortality before age 15 years was one-third that in the bush. The sedentary children were apparently more decrepit but less frail, despite the fact that they had no better access to effective medical care than did nomadic children. The distinction between frailty and decrepitude helps explain a paradox first noted by Howell [1979]: nomadic !Kung appear to be in good condition but have mortality rates at all ages that are substantially higher than those of surrounding settled groups.

Pennington (1992, Pennington and Harpending 1992), using data reported by Howell [1979], finds the same response to sedentization in Ngamiland. Herero pastoralists entered !Kung areas in Ngamiland in the mid-1950s, and many !Kung settled at Herero cattleposts, where milk and other foods were widely available. The result of settlement, as Pennington shows, was a drop in infant mortality of about 50%. The largest decline was in the age-interval 1–4 years, corresponding to the usual age at weaning, when mortality dropped to one-fourth its level before arrival of the Herero. Not only were these !Kung newly sedentary, thus meeting the conditions of Cohen's model, but they were living in close contact with cattle, a source of zoonoses such as tuberculosis. An archaeological osteologist of the future might well find much more skeletal evidence of decrepitude following settlement and yet have no way to infer that this increased decrepitude was accompanied by a dramatic decline in early childhood mortality. This is particularly true if the age composition of the sample is ignored, as happens all too often in paleopathological studies.

By presenting this example, we do not mean to imply that sedentization is always accompanied by a decline in mortality, for evidence from other areas suggests that the opposite sometimes occurs. In general, we consider it extremely unlikely that sedentization always has simple, predictable effects on mortality or, for that matter, fertility. In fact, we would go farther and suggest that the transition to settled life differs in important respects from one location to another. Among the Northern Cree, settlement led to a rapid increase in fertility, but only because the Canadian government supplied formula and milk, thus shortening the duration of lactational infecundability (Romaniuk 1981). In other locations, permanent settlement may result in a marked and immediate rise in mortality, but only because certain infectious diseases such as cholera happen to be epidemic in the region at the time of settlement. Very possibly, the development of settled agriculture in the Old World, especially where domesticated animals and nonhuman milk were available, had different consequences for fertility and mortality than the parallel process in the New World. We will never know unless we reject the claim that a uniform response always occurs and its direction is self-evident from existing data. In sum, it is unrealistic to think that the transition from foraging to agriculture must always have had the same demographic and biomedical effects. Indeed, we believe that future research should abandon the concept of "sedentization" as a single process with uniform biological effects and concentrate instead on the specific mechanisms governing demography and health in each population.

All the problems we have raised can be summed up very simply: data do not speak for themselves. This is an elementary lesson that has never had the impact it deserves. The information contained in empirical data can be extracted and interpreted only when specific models are applied to them. To borrow a wonderful aphorism from Nathan Keyfitz [1975], "No model, no understanding." In one form or another, everyone uses models, even if only vague and nonmathematical ones, when thinking about data. All too often, however, these models are implicit and thus unavailable for critical evaluation. It has seemed so obvious that early ages at death imply high mortality or that diseased skeletons imply diseased populations that little thought has been given to the assumptions involved in these inferences. We have tried to show that even very simple alternative models can lead to conclusions that are the exact opposite of our immediate intuitions. Of course, as several of our commentators point out, this problem of overemphasis on data at the expense of theory and analysis is by no means restricted to archaeological osteology. Papers similar to ours could have been written for demography, epidemiology, genetics, paleoanthropology, evolutionary ecology, or any other field of empirical research. But that does not lessen the importance of the message for paleodemography and paleopathology. The challenge now is for researchers to come up with reasonable biological models that will help us make sense of the way skeletal samples are formed and, hence, what they mean.
References Cited


tion. Identification of pathological conditions in human skeletal remains. Smithsonian Institution Contributions to Anthropology 28.


