NBER WORKING PAPER SERIES

LIVE LONG, LIVE WELL: QUANTIFYING THE HEALTH OF HETEROGENEOUS POPULATIONS

John Mullahy

Working Paper 7895
http://www.nber.org/papers/w7895

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2000

Live Long, Live Well: Quantifying the Health of Heterogeneous Populations
John Mullahy
NBER Working Paper No. 7895
September 2000
JEL No. I1

ABSTRACT

Various health-, quality-, and disability-adjusted life year or life expectancy (HALY, QALY, DALY; HALE, QALE, DALE) measures have become gold standards for defining outcomes in technology evaluation, population health monitoring, and other evaluative efforts. As such, it is critical that the analytical framework within which these measures are used for descriptive and evaluative purposes be theoretically consistent and statistically rigorous. For instance, widely-accepted definitions of cost-effectiveness ratios and other technology evaluation criteria that are based on expectations of the respective cost and outcome measures must as such be defined in terms of expected HALYs or QALYs. Similarly, measures like HALEs or QALEs used for population health monitoring are typically concerned with population expectations of such measures (or their corresponding totals). This paper demonstrates that estimation of such expectations necessarily requires consideration of the population variation in and covariation between quality and longevity. From the perspective of several different environments characterizing such heterogeneity, quantification or estimation of measures like QUALs are recondidered. An empirical example of the central issues is provided by means of an analysis of the Years of Healthy Life (YHL) measure drawn from the U.S. National Health Interview Survey.

John Mullahy
Departments of Preventive Medicine and Economics
787 WARF, 610 Walnut St.
University of Wisconsin-Madison
Madison, WI 53705
and NBER
jmullahy@facstaff.wisc.edu

If thou wouldst live long, live well.

Benj. Franklin,
Poor Richard's Almanack, 1739

I. INTRODUCTION

For settings as diverse as large countries, memberships in sizable HMOs, small clinical trials, and local policy interventions, one finds increasing interest in quantifying the health status of the "population" of concern. Kindig, 1997, argues compellingly that for many practical purposes it is useful to be able to reference a scalar summary measure of the health of a population at a point in time, much like GDP or CPI measures characterize the output or price levels of a nation's economy at points in time. A recent report of the Institute of Medicine in the U.S. (Institute of Medicine, 1998), noted that "The development and application of summary measures of population health present complex and intriguing methodological, ethical, and political challenges." The consideration of some central conceptual and empirical aspects of the pursuit of summary measures of population health is the main purpose of this paper.

Some Recent Context

Recognition of heterogeneity in the health of broad populations' members has become widespread, with some consequent impetus for policy intervention (see Shepard and Zeckhauser, 1982, for an early theoretical treatment). With the stated objective of eliminating "disparities in six areas of health status experienced by racial and ethnic minority populations while continuing the progress we have made in improving the overall health of the American people," the U.S. National Institutes of Health have launched a formal Program to Address Health Disparities. The NIH definition of "health disparities" is "differences in the incidence, prevalence, mortality, and

burden of diseases and other adverse health conditions that exist among specific population groups." Of course, "disparities" amounts to the same thing as heterogeneity; the latter term is preferred here given its neutral rather than negative connotation.

In assessing the performance of the world's health care systems in delivering health product, WHO's recent *World Health Report 2000* recognized that summarizing the health of heterogeneous populations in a single measure is problematic. Among other things, the WHO report notes that both "the overall level of health" as well as "the distribution of health in the population" must be measured to be able to assess the objectives of any given health system. Moreover, the WHO report recognizes that for any given individual in a population health itself has a multiattribute character, and thus proceeds to characterize health status in terms of disability-adjusted life years/expectancy.

In a similar vein, the much-publicized *Healthy People 2010* initiative of the U.S. Department of Health and Human Services is designed to achieve two overarching goals: (1) increase quality and years of healthy life; and (2) eliminate health disparities among different segments of the population. While interesting (and controversial; see Kenkel, 2000, for instance) in their own right, these HP2010 goals jointly serve to highlight key aspects of the subsequent analysis. Specifically, the main concern here is with quantification of health measures that have a multiattribute character -- e.g. goal #1, regarding longevity ("live long") and quality of life ("live well") -- *and* that are simultaneously distributed heterogeneously in the population (e.g. goal #2, regarding disparities of health status in the population).[1] Indeed, a recent report by the U.S.

[1] Aggregate measures like the "Years of Healthy Life" (YHL) (continued)

National Institute on Aging (NIA, 2000) confirms that Americans are living both longer and healthier lives than in past generations, as disabilities have become relatively less common among older Americans over time.

It should be emphasized that the conceptual and analytical frameworks presented here are applicable not just in situations where quantification of the health status of large populations is of concern but also are suitable when quantification or estimation of the health of more narrowly-defined "populations" is the objective. A leading example of such narrower populations would be the treatment and control groups in a clinical trial within which one outcome of interest might be health status measured in some multiattribute manner -- for instance, quality of life and survival -- like the measures considered below (e.g. Lamas et al., 1998; Hlatky et al., 1997). To be sure, measurement of inherently multiattribute health status has attracted increasing attention in the clinical literature (Testa and Simonson, 1996; Wright and Weinstein, 1998).

The Issues and Plan for the Paper

A common feature of the work described above is its reliance on some summary measure or measures of the health status of the population in question. Such efforts must thus confront directly the issue of how to "map" from a distribution of health that in almost any interesting exercise will be heterogeneous in this population and may also be multiattribute in its character *into* a summary (scalar) measure of the health of this population (see Wolfson, 1999, for an excellent and comprehensive discussion).

---

measure (Erickson et al., 1995) have been designed to monitor in the aggregate these objectives in the United States (the YHL measure will the the focus of the empirical exploration undertaken in section V).

This paper sets out to develop an analytical framework for characterizing such summary measures and for assessing some properties of empirical strategies used to estimate or quantify these measures.

The roadmap for the remainder of the paper is as follows. Section II presents some fundamental conceptual and measurement issues, addressing the first-order question of what a summary measure of a population's health might entail when the health status of a population's members is simultaneously multiattribute in scope and heterogeneous in its distribution across the population. Section III considers from an analytical perspective the implications of quantifying population health when its multiattribute constituents are both heterogeneous in a population *and* may themselves covary across this population. It is suggested here that the concept of a statistical functional provides a conceptually useful typology for quantification of a population's health status. Related discussion on univariate and multivariate stochastic dominance then points the way toward more practical implementation of health measures based on low-order moments, with particular focus on population health measures characterized by means or conditional means of scalar outcomes. In this light Section IV considers conceptually the special though leading case of health- or quality-adjusted life expectancy (HALE, QALE) and health- or quality-adjusted life year (HALY, QALY) type measures of health. Key statistical properties of these measures are discussed, and the implications of some *ad hoc* approaches to estimation are demonstrated. Section V examines empirically issues involving standard measurement and estimation strategies in the context of the YHL measure and implemented with data from the 1994 U.S. National Health Interview Survey and 1993 U.S. life table data. Section VI concludes.

II. CONCEPTUAL AND MEASUREMENT ISSUES

Suppose the health status of each member of a defined population comprising N individuals at baseline[2] can be characterized by an m-vector $\mathbf{a}_i=[a_{i1},...,a_{im}]$, i=1,...,N, of measurable health "attributes" a_{ij}. (To fix ideas for the case of m=2, it may be that a_{i1} is functional status or quality of life, while a_{i2} might be life expectancy or survival.) Neither the precise nature of each attribute nor the peculiar manner in which each is measured need be of concern at this juncture, although some particular measurement issues will be of concern later on.

A scalar summary measure of "health" at the individual level is given by the mapping (the "aggregator function") $h_i=h(\mathbf{a}_i)$, though it should be emphasized that there is no *a priori* reason that any particular h_i should be more interesting than the constituent $\mathbf{a}_i$. The discussion for the remainder of this section presumes that the h_i are measurable (i.e. $\mathbf{a}_i$ observed and functional form h(.) known), although much of what follows thereafter is devoted to assessing the problems that may arise when the h_i are not directly measurable (e.g. not all elements of $\mathbf{a}_i$ observable).

Suppose moreover that each individual in this baseline

[2] The reason "baseline" is emphasized here is that any population whose health is monitored over time will experience attrition due to mortality, emigration, noncompliance, etc.. Accommodating such attrition in exercises like this is likely to be of some empirical importance. For obvious reasons, restricting attention to populations and samples of "survivors" is likely to be problematic in situations where population health is of concern: One would be hard pressed to claim that a population of size N=3 at baseline (t=0) having scalar health outcomes (say) $\{h_1=.85, h_2=.90, h_3=.95\}$ has *worse* health than a population of size N=1 having health outcome $\{h_1=.975\}$ at t=1 when individuals 2 and 3 died between periods 0 and 1.

population can be described by a two mutually-exclusive vectors of observable covariates -- $\mathbf{x}_i$ and $\mathbf{z}_i$ -- and by a vector of unobservables that can be summarized by the unobservable scalar Θ_i. In essence, the particular identity of an individual member of the population is determined by the triple $(\mathbf{x}_i,\mathbf{z}_i,\Theta_i)$. As will be discussed below, considerations of time will also be germane in some instances, so that the "i" subscripts might better be thought of as "i,t" subscripts, but this detail will be omitted unless needed for clarity. The population joint distribution of $(\mathbf{a}_i,\mathbf{x}_i,\mathbf{z}_i,\Theta_i)$ is given by $G(\mathbf{a},\mathbf{x},\mathbf{z},\Theta)$, which has the corresponding conditional distribution $G(\mathbf{a}|\mathbf{x},\mathbf{z},\Theta)$ (the unsubscripted $(\mathbf{a},\mathbf{x},\mathbf{z},\Theta)$ are typical elements).[3]

The remainder of this section sketches a variety of conceptual approaches to population health measurement based on these population distributions G(.) of health attributes (**a**) or health status (h). Some of this discussion will be familiar to readers having exposure to the literature on economic inequality (see Litchfield, 1999, for a survey); yet despite the parallels, it should be emphasized that the main purposes of the inequality literature are in many regards different from the purposes of this exposition.

Population Health Functionals

A most general summary scalar characterization (H) of the health -- or perhaps the social value associated *with* the health -- of this baseline population, conditional on its observed and its unobserved covariates, is given by the value of the *functional* $H=F[G(\mathbf{a}|\mathbf{x},\mathbf{z},\Theta)]$ or $H=F[G(h|\mathbf{x},\mathbf{z},\Theta)]$ (see Allen, 1938,

[3] A minor variation would be to view **a** as being determined nonstochastically once $(\mathbf{x},\mathbf{z},\Theta)$ are given, i.e. $\mathbf{a}=\mathbf{a}(\mathbf{x},\mathbf{z},\Theta)$, so that $G(\mathbf{a}|\mathbf{x},\mathbf{z},\Theta)$ would be degenerate.

pp. 521-523, for a useful exposition of functionals).[4] A functional "values" functions defined on the same domain in much the same way that a function "values" the arguments to a function. Allen, 1938, informally characterizes a functional as a limiting case of a function when the number of arguments in the function is permitted to go to infinity.

With reference to figure 1, the functional F[.] would value -- by assigning a larger and smaller value to population health H -- the two population distributions of the scalar health outcome h, $G_1(h)$ and $G_2(h)$. Importantly, though, this mapping is not dependent *per se* on the mean, variance, order statistics, quantiles, or any other particular feature(s) of the $G_i(h)$. Rather, it is based on the entirety of the respective probability distributions, i.e. the positions of all the points constituting G_1 vs. all the points constituting G_2. How the valuation mechanism -- the functional -- is structured would depend on the analyst's or the policymaker's sense of what constitutes a scalar summary measure of the health of a population.

While conceptually an ideal setup for quantifying the health of a heterogeneous population, the formidable and obvious practical problem here is designing an operational way to rank alternative functions, i.e. what is the "functional" form? Nonetheless, while likely to be of little practical use, it is still of great conceptual utility to conceive of population health quantification in terms of mappings from the space of population health or health attribute distributions to scalar measures of health via the tool of a functional.

[4] The functional is a concept used commonly, e.g., in expected and non-expected utility analysis (see, e.g., Machina, 1988, on "preference functionals").

Stochastic Dominance

A more concrete way alternative distributions of population health can be compared is to invoke criteria of first- and second-order stochastic dominance in the scalar case -- i.e. pertaining to distributions G(h;.)-- and corresponding notions of multivariate stochastic dominance for considerations of distributions G(**a**;...). Because the literature on stochastic dominance (at least for the univariate case) is rather well developed the discussion here will be brief. The important point to carry away here is that a stochastic dominance approach to ordering population distributions of health or health attributes is in many respects a middle ground between the unstructured approach of population health functionals and the more restrictive (albeit more practical) approaches based on moments, quantiles, order statistics, and tail probabilities sketched out in the next section.

In comparing two distributions defined on a scalar variate (e.g. h), say $G_1(h)$ and $G_2(h)$, G_1 exhibits first-order stochastic dominance over G_2 if $G_2(h) \geq G_1(h)$ for all h, with $G_2(h) > G_1(h)$ for at least some h, while G_1 exhibits second-order stochastic dominance over G_2 if $\int_{-\infty}^{r} G_1(h)dh \leq \int_{-\infty}^{r} G_2(h)dh$, with strict inequality for at least some h. In terms of these stochastic dominance measures, one distribution of population health would be judged "better" than an alternative if it exhibited an appropriate j-th order stochastic dominance.

Multivariate stochastic dominance is a much less well developed concept, but would be the appropriate concept in assessing the relative merits of competing distributions of health attributes, say $G_1(\mathbf{a}|\mathbf{x},\mathbf{z},\Theta)$ and $G_2(\mathbf{a}|\mathbf{x},\mathbf{z},\Theta)$. Arguments fully paralleling those developed by Atkinson and Bourguignon, 1982, in their analysis of multidimensioned distributions of

economic status would be appropriate in this regard; indeed Atkinson and Bourguignon pursue an example that entails some considerations of health (life expectancy). The analytics of multivariate stochastic dominance are more formidable than those applicable in the univariate setting, however.

Functions of Moments, Quantiles, Order Statistics, etc.

Virtually any particular summary measure of population health that can be or that has been conceptualized will be a special case of the functional F[.], and will typically be represented by some *function* of the moments, quantiles, order statistics, or tail probabilities of $G_{\mathbf{a}}(\mathbf{a})$ or of $G_h(h(\mathbf{a}))$. Ignoring for the moment conditioning on some set Ω, let $\mu_{\mathbf{a}}$ and $\Sigma_{\mathbf{a}}=[\sigma_j]$ denote, respectively, the finite m-vector and m×m matrix of population means and covariances of the marginal distribution $G_{\mathbf{a}}(\mathbf{a})$, i.e. $\mu_{\mathbf{a}}=E_{G_{\mathbf{a}}}[\mathbf{a}]$ and $\Sigma_{\mathbf{a}}=E[(\mathbf{a}-\mu_{\mathbf{a}})(\mathbf{a}-\mu_{\mathbf{a}})']$. Similarly let ξ_h and ν_h denote the scalar population mean and variance of the marginal distribution $G_h(h(\mathbf{a}))$, and $\xi_h^{(p)}$ be the p-th raw moment of $G_h(h(\mathbf{a}))$.

Then some particular characterizations would be, e.g., a general function defined on low-order moments of $G_h(h(\mathbf{a}))$,

$$H = \varphi(\xi_h, \xi_h^{(2)}, \ldots, \xi_h^{(r)}); \tag{1}$$

mean-variance measures

$$H = \alpha'\mu_{\mathbf{a}} + \beta'\Sigma_{\mathbf{a}}\beta \tag{2}$$

or

$$H = \gamma\xi_h + \delta\nu_h, \tag{3}$$

where α and β are prespecified k-vectors and γ and δ are scalars to be specified; θ-quantiles,

$$H = \mathcal{Q}_\theta(G_h(h(\mathbf{a}))), \tag{4}$$

where $\int_{-\infty}^{\mathcal{Q}_\theta} dG_h(h(\mathbf{a})) = \theta$; tail probabilities or "distiles",

$$H = \Pr(h<h^*) \text{ or } \Pr(h>h^*) \tag{5}$$

which are encountered in practice as, e.g., percentage of births that are low birthweight or percentage of adults that have BMI exceeding an obesity criterion; or Rawlsian maximin (order statistic) measures defined on the set of h_i,

$$H = \text{argmin}\{h(\mathbf{a}_1),\ldots,h(\mathbf{a}_N)\}. \tag{6}$$

III. HETEROGENEITY, MEANS, AND COVARIANCES

Decision Criteria Defined in Terms of Population Expectations

Much of what follows will be based on the working assumption that conditional or unconditional expectations (means) *of some quantity* are the main concern in the population health measurement exercise. As noted above, this is by no means a necessary focus and may -- for both the reasons indicated above as well as for some additional reasons spelled out below -- distract attention from other aspects of the distribution of population health that may be of concern. Without reference to a particular form of a social welfare function, there is no way to judge whether a mean or some other parameter is the "appropriate" parameter to summarize the health of a population.

One prominent area where conditional and unconditional expectations play a key role is in medical technology evaluation, i.e. cost-effectiveness or cost-utility analysis (CEA, CUA). It is common practice in assessing a new technology's (T) desirability relative to a standard or baseline practice (B) in terms of cost (c) and health effectiveness (e) to define parameters like the incremental cost-effectiveness ratio (ICER) or the incremental net benefit function (INHB; Stinnett and Mullahy, 1998) in terms of the underlying marginal population means of costs and health outcomes, e.g. $ICER=[(\mu_{c_T}-\mu_{c_B})/(\mu_{e_T}-\mu_{e_B})]$ and $INHB=[(\mu_{e_T}-\mu_{e_B})-(\mu_{c_T}-\mu_{c_B})/\lambda]$ where $\mu_{c_i}=E[c_i|\Omega]$ and $\mu_{e_i}=E[e_i|\Omega]$, $i\in\{T,B\}$, λ represents the social value of the health increment, and Ω is some conditioning set involving $[\mathbf{x},\mathbf{z}]$.[5] While it is by no means necessary to assert such a means-based definition, these definitions would appear to be the concepts referred to when most analysts consider statistical characterizations of ICERs and INHBs. As such -- and despite the interesting intellectual debates surrounding alternative definitions (e.g. whether the ICER should be based on the ratio of means or on the means of the ratios (see Stinnett and Paltiel, 1997, for discussion)) -- this paper will adopt the means-based approach as the analytical centerpiece.[6]

Of course, if interest is more in monitoring population health status/outcomes than in technology assessment *per se*, then the relevant comparison might simply be based on the differences

[5] Phelps, 1997, is an excellent discussion of the importance of conducting *conditional* technology evaluation.

[6] While not a primary concern of this paper, it might be noted at this juncture that the statistical/inferential properties of the analogy estimators of ICER and INHB so defined have been assessed in a series of recent papers (e.g. Chaudhary and Stearns, 1996; Mullahy and Manning, 1995; Willan and O'Brien, 1996).

$E[e_T - e_B|\Omega]$ or, in the previous notation, $E[H_T - H_B|\Omega]$ themselves, where H_T and H_B represent in such instances measures of health status in different populations, or in a given population at different points in time or under different imagined policy initiatives.

Covariances of Health Attributes Across the Population

As a general matter it would be expected that the population variance-covariance of the health attributes $\mathbf{a}$, $\Sigma_{\mathbf{a}}$, will be nondiagonal. That is, at least for some attributes in $\mathbf{a}_i$ it would be unlikely if individuals having relatively high a_{ij} were not also found on average to have relatively high (or, possibly, low) a_{ik} for some (j,k) pairs.

Take for the moment a_{ij} to be a measure of quality of life and a_{ik} to be a measure of longevity. Should it turn out that individuals having relatively high propensities to "live long" also have relatively high propensities to "live well," then consideration of such correlation is essential in properly characterizing the outcome measures. This section describes some simple stochastic frameworks within which such considerations might be addressed formally, and then proceeds subsequently to present some central implications of these results for conducting empirical CEA/CUA,[7] undertaking population health monitoring, or

[7] Such an endeavor is warranted in part because of a potentially important friction between theoretical CEA and empirical CEA. In particular, the use of QALY-type measures as von-Neumann-Morgenstern utility functions has been shown to depend on a set of conditions that require *inter alia* a form of preference independence between quality of life and longevity (Pliskin et al., 1980; Bleichrodt and Johannesson, 1997). QALY measures thus rationalized provide a relatively more solid basis for welfare analysis. Yet regardless of whether individuals' utility functions are structured in accordance with these conditions, QALY measures will be used as outcome measures in CEA. Assessing
(continued)

formulating public policies involving health interventions.

Suppose k=2 so that $h_i=h(a_{i1},a_{i2})$, and suppose h(.,.) is continuously differentiable. Consider a measure of population health akin to (3) with $\gamma=1$ and $\delta=0$, i.e. H is expected or mean population health ξ_h. In many practical applications, focus will be on such mean outcomes; not least in this regard are randomized trials in which differences in mean outcomes between treatment and control will often be the basis of regulatory efficacy claims (e.g. FDA NDAs)

To understand how the structure of H depends in this instance on the variance-covariance structure $G_{\mathbf{a}}(\mathbf{a})$, consider the second-order Taylor expansion of $h(a_{i1},a_{i2})$ around the vector $\mu_{\mathbf{a}}=[\mu_{a_1},\mu_{a_2}]$:

$$h(a_{i1},a_{i2}) \approx h(\mu_{a_1},\mu_{a_2}) + \left[h_1, h_2\right] \times \begin{bmatrix} a_1 - \mu_{a_1} \\ a_2 - \mu_{a_2} \end{bmatrix} + \frac{1}{2}\left[a_1 - \mu_{a_1}, a_2 - \mu_{a_2}\right] \times \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} \times \begin{bmatrix} a_1 - \mu_{a_1} \\ a_2 - \mu_{a_2} \end{bmatrix}, \quad (7)$$

where h_j and h_{jk} are the first- and second-order partial derivatives of h(.,.) evaluated at $\mu_{\mathbf{a}}$. It follows that

---

the empirical implications of how quality of life and longevity may be related in the population -- i.e. their lack of independence in the population -- is ultimately the major concern of this paper.

$$\xi_h = E_{G_{\mathbf{a}}}[h(a_{i1},a_{i2})] \approx h(\mu_{a_1},\mu_{a_2}) + \frac{1}{2}\left[h_{11}\sigma_{11} + h_{22}\sigma_{22}\right] + h_{12}\sigma_{12}. \quad (8)$$

Thus, to the extent that h(.,.) manifests nonlinearities in its arguments (h_{ij} nonzero), then H -- characterized thusly as mean population health -- will depend not just on the population means of the component health attributes μ_{a_1} and μ_{a_2} via the term $h(\mu_{a_1},\mu_{a_2})$, but also on the variance-covariance structure of $G_{\mathbf{a}}(\mathbf{a})$, i.e. $\Sigma_{\mathbf{a}}$. The general result for m≥2 is

$$\xi_h \approx h(\mu_{a_1},\ldots,\mu_{a_m}) + \frac{1}{2}\sum_{j=1}^{m} h_{jj}\sigma_{jj} + \sum_{j=1}^{m-1}\sum_{k=j+1}^{m} h_{jk}\sigma_{jk}. \quad (9)$$

Among other things (8) and (9) demonstrate that as a general matter, information on (e.g. estimates of) more than just the marginal means of the health attributes will be required to quantify H when the latter is defined in terms of mean population health.

Note that (8) and (9) are exact if h(.) is linear, quadratic, or multiplicative in the sense of containing only first-order interactions. Specifically, for the case with m=2 where $h(a_{i1},a_{i2})=a_{i1}\times a_{i2}$ one has exactly

$$\xi_h = \mu_{a_1}\times\mu_{a_2} + \sigma_{12}, \quad (10)$$

which, upon inspection, is simply a restatement of the definition of a covariance, i.e.

$$\mathrm{cov}(a_1,a_2|\Omega) = E[a_1\times a_2|\Omega] - E[a_1|\Omega]\times E[a_2|\Omega], \quad (11)$$

where Ω is some relevant conditioning information. Yet equation (10) and variants thereon will be shown below to play a fundamentally important role in understanding the extent to which QALE- or QALY-type summary measures of health -- as commonly implemented -- in fact quantify a population's health as they are ostensibly designed to do. Specifically, the remainder of the paper will focus largely on population health measures along the lines of $E[a_1 \times a_2 | \Omega]$, where a_1 will generally represent some measure of the quality of life at a point in time while a_2 will be some measure of survival, life expectancy, longevity, etc., whose precise definition will be context-specific.

Some Implications of Multiplicative Functional Forms

Suppose, thusly, one takes $a_1=q$ to be "quality of life" and $a_2=\ell$ to be some measure of "longevity" (defined suitably at the individual level). Instead of $E[q \times \ell | \Omega]$, however, suppose one argued for $E[q|\Omega] \times E[\ell|\Omega]$ as a population health measure. Then by definition this measure of health would be

$$E[q|\Omega] \times E[\ell|\Omega] = E[q \times \ell|\Omega] - cov(q,\ell|\Omega). \qquad (12)$$

There is nothing inherently incorrect about such a strategy. Yet it is important to emphasize that such a "product of means" measure handles in a different way certain features of a population's joint distribution of health attributes that *may* be of interest from a welfare perspective than does the "mean product" measure.

For example, an Ω-population of size N=3 with (q,ℓ) outcomes $\{(0.5,6),(0.9,10),(1.0,20)\}$ would, by the standards of

(12), be judged to have the same level of population health (equal to 9.6) as would the population {(0.75,13),(0.8,12),(0.85,11)}, whereas the health of the former population would be judged to better than the health of the latter population (10.67 vs. 9.57) under the $E[q \times \ell | \Omega]$ measure. For all practical purposes, only if q and ℓ are independently distributed in the population (although zero correlation would in fact suffice) would these distinctions become irrelevant. Since such independence would seem tenuous to maintain *a priori* for many measurement settings that can be imagined, allowance for the implications of nonzero correlation seems the prudent analytical course.[8]

The seemingly simple multiplicative functional form $q \times \ell$ in fact provides considerable structure (and, therefore, considerable restrictions) on the manner in which the various attributes combine to produce "health" in a heterogeneous population. For given marginal means $E[q|\Omega]$ and $E[\ell|\Omega]$, expectations of the multiplicative form will reward (in terms of their measured population health status) populations having relatively high $cov(q,\ell|\Omega)$ relative to those having low or

[8] To forestall potential confusion, a point regarding "independence" might be clarified at this juncture. The issues described here are inherently statistical in nature. As noted in the introduction and directly above, the concerns about issues such as independence between quality and longevity when using QALYs or related measures as utility or welfare measures are essentially concerns about the structures of individuals' utility functions (e.g. Pliskin et al., 1980; Bleichrodt and Johannesson, 1997). "Independence" used in such contexts connotes something potentially very different than would "independence" used, for instance, to characterize a population joint probability distribution (e.g. $G(q,\ell)=G(q)\times G(\ell)$).

negative cov(q,$\ell|\Omega$). More general and flexible forms (e.g. the second-order or quadratic form (8)) would allow one in principle to mitigate or offset the implications of such reward structures, but whether the available data would be up to the task of implementing such measures remains to be seen.

To close this section it is worth noting that with multiplicative mappings of attributes to health -- and for many other functional forms that might be imagined as well, as suggested by (9) -- it is intuitive that the population health measure would demonstrate some reliance on the variance-covariance structure of the attributes, $\Sigma_{\mathbf{a}}$. If all the elements of **a** were perfectly correlated, then information on any single element would convey just as much information as would the entire vector. Conversely, to work under a maintained assumption of a diagonal $\Sigma_{\mathbf{a}}$ -- i.e. all elements perfectly uncorrelated -- would be empirically unrealistic owing to the likely joint reliance of the a_{ij} on a deep set of individual biophysical characteristics ("health capital"). As such, the prominence of $\Sigma_{\mathbf{a}}$ in many interesting measures of population health should be considered logical rather than surprising.

IV. HALE/QALE- AND HALY/QALY-TYPE MEASURES

The last decade witnessed a burgeoning use of what will be referred to here generically as health-adjusted life year ("HALY") or health-adjusted life expectancy ("HALE") measures in medical technology evaluation and population health monitoring.[9]

[9] See Fryback, 1997, for an excellent comprehensive overview. The use of Quality-Adjusted Life Years (QALYs), Disability-Adjusted Life Years (DALYs), Health-Adjusted Life Years (HALEs), Years of Healthy Life (YHLs), etc., as the central outcome measures in cost-effectiveness analysis (CEA) -- or, perhaps more accurately, cost-utility analysis (CUA) -- is both well-
(continued)

An enormous amount of intellectual energy has been dedicated to understanding the conceptual features of such measures, their relationships (or lack thereof) to welfare economics, etc. Moreover, much intellectual effort has been dedicated to developing sound empirical methods for estimating both *components* -- the health adjustment and the life year or life expectancy measure -- of HALYs or HALEs (see Johannesson et al., 1996, for a comprehensive survey).[10]

Yet despite all this scholarly effort, remarkably little attention has been devoted to understanding the conceptual, empirical, and statistical properties of HALY or HALE measures themselves. The "numbers" used in such enterprises to measure life's quality and length and their ultimate aggregation into the HALY/HALE measure necessarily derive from some source. Either they are asserted from expert opinion -- thus rendering statistical analysis irrelevant -- or they are derived in some manner from some data source or sources, in which case their classical statistical properties (bias, consistency, efficiency,

---

entrenched in practice and advocated formally by the recent report of the U.S. Panel on Cost-Effectiveness in Health and Medicine (PCEHM) (Gold et al., 1996). Quality-adjusted measures of longevity and life expectancy have also become a centerpiece of efforts to monitor the health status of large populations and, ultimately, to base policy recommendations thereon (Erikson et al., 1995; Cutler and Richardson, 1997).

[10] For the health- or quality-adjustment component, methods like the standard gamble (SG), time tradeoff (TT0), rating scale (RS), and others have been advocated and debated. For the survival, life expectancy, or longevity component (the "LY" part of HALY), a substantial share of the research agenda has been devoted to developing statistical methods for estimating the appropriate hazard or survival models, with issues like whether proportional or additive hazard models may be more appropriate in particular circumstances occupying the center stage (see, for example, Beck et al., 1982a, 1982b). There has also ensued a debate of considerable fervor about the merits, shortcoming, differences, and similarities of alternative HALY measures (e.g. QALYs vs. DALYs).

etc.) become pertinent considerations. This section assumes the second perspective and develops a framework within which the properties of various HALE/QALE or HALY/QALY measures can be assessed.

The entire section works from two fundamental assumptions: first, that health at the individual level is a multiplicative function in attributes a_{i1} and a_{i2}, i.e. $h_i=a_{i1}\times a_{i2}$; and second, that the summary measure of population health is a mean or conditional mean measure. As noted above, while neither of these assumptions is necessary to pursue measurement of population health, they do characterize what would be more-or-less agreed to be "standard practice."

The key issue thus becomes the practical one of estimating population health measures on the basis of available data and assessing such estimates by classical statistical standards, bias being of greatest concern here. That is, do the estimates tend on average to describe the population parameters they are supposed to mimic? As will be suggested in what follows, the particulars of how data on both quality of life and longevity/survival/life expectancy are assembled and utilized -- given that data on one or both attributes are often not available at the individual level -- will play a key role in assessing the "bias" properties of the population health estimates so obtained.

HALE/QALE-Type Measures

One aspect of population health measures based on "life expectancy" (ℓ) is that they are inherently not measurable at the individual level. "Life expectancy" is not observable at an individual level as are outcomes like survival or systolic blood pressure. Rather, life expectancy, conditional on some set Ω (e.g. age, gender, race), is typically estimated actuarially on the basis of recent average mortality experiences of "like"

populations, with the resulting estimates often summarized in the form of life tables (e.g. Anderson, 1999). Measures akin to Health- or Quality-Adjusted Life Expectancy (HALE, QALE) will -- even if ostensibly individual-level -- necessarily entail averages of one or more measures taken over some related population. This holds true regardless of whether individual-level measures of quality of life are available.

It may be that in some situations individual-level measures of quality of life are accessible in the data of interest, while in other instances it will be necessary to rely on like experiences, regression predictions, published results, etc. to obtain factors that provide for estimates of the "QA" or "HA" part of the QALE or HALE measure. Rosenberg et al., 1998, note that: "Different approaches for calculating HALE use quality numbers based on an average of one measurement at one time-point over all individuals, an average over all individuals longitudinally over time, or perhaps, measurements taken from individuals at repeated time intervals." As will be seen in the next section, the Years of Healthy Life (YHL) measure is an example of this type of measure: Life expectancy data are obtained from life tables, while quality adjustments are obtained on the basis of a large sample of individual-level responses to two survey items that map into the so-called HALex index, with the (estimated) parameters of the mapping themselves usefully conceived as arising from some averaging procedure.

The goal of this subsection is to present an analytical structure within which the merits of alternative computation/estimation strategies for HALE- or QALE-type measures can be assessed given a population heterogeneous in $(q,\ell,\mathbf{x},\mathbf{z},\Theta)$ *and* one for which conditional or unconditional covariance between q and ℓ may be nonzero $(\text{cov}(q,\ell|...)\neq 0)$. It should be stressed that the effort here is to develop a framework that is quite

general. Specific implementations will typically require additional considerations.[11]

The general result is as follows. Suppose the population health measure that is the objective of estimation is

$$H = E\{\ E[q|\Omega_q] \times E[\ell|\Omega_\ell]\ \mid \Omega\ \} \qquad (13)$$

$$= E\{\ \zeta(\Omega_q) \times \psi(\Omega_\ell)\ \mid \Omega\ \},$$

where Ω_q, Ω_ℓ, and Ω represent possibly different conditioning information, such that $\{\Omega_q \cup \Omega_\ell \cup \Omega\} \subseteq \{\mathbf{x}, \mathbf{z}, \Theta\}$. (Whether this is the ideal measure or the measure that is necessarily specified owing to data restrictions is not essential here.) In particular, Ω_q and Ω_ℓ may be chosen such that q and/or ℓ are completely

---

[11] Indeed, it is instructive to read how one particular HALE measure is actually computed; the following excerpt from the Rosenberg-Fryback-Lawrence, 1998, study based on the Beaver Dam Health Outcomes Study data:

> The quality and mortality numbers are combined to form the age-specific HALE by the following method. Assume that the HALE is desired for males aged 55. One-year mortality rates at each age ($q_{55}, q_{56}, \ldots$) would be estimated until a limiting age, that where the probability of death is equal to 1. HRQOL numbers at each corresponding age ($h_{55}, h_{56}, \ldots$) would also be determined. Then: The mortality rates would be combined to form survival rates ($_1p_{55}, _2p_{55}, \ldots$) to each age, where $_tp_\chi$ is the probability that a person aged χ will survive to age χ+t.
>
> $$HALE_{55} = \sum_{t=1}^{\infty} h_{55+t-1} \times {}_tp_{55}$$
>
> The HALE for the community would involve a weighting of the age-adjusted HALEs over those in the community.

determined given Ω_q and/or Ω_ℓ, i.e. "q"=E[q|Ω_q] and/or "ℓ"=E[ℓ|Ω_ℓ] so that E[q×ℓ|Ω] can be taken here to be one special case of (13) (see footnote 3 above).

Equation (13) can be reexpressed as

$$E\{E[q|\Omega_q]\times E[\ell|\Omega_\ell]\,|\Omega\} = \qquad (14)$$

$$[E\{\Omega_q\cup\Omega_\ell|\Omega\}\zeta(\Omega_q)]\times[E\{\Omega_q\cup\Omega_\ell|\Omega\}\psi(\Omega_\ell)] +$$

$$\mathrm{cov}\{\Omega_q\cup\Omega_\ell|\Omega\}(\zeta(\Omega_q),\psi(\Omega_\ell)),$$

where $E\{\nu\}[.]$ denotes the expectation of [.] taken over the distribution of ν. In words, the expectation of interest is the sum of two components: first is the product of the expectations of $\zeta(\Omega_q)$ and $\psi(\Omega_\ell)$ taken over $\Omega_q\cup\Omega_\ell$ given Ω; second is the covariation between $\zeta(\Omega_q)$ and $\psi(\Omega_\ell)$ that arises due to covariance between Ω_q and Ω_ℓ in the Ω-population. Note that if $\{\Omega_q\cup\Omega_\ell\}\subseteq\Omega$ then $E\{E[q|\Omega_q]\times E[\ell|\Omega_\ell]\,|\Omega\}$ reduces simply to $\zeta(\Omega_q)\times\psi(\Omega_\ell)$.

Concrete examples may be useful. First, suppose $\Omega_q=\{\mathbf{x},\mathbf{z},\Theta\}$, $\Omega_\ell=\{\mathbf{x},\mathbf{z}\}$, and $\Omega=\{\mathbf{x}\}$. Then (13) entails, *inter alia*, a consideration of how **z** and Θ covary in the **x**-population. Alternatively, suppose $\Omega_q=\{\mathbf{x}\}$, $\Omega_\ell=\{\mathbf{z}\}$, and $\Omega=\{\mathbf{x},\mathbf{z}\}$. Then there is clearly no covariation between $\zeta(\mathbf{x})$ and $\psi(\mathbf{z})$ over $\{\mathbf{x},\mathbf{z}\}$ since conditioning is on $\{\mathbf{x},\mathbf{z}\}$ in the first place, so the expectation defining H in (13) is simply the product of the expectations $\zeta(\mathbf{x})$

and $\psi(\mathbf{z})$.

Of course, the key issue here is that just because $E\{E[q|\Omega_q]\times E[\ell|\Omega_\ell]\,|\Omega\}$ is the target measure, it is not necessarily feasible to implement it. That is, measures or estimates of $E[q|\Omega_q]$ and/or $E[\ell|\Omega_\ell]$ may not be available. What "biases" relative to the target measure $E\{E[q|\Omega_q]\times E[\ell|\Omega_\ell]\,|\Omega\}$ would arise would clearly depend on the specific measures or averages used in computation. Obviously if the measure computed was obtained as the product of the averages $E\{\Omega_q\cup\Omega_\ell|\Omega\}\zeta(\Omega_q)$ and $E\{\Omega_q\cup\Omega_\ell|\Omega\}\psi(\Omega_\ell)$, then from (13) estimate would be biased owing to the covariance term:

$$[E\{\Omega_q\cup\Omega_\ell|\Omega\}\zeta(\Omega_q)]\times[E\{\Omega_q\cup\Omega_\ell|\Omega\}\psi(\Omega_\ell)] = \qquad (15)$$

$$E\{E[q|\Omega_q]\times E[\ell|\Omega_\ell]\,|\Omega\} - \mathrm{cov}\{\Omega_q\cup\Omega_\ell|\Omega\}(\zeta(\Omega_q),\psi(\Omega_\ell)).$$

For a more concrete example, take $\Omega_q=\Omega_\ell=\{\Theta\}$ and $\Omega=\{\mathbf{x},\mathbf{z}\}$, with $\mathbf{x}$ denoting age and $\mathbf{z}$ denoting sex. Thus, the target measure amounts to $E[q\times\ell|\mathbf{x},\mathbf{z}]= E[q|\mathbf{x},\mathbf{z}]\times E[\ell|\mathbf{x},\mathbf{z}]+\mathrm{cov}(q,\ell|\mathbf{x},\mathbf{z})$. However, suppose as above that the estimate actually used is based on $E\{E[q|\mathbf{x}]\times E[\ell|\mathbf{z}]\,|\mathbf{x},\mathbf{z}\}=E[q|\mathbf{x}]\times E[\ell|\mathbf{z}]$, i.e. mean q given age (marginal over sex and Θ) times mean ℓ given sex (marginal over age and Θ). As a general matter, there is no reason to expect that $E[q|\mathbf{x}]\times E[\ell|\mathbf{z}]$ would approximate well $E[q\times\ell|\mathbf{x},\mathbf{z}]$, with an intuition being that the over-averaging entailed in the former

would tend to mitigate variation in $E[q \times \ell | \mathbf{x}, \mathbf{z}]$ over $\{\mathbf{x}, \mathbf{z}\}$. The following exhibit demonstrates this phenomenon by means of a simple example:

Exhibit 1(a)

Cell triples are $(q, \ell, Pr(x,z))$		**x**	
		0	1
z	0	(.8,10,.25)	(1,30,.25)
	1	(.8,5,.25)	(.9,20,.25)

Exhibit 1(b)

Cell entries are $E[q \times \ell \mid x,z]$		**x**	
		0	1
z	0	8	30
	1	4	18

Exhibit 1(c)

Cell entries are $E[q \mid x] \times E[\ell \mid z]$		**x**	
		0	1
z	0	16	19
	1	10	11.875

The upshot of the discussion in this subsection is that unless the analyst has access to precisely the kind of data on q and ℓ that correspond to the particular measure of population health of interest, estimation or averaging methods used as proxies are in general unlikely to describe accurately the outcome of interest when populations are heterogeneous in their quality of life and in their life expectancy. The covariation between q and ℓ that one might expect to find in a population is ultimately the main source of such discrepancies.

HALY/QALY-Type Measures

At any time t, a "HALY" or "QALY" can appropriately be thought of a the product of a 0-1 individual-specific survival indicator at time t (s_t) and the individual's quality of life at period t (q_t). One imagines then a joint distribution $G(\mathbf{q},\mathbf{s},\mathbf{x},\mathbf{z},\Theta)$ where $\mathbf{q}$ and $\mathbf{s}$ are vectors of health outcome attributes comprising $\{q_t\}$ and $\{s_t\}$ respectively.[12] It is common practice to normalize $q_t=0$ as the quality of life equivalent to being dead so that $s_t=0$ is sufficient for q_t to be zero, but is not necessary. Conditional on some set $\Omega\subseteq\{\mathbf{x},\mathbf{z}\}$, the "expected QALY" measure (ignoring discounting for present purposes) for an individual whose quality of life and survival status would in principle be measured[13] over T+1 time periods from baseline (t=0) to termination (t=T) is given by

$$\begin{aligned} E[QALY|\Omega] &= \sum_{t=0}^{T} E\left[q_t \times s_t \mid \Omega\right] \\ &= \sum_{t=0}^{T} E\left[q_t \mid \Omega\right], \end{aligned} \tag{16}$$

with the second equality following from the normalization that "$s_t=0$"⇒"$q_t=0$". A quantity akin to (16) is presumably what the U.S. Panel on Cost-Effectiveness in Health and Medicine contemplated when it defined the QALY measure as

[12] A continuous-time setting where, e.g., survival time is measured as a continuous scalar variable can be captured here by imaging small time increments.

[13] "in principle" is noted here since QALY-type measures will typically be most useful in *ex ante* contexts where individual-level survival is not known.

> *...the sum of the quality weights for the various health states...multiplied by the duration (in years or fractions of years) of each health state. This is the number of QALYs gained without discounting.* (PCEHM, p. 92)

(See also Glasziou et al., 1990, and Zhao and Tsiatis, 1997.)

Ignoring "t" subscripts, the typical term in the summand (16) can be usefully decomposed as

$$E[q\times s|\Omega] = E[q\times s|\Omega,s=1]\times Pr(s=1|\Omega) + E[q\times s|\Omega,s=0]\times Pr(s=0|\Omega) \quad (17)$$

$$= E[q|\Omega,s=1]\times Pr(s=1|\Omega). \quad (18)$$

As such, and in particular owing to the 0-1 measurement of s and the "s=0"⇒"q=0" normalization, the "expected QALY" $E[q\times s|\Omega]$ can be obtained as the product of the mean q among survivors with Ω ($E[q|\Omega,s=1]$) and the probability of survival at time t for those with Ω ($Pr(s=1|\Omega)$). This, of course, is tantamount to how QALYs are measured when individual-level data are available in (say) clinical trials, i.e. as areas under quality-adjusted survivor curves (the thin curve in figure 2).

An enormous advantage of the normalizations leading to the result (18) is that separate (consistent) estimates of $E[q|\Omega,s=1]$ and $Pr(s=1|\Omega)=E[s|\Omega]$ -- even though possibly obtained from different sources of data -- will suffice to estimate (consistently, owing to Slutsky's theorem) the proper conceptual measure, $E[q\times s|\Omega]$. Since researchers do not always have the luxury of access to data sources where both q and s are observed for individual subjects, this result is obviously of some practical importance. As a general matter, combining separate estimates of component expectations will not afford such a solution. Indeed, it must be emphasized that this result

requires estimation of $E[q|\Omega,s=1]$, *not* of $E[q|\Omega]$, in order to work.

Quite generally, it is useful to note that there is a nonzero covariance (given Ω) between q and s:

$$E[q|\Omega]\times E[s|\Omega] = \{E[q|\Omega,s=1]\times Pr(s=1|\Omega) + E[q|\Omega,s=0]\times Pr(s=0|\Omega)\} \times Pr(s=1|\Omega), \quad (19)$$

implying

$$cov(q,s|\Omega) = E[q|\Omega,s=1] \times Pr(s=1|\Omega) \times \{1-Pr(s=1|\Omega)\} > 0. \quad (20)$$

That the covariance is necessarily positive can be seen by examining the sample space depicted in figure 3. A linear regression of q on s will obviously have a positive slope; recognizing that the sign of the slope of the linear regression of q on s has the same sign as cov(q,s), it follows that $cov(q,s|\Omega)>0$.[14]

V. AN EXAMPLE: MEASURING POPULATION HEALTH USING YEARS OF HEALTHY LIFE

The Years of Healthy Life (YHL) measure was developed and reported by researchers at the National Center for Health Statistics (Erikson et al., 1995, henceforth "the YHL report"). Two heralded features of the YHL measure are its ability to monitor continually the health of the U.S. population (since its

[14] That this covariance is necessarily positive *at any point in time* in the population is a matter separate from the fact that medical technology interventions may result in changes in both the survivor function as well as in its quality-adjusted counterpart such that increases in survival probabilities may arise either in lockstep with or at the expense of increases in quality of life.

empirical basis is a battery of unchanging questions from the annual National Health Interview Surveys (NHIS)) and its computational ease (since it is based on a small number of items from the NHIS in conjunction with standard life table information). See Gold et al., 1997, for additional discussion.

An example of the output from the YHL measurement algorithm is presented in table 1 (excerpted and abridged from Erikson et al., 1995). Columns 1, 2, and 3 display the five-year age increments, the number of individuals alive at the beginning of the age increment from an imaginary birth cohort of 100,000 individuals, and the stationary population in the full five-year age range, respectively. Columns 1-3 are derived from NCHS life tables. Column 4 depicts the within-age-interval sample mean QOL index (the so-called Health and Limitations Index, or HALex), derived from two survey items on the NHIS sample regarding self-perceived health status (EVGFP) and limitations due to disability (and supplemented with information on the institutionalized and military populations). Column 5 is the product of columns 3 and 4, and column 6 is the bottom-up cumulative of column 5 (i.e. $C6_j = C5_j + \sum_{k=j+1}^{\bar{J}} C5_k$). Column 7 is column 6 divided by column 1, and column 8 is taken from life tables.

The key issue here concerns the computation used to generate the figures in column 5 where the ostensible objective is to obtain the total number of healthy life years within each age interval. That is, the QOL scores for each age interval are obtained as *averages within the age interval* of the individual QOL scores within the interval. For example, for ages 0-5 the sample average QOL score of 0.94 accounts for a large percentage of young children in "perfect health" (QOL=1.0) and a small percentage in sub-perfect health.

An Analysis of Health-Adjusted Life Expectancy

Unfortunately the NHIS data do not permit a comparison with the YHL report's findings in light of the previous analyses since NHIS is a residential survey and the YHL made a series of modifications to accommodate institutionalized populations. Instead this subsection describes an illustrative analysis of the computation of health-adjusted life expectancy based on HALex data from the 1994 NHIS (sample size 47,719) combined with U.S. life table data from 1993. The main thrust of this exercise is to demonstrate the sensitivity of estimates of population health summary measures (here HALE) to alternative strategies for obtaining HALex values (q) and life expectancy (ℓ) measures. Specifically, different averaging strategies for both q and ℓ across different subpopulations are considered. The HALE measures developed here are purely illustrative; more sophisticated measures based on specific demographic considerations like mortality could be developed in extensions of this work.

As indicated above, life expectancy is not a variable "observed" at an individual level. Rather, life expectancy measures are inherently population or sub-population averages based on recent mortality experiences of similar populations. The life expectancy measures used here from 1993 U.S. life tables are available for the 272 subpopulations defined by age (in years; 18-85 are used here), sex, and race (white vs. nonwhite). Conversely, the quality-adjustment measures (q) from the HALex are, in some sense, "observed" at an individual level, although as noted earlier the health state scores are based on parameters from the HUI Mark-I index, with these parameters themselves being summaries or averages. So for purposes of this analysis q is treated as an individual-level measure, but in fact it may not be as "individual" in nature as would, say, a systolic blood

pressure readind or an FEV_1 measure.

For purposes of this discussion, the conditioning information $(\mathbf{x},\mathbf{z})=\{age,sex,race\}$; the particular characterizations of $\mathbf{x}$ and $\mathbf{z}$ will be specified below. From the earlier discussion, the ideal measures of population health would be parameters like $E[q\times\ell|\mathbf{x}]$ or $E[q\times\ell|\mathbf{x},\mathbf{z}]$. Of course, ℓ is not observed at the individual level; the measures available from the life tables are $E[\ell|\mathbf{x}]$ or $E[\ell|\mathbf{x},\mathbf{z}]$. As such, the full range of individual-level conditional variation in ℓ *and* conditional covariation between q and ℓ cannot be exploited. For instance, were the objective estimation of $E[q\times\ell|\mathbf{x},\mathbf{z}]=E[q\times\ell|age,sex,race]$ and were the analysis based on the measures $E[\ell|\mathbf{x},\mathbf{z}]=E[\ell|age,sex,race]$, then the estimand would be

$$E[q\times E[\ell|\mathbf{x},\mathbf{z}]|\mathbf{x},\mathbf{z}] = E[q|\mathbf{x},\mathbf{z}]\times E[\ell|\mathbf{x},\mathbf{z}] \qquad (21)$$
$$= E[q\times\ell|\mathbf{x},\mathbf{z}] - cov(q,\ell|\mathbf{x},\mathbf{z}),$$

so that the ideal measure is over- or under-stated to the extent that the covariance between q and ℓ in the sub-population defined by $(\mathbf{x},\mathbf{z})$ -- owing to sub-population heterogeneity in Θ -- is negative or positive. Such discrepancies might be taken as indicators of the "bias" resulting from the use of "average" or "proxy" information. For purposes of this particular exercise, it is important to recognize that much of the population heterogeneity in life expectancy (were it in fact "measurable" at an individual level) is likely be attributable to factors beyond

{age,sex,race}, i.e. Θ.[15] As such, the empirical results to be discussed now should be interpreted thusly.

Two sets of measures of population health are of concern. The objective of the first exercise is estimation of the age/sex-specific mean QALE $E[q\times\ell|\mathbf{x}]$, with $\mathbf{x}$={age,sex} and $\mathbf{z}$={race}. For the reasons discussed above, the "best" one can do here to exploit heterogeneity in $q\times\ell$ beyond that owing to $\mathbf{x}$ is computation based on $E[q\times E[\ell|\mathbf{x},\mathbf{z}]|\mathbf{x}]$ since the life table data on ℓ are available only at the {age,sex,race} level. This exercise will illustrate how results based on this approach compare with those obtained when alternatives based on $E\{E[q|\mathbf{x}]\times E[\ell|\mathbf{x}]|\mathbf{x}\}=E[q|\mathbf{x}]\times E[\ell|\mathbf{x}]$ are used. The second exercise is to estimate the unconditional population HALE/QALE, with the same basic considerations about alternative averaging approaches assessed as well.

The detailed age-sex results are presented in table 2. It is seen here that for virtually all age-sex sub-populations the empirical covariance between q and $E[\ell|\mathbf{x},\mathbf{z}]$ is positive, as expected, but small in magnitude. That these covariances are small -- with the corollary implication that the differences between $E[q\times E[\ell|\mathbf{x},\mathbf{z}]|\mathbf{x}]$ and $E[q|\mathbf{x}]\times E[\ell|\mathbf{x}]$ are small -- should not be surprising since the only heterogeneity and covariation that are effectively being exploited in the computation of $E[q\times E[\ell|\mathbf{x},\mathbf{z}]|\mathbf{x}]$ are those due to the fact that q and ℓ covary by

[15] Indeed, a linear regression of q on a quadratic in {age,sex,race} gives an R^2 of only 0.11 in the full sample. Whether {age,sex,race} captures more or less of the variation in ℓ than in q (so measured) is not known.

race within each {age,sex} subpopulation.

The second exercise, computation of the unconditional $E[q \times E[\ell|\mathbf{x},\mathbf{z}]]$ vs. proxying by $E[q] \times E\{E[\ell|\mathbf{x},\mathbf{z}]\}$ results in estimates $E[q \times E[\ell|\mathbf{x},\mathbf{z}]]=28.63$ and $E[q] \times E\{E[\ell|\mathbf{x},\mathbf{z}]\}=27.73$, with an implied $cov(q,E[\ell|\mathbf{x},\mathbf{z}])=0.9$. The covariance is obviously larger in this instance since covariation between q and ℓ across age, sex, and race subpopulations influence the computations.

A Simulation

This subsection provides a simulation exercise based on the specific information from the YHL report. For simplicity of exposition, the focus here is on the particular measure "Years of Healthy Life for the Population Ages 85+."[16] (Note the relevant conditioning **x** here is simply a single age category.) The idea is to simulate under alternative correlation assumptions a set of individual-level samples of N=31,892 observations consistent with the marginals and averages appearing in the last row of table 1, and then to assess how the corresponding measures (averages, totals) of quality-adjusted life expectancies taken over the individuals in these pseudo-samples relate to the correlation assumptions.

Specifically, let u_{i1}, i=1,...,31892, be pseudo-random N(0,1) variates and let $u_{i2}=\alpha u_{i1}+u_{i3}$, where the u_{i3} are also N(0,1) pseudo-random variates independent of u_{i1}. The population correlation between the u_{i1} and u_{i2} is then given by

[16] The arguments advanced here generalize, *mutatis mutandis*, to the YHL measure for the entire population as examined by Erikson et al., but are most easily exposited for a this particular category so that cumulation can be avoided.

$\rho_{12}=\frac{\alpha}{\sqrt{\alpha^2+1}}$. (Note that u_{i1} and u_{i3} are drawn only once.) Assume now that life expectancy is generated off the u_{i1} as the lognormal variates $L_i=\exp(\mu_1+u_{i1})$, where μ_1 is chosen to make the mean of these lognormal variates square with the sample mean life expectancy 6.07=193,523÷31,892. The corresponding quality-of-life scores are generated off the u_{i2} as the probits $Q_i=\Phi(\mu_2+u_{i2})$, where μ_2 is chosen to force the mean of these probits to equal the sample mean quality-of-life score 0.51. The sample correlation between the Q_i and L_i is denoted $\hat{\rho}_{QL}$. Finally, the individual-level YHLs are given by the product $Q_i \times L_i$. Three prototypical joint distributions for positive, zero, and negative ρ are displayed in figures 4-6.

The results of the simulation are summarized in table 3. Column 1 displays the population correlations ρ_{12}. Column 2 reports the sample correlations $\hat{\rho}_{QL}$. Columns 3, 4, and 5 report the sample means of the corresponding quality-of-life scores, life expectancies, and YHLs, respectively. The row marked in boldface font corresponding to $\rho_{12}=0$ provides a useful baseline reference. Here are seen an empirical $\hat{\rho}_{QL}$ correlation close to zero and sample means of Q_i, L_i, and YHL_i virtually identical to those reported in (or implied by) the last row of table 1.

The findings of primary interest are displayed in the last column for the nonzero ρ_{12} correlations. Even though the sample means of Q_i (and, of course, L_i) are essentially the same under each correlation structure, the means of YHL_i vary impressively over the different assumed correlations. Even empirical correlations $\hat{\rho}_{QL}$ on the order of ±0.4 apparently result in

sizable divergences of the mean YHLs from the value obtained under the naive zero-correlation assumption.

The bottom line is that (sub-)population heterogeneity in q and ℓ in conjunction with nonnegligible covariance between q and ℓ have potentially dramatic implications for health status measurement.

VI. CONCLUSION

For cost-effectiveness analysis, population health monitoring, and other important practical pursuits, {H,Q,D}-AL-{E,Y} measures have become the standard vehicles for quantifying outcomes. As such, it is critical to work within a methodologically rigorous framework when such measures are used for evaluative purposes. In the context of a very general stochastic framework, this paper has exposited a range of analytical tools for quantifying population health outcomes -- functionals, stochastic dominance, and parametric functions of moments, order statistics, quantiles, and tail probabilities -- and pursued in detail various features of expectations-based methods. Estimating such expectations, while *conceptually* straightforward, will often require in practice consideration of the covariance structures of G(.), thus rendering empirical implementation perhaps less straightforward than might meet the eye. Moreover, as suggested above, whether multiplicative functional forms that map attributes into health status are desirable -- from the perspective of characterizing usefully the population distribution of health -- should be a paramount consideration in designing measures of population health. Finally, it should be emphasized that only a handful of approaches to definition and estimation have been presented and assessed here, and that other approaches might be considered under particular circumstances. Whether or not the "covariance"

issues described above will pertain in such applications will depend on the particular definition and estimator, but before approaching any such alternative analytical structure, analysts might do well to consider the prospect that without due caution "lurking covariances" may ensnare empirical analyses. At a minimum, it is hoped that the analysis undertaken here will point the way to further developments in the important field of empirical population health research.

ACKNOWLEDGMENTS

Thanks are owed to Denny Fryback, Dave Kindig, Will Manning, Bernie O'Brien, Sally Stearns, Aaron Stinnett, and participants in the Population Health Doctoral Research Seminar at UW-Madison for insightful discussions, comments, and suggestions, and to Michael Mullahy for helpful research assistance. This paper was prepared for presentation at the Ninth European Workshop on Econometrics and Health Economics, Amsterdam, Netherlands, September 2000

REFERENCES

Allen, R.G.D. 1938. *Mathematical Analysis for Economists*. New York: St. Martin's Press.

Anderson, R.N. 1999. *United States Life Tables, 1997*. National Vital Statistics Reports 47 (No. 28). Washington: NCHS/CDC.

Atkinson, A.B. and F. Bourguignon. 1982. "The Comparison of Multi-Dimensioned Distributions of Economic Status." *Review of Economic Studies* 49: 183-201.

Beck, J.R. et al. 1982a. "A Convenient Approximation of Life Expectancy (The "DEALE"). I. Validation of the Method." *American Journal of Medicine* 73: 883-888.

Beck, J.R. et al. 1982b. "A Convenient Approximation of Life Expectancy (The "DEALE"). II. Use in Medical Decision-Making." *American Journal of Medicine* 73: 889-897.

Bleichrodt, H. and M. Johannesson. 1997. "The Validity of QALYs: An Experimental Test of Constant Proportional Tradeoff and Utility Independence." *Medical Decision Making* 17: 21-32.

Chaudhary, M.A. and S.C. Stearns. 1996. "Estimating Confidence Intervals for Cost-Effectiveness Ratios: An Example from a Randomized Trial." *Statistics in Medicine* 15: 1447-1458.

Cutler, D.M. and E. Richardson. 1997. "Measuring the Health of the United States Population." *Brookings Papers on Economic Activity - Microeconomics 1997*: 217-271.

Erikson, P. et al. 1995. *Years of Healthy Life*. U.S. National Center for Health Statistics, Statistical Notes No. 7. Washington: USDHHS Publication No. (PHS) 95-1237.

Fryback, D.G. 1997. "Health-Related Quality of Life for Population Health Measures: A Brief Overview of the HALY Family of Measures." Presented at the Workshop on Summary Measures of Population Health Status, Institute of Medicine, December 12, 1997.

Glasziou, P.P. et al. 1990. "Quality Adjusted Survival Analysis." *Statistics in Medicine* 9: 1259-1276.

Gold, M.R. et al., eds. 1996. *Cost-Effectiveness in Health and Medicine*. New York: Oxford University Press.

Gold, M.R. et al. 1997. "Toward Consistency in Cost-Utility Analyses: Using National Measures to Create Condition-Specific Values." *Medical Care* (forthcoming).

Hlatky, M.A. et al. 1997. "Medical Care Costs and Quality of Life after Randomization to Coronary Angioplasty or Coronary Bypass Surgery." *New England Journal of Medicine* 336: 92-99.

Institute of Medicine. 1998. *Summarizing Population Health: Directions for the Development and Application of Population Metrics*. Washington: National Academy Press.

Johannesson, M. et al. 1994. "A Note on QALYs, Time Tradeoff, and Discounting." *Medical Decision Making* 14: 188-193.

Johannesson, M. et al. 1996. "Outcome Measurement in Economic Evaluation." *Health Economics* 5: 279-296.

Kenkel, D.S. 2000. "The Costs of *Healthy People*." Mimeo: Cornell University.

Kindig, D.A. 1997. *Purchasing Population Health: Paying for Results*. Ann Arbor: University of Michigan Press.

Kuntz, K.M. and M.C. Weinstein. 1995. "Life Expectancy Biases in Clinical Decision Modeling." *Medical Decision Making* 15: 158-169.

Lamas, G. et al. 1998. "Quality of Life and Clinical Outcomes in Elderly Patients Treated with Ventricular Pacing as Compared with Dual-Chamber Pacind." *New England Journal of Medicine* 338: 1097-1104.

Litchfield, J.A. 1999. "Inequality: Methods and Tools." World Bank working paper.

Machina, M. 1988. "Generalized Expected Utility Analysis and the Nature of Observed Violations of the Independence Axiom." In P. Gardenfors and N.-E. Sahlin, eds., *Decision, Probability, and Utility: Selected Readings*. Cambridge: Cambridge University Press.

Manski, C.F. 1988. *Analog Estimation Methods in Econometrics*. London: Chapman and Hall.

Mullahy, J. and W.G. Manning. 1995. "Statistical Issues in Cost-Effectiveness Analyses." in F. Sloan, ed. *Valuing Health Care: Costs, Benefits, and Effectiveness of Pharmaceuticals and Other Medical Technologies*. New York: Cambridge University Press.

Mullahy, J. and A.A. Stinnett. 2000. "Regression Models for Quality of Life Outcomes." Draft manuscript.

Phelps, C.E. 1997. "Good Technologies Gone Bad: How and Why the Cost-Effectiveness of a Medical Intervention Changes for Different Populations." *Medical Decision Making* 17: 107-117.

Pliskin, J.S. et al. 1980. "Utility Functions for Life Years and Health Status." *Operations Research* 28: 206-224.

Shepard, D.S. and R.J. Zeckhauser. 1982. "The Choice of Health Policies with Heterogeneous Populations." in V.R. Fuchs, ed., *Economic Aspects of Health*. Chicago: University of Chicago Press for NBER.

Sox, Jr., H.C. et al. 1988. *Medical Decision Making*. Boston: Butterworths.

Stinnett, A.A. and J. Mullahy. 1998. "Net Health Benefits: A New Framework for the Analysis of Uncertainty in Cost-Effectiveness Analysis." *Medical Decision Making* 18:S68-S80.

Stinnett, A.A. and A.D. Paltiel. 1997. "Estimating C/E Ratios under Second-Order Uncertainty: The Mean Ratio versus the Ratio of Means." *Medical Decision Making* 17:483-489.

Testa, M.A. and D.C. Simonson. 1996. "Current Concepts: Assessment of Quality-of-Life Outcomes." *New England Journal of Medicine* 334: 835-840.

U.S. National Institute on Aging. 2000. *Older Americans 2000: Key Indicators of Well-Being*. Washington: Federal Interagency Forum on Aging-Related Statistics.

Willan, A.R. and B.J. O'Brien. 1996. "Cost-Effectiveness Ratios

in Clinical Trials: From Deterministic to Stochastic Models." *Health Economics* 5: 297-305.

Wolfson, M.C. 1999. "Measuring Health -- Visions and Practicalities." *Statistical Journal of the U.N. Economic Commission for Europe* 16: 1-17.

World Health Organization. 2000. *The World Health Report 2000: Health Systems: Improving Performance*. Geneva: WHO.

Zhao, H. and A.A. Tsiatis. 1997. "A Consistent Estimator for the Distribution of Quality Adjusted Survival Time." *Biometrika* 84: 339-348.

Figure 1

Prototypical Distributions (Densities) of a Scalar Health Measure

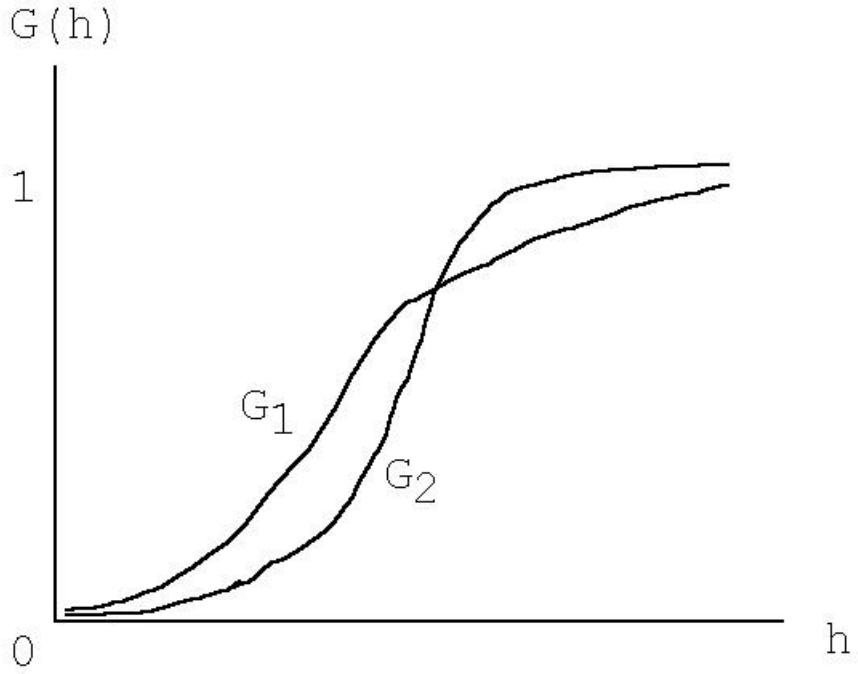

Figure 2

Quality-Adjusted Survivor Curves

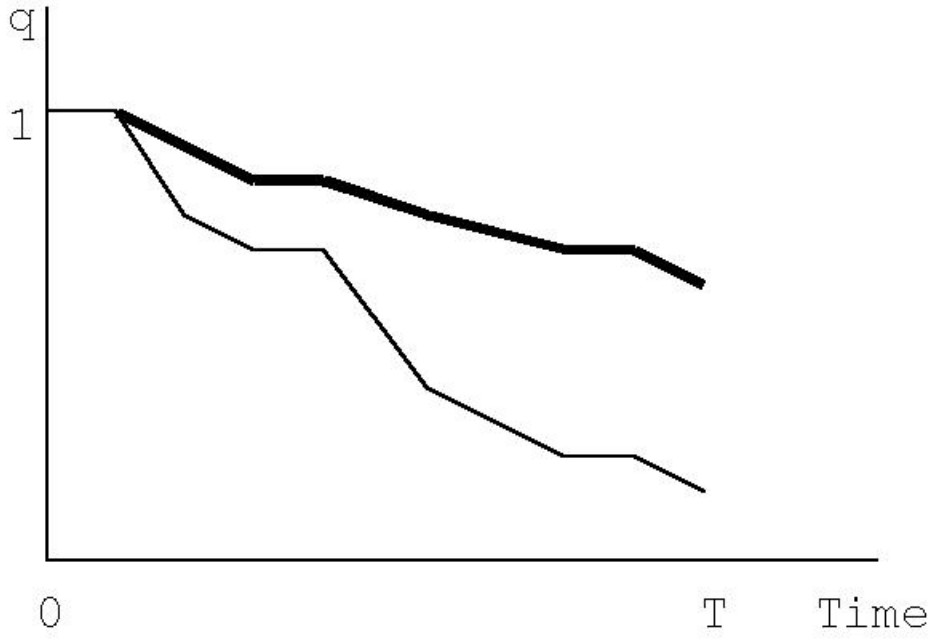

Figure 3

Joint Sample Space for (q,s) in QALY Example

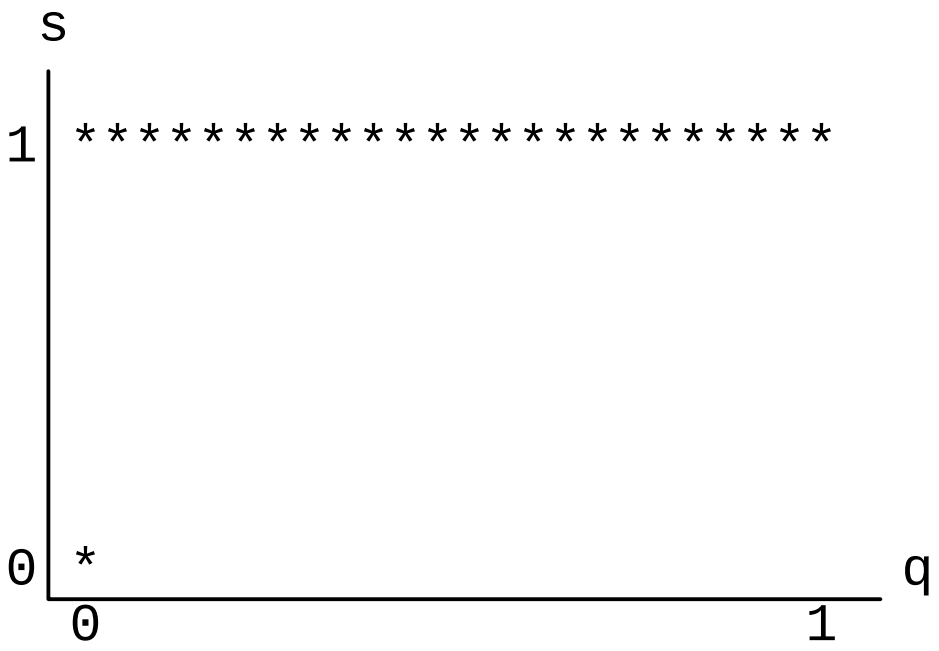

Figure 4

$g(q,\ell)$ with $\rho_{q\ell} = +.42$

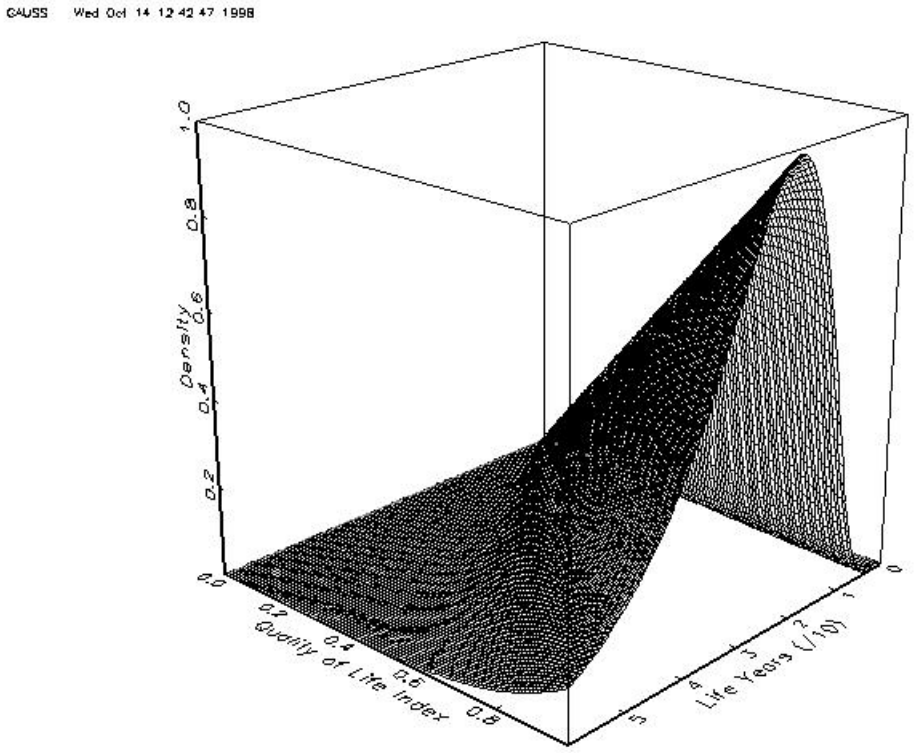

Figure 5

$g(q,\ell)$ with $\rho_{q\ell} = 0$

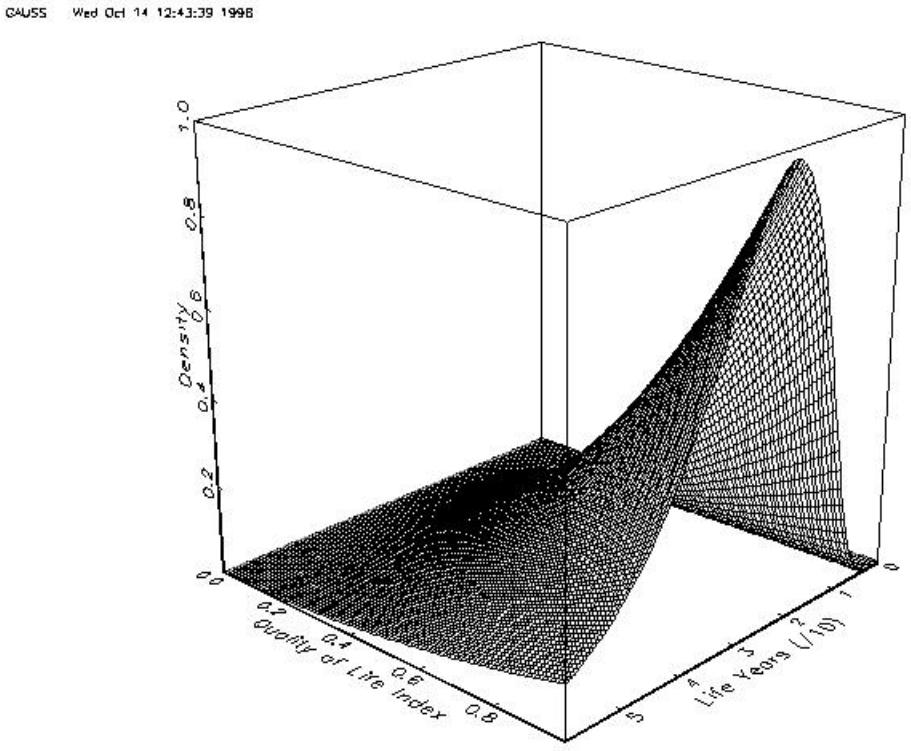

Figure 6

$g(q,\ell)$ with $\rho_{q\ell} = -.51$

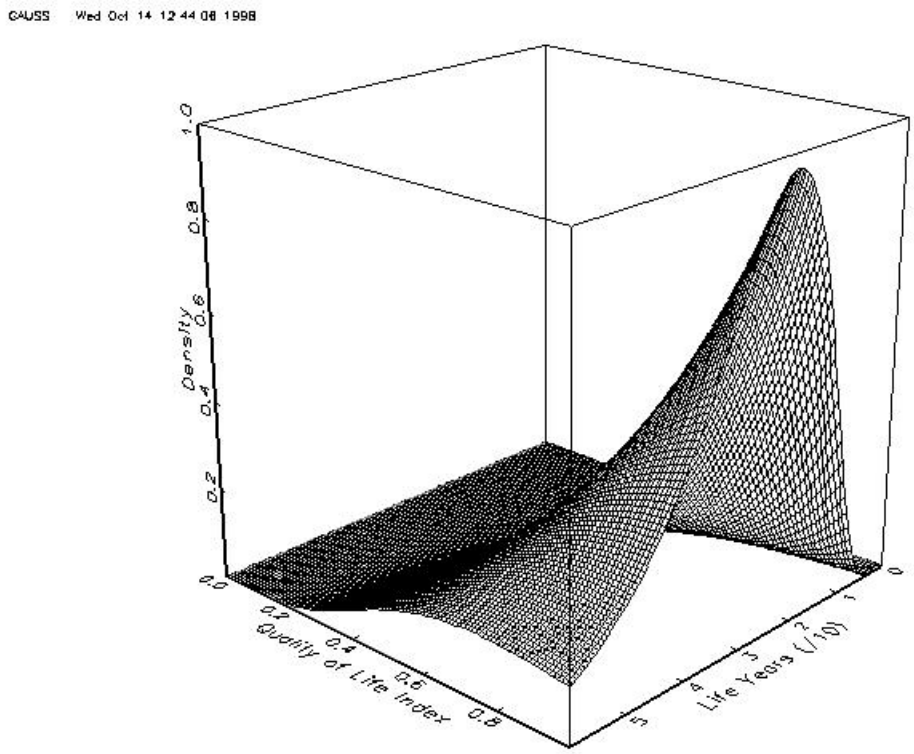

Table 1

YHL for Selected Age Intervals
(Excerpted and Abridged from Erikson et al., 1995)

				Quality-Adjusted Stationary Population...			
Age Interval	**# Living at Beginning of Interval of 100k Born Alive**	**Stationary Population in Interval**	**Average HRQOL of Persons in Interval**	**...in Interval**	**...in This and Subsequent Intervals**	**YHL Remaining**	**LY Remaining**
(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
0-5 Years	100,000	495,073	0.94	465,369	6,403,748	64.0	75.4
5-10 Years	98,890	494,150	0.93	459,560	5,938,379	60.1	75.1
:	:	:	:	:	:	:	:
80-85 Years	47,168	197,857	0.63	124,650	223,347	4.7	10.9
85 Years	**31,892**	**193,523**	**0.51**	**98,697**	**98,697**	**3.1**	**8.3**

Table 2

HALE/QALE Estimates from 1994 NHIS and 1993 U.S. Life Tables

Age	Sex	E[qxE[ℓ|A,S, Race]|A,S]	E[q|A,S]x E[ℓ|A,S]	cov(q,ℓ|A,S)	E[ℓ|A,S]	E[q|A,S]	N. Obs.
18	F	54.84	54.82	0.0270	61.40	0.89	241
	M	49.98	50.00	-0.0122	55.18	0.91	149
19	F	53.83	53.81	0.0252	60.41	0.89	313
	M	49.81	49.79	0.0141	54.08	0.92	187
20	F	53.06	53.04	0.0141	59.65	0.89	329
	M	49.58	49.58	0.0054	53.38	0.93	195
21	F	52.30	52.29	0.0038	58.65	0.89	360
	M	48.18	48.18	0.0015	52.32	0.92	230
22	F	51.30	51.27	0.0297	57.65	0.89	400
	M	47.26	47.28	-0.0105	51.44	0.92	270
23	F	50.62	50.62	0.0045	56.67	0.89	477
	M	46.26	46.23	0.0332	50.41	0.92	285
24	F	49.80	49.79	0.0143	55.79	0.89	500
	M	45.56	45.53	0.0267	49.85	0.91	291
25	F	48.95	48.94	0.0085	54.78	0.89	524
	M	44.94	44.92	0.0164	48.72	0.92	315
26	F	47.92	47.91	0.0130	53.88	0.89	531
	M	43.51	43.49	0.0255	47.66	0.91	325
27	F	47.58	47.56	0.0216	52.86	0.90	520
	M	42.47	42.44	0.0293	46.97	0.90	359
28	F	46.07	46.06	0.0085	51.91	0.89	560
	M	41.78	41.79	-0.0102	46.08	0.91	327
29	F	45.08	45.06	0.0122	50.99	0.88	606
	M	41.09	41.06	0.0287	45.11	0.91	315
30	F	43.96	43.95	0.0132	49.99	0.88	704
	M	40.19	40.18	0.0123	44.14	0.91	430
31	F	43.21	43.19	0.0220	48.95	0.88	664
	M	38.86	38.84	0.0157	43.30	0.90	401
32	F	42.15	42.13	0.0185	48.12	0.88	693
	M	38.42	38.40	0.0184	42.49	0.90	390
33	F	41.84	41.82	0.0167	47.15	0.89	682
	M	37.05	37.04	0.0118	41.58	0.89	407
34	F	40.67	40.65	0.0244	46.08	0.88	685
	M	35.94	35.90	0.0439	40.48	0.89	402
35	F	39.81	39.79	0.0164	45.28	0.88	687
	M	35.42	35.41	0.0084	39.71	0.89	412
36	F	39.16	39.15	0.0102	44.27	0.88	695
	M	33.94	33.91	0.0308	38.80	0.87	421

Age	Sex	E[q×E[ℓ|A,S, Race]|A,S]	E[q|A,S]× E[ℓ|A,S]	cov(q,ℓ|A,S)	E[ℓ|A,S]	E[q|A,S]	N. Obs.
37	**F**	37.27	37.24	0.0320	43.30	0.86	717
	M	34.03	34.02	0.0051	38.11	0.89	463
38	**F**	36.66	36.63	0.0277	42.32	0.87	696
	M	32.44	32.44	-0.0082	37.02	0.88	379
39	**F**	35.61	35.59	0.0221	41.49	0.86	681
	M	31.40	31.37	0.0330	36.16	0.87	349
40	**F**	35.31	35.29	0.0204	40.52	0.87	624
	M	30.49	30.49	-0.0009	35.30	0.86	403
41	**F**	33.60	33.57	0.0266	39.51	0.85	604
	M	29.30	29.27	0.0393	34.38	0.85	371
42	**F**	33.12	33.09	0.0282	38.61	0.86	625
	M	28.73	28.69	0.0471	33.62	0.85	363
43	**F**	32.36	32.32	0.0365	37.60	0.86	599
	M	27.45	27.44	0.0138	32.67	0.84	359
44	**F**	31.19	31.17	0.0265	36.70	0.85	552
	M	27.57	27.57	0.0075	31.77	0.87	342
45	**F**	30.41	30.38	0.0370	35.88	0.85	559
	M	26.05	25.99	0.0666	30.85	0.84	346
46	**F**	29.75	29.72	0.0295	34.89	0.85	568
	M	25.53	25.51	0.0160	30.04	0.85	320
47	**F**	28.55	28.54	0.0124	34.06	0.84	564
	M	24.85	24.83	0.0187	29.20	0.85	373
48	**F**	27.68	27.67	0.0171	33.14	0.83	468
	M	24.16	24.16	-0.0009	28.32	0.85	278
49	**F**	26.07	26.04	0.0348	32.17	0.81	439
	M	23.32	23.30	0.0124	27.56	0.85	290
50	**F**	25.92	25.90	0.0235	31.35	0.83	470
	M	22.57	22.53	0.0356	26.62	0.85	273
51	**F**	24.52	24.49	0.0291	30.36	0.81	448
	M	20.68	20.65	0.0247	25.81	0.80	311
52	**F**	23.64	23.61	0.0360	29.52	0.80	396
	M	20.45	20.46	-0.0089	24.86	0.82	270
53	**F**	23.52	23.51	0.0083	28.65	0.82	376
	M	19.67	19.64	0.0282	24.10	0.82	239
54	**F**	21.95	21.91	0.0434	27.80	0.79	386
	M	18.96	18.94	0.0216	23.26	0.81	233
55	**F**	20.98	20.93	0.0514	26.91	0.78	345
	M	17.69	17.68	0.0050	22.43	0.79	243
56	**F**	20.55	20.52	0.0376	25.96	0.79	367
	M	17.24	17.24	0.0039	21.72	0.79	242
57	**F**	19.50	19.48	0.0255	25.23	0.77	354
	M	16.58	16.57	0.0102	21.00	0.79	235

Age	Sex	$E[q \times E[\ell \mid A,S,Race] \mid A,S]$	$E[q \mid A,S] \times E[\ell \mid A,S]$	$cov(q,\ell \mid A,S)$	$E[\ell \mid A,S]$	$E[q \mid A,S]$	N. Obs.
58	F	18.78	18.75	0.0345	24.28	0.77	380
	M	16.18	16.16	0.0190	20.21	0.80	198
59	F	18.56	18.53	0.0292	23.55	0.79	376
	M	14.82	14.80	0.0186	19.42	0.76	243
60	F	17.31	17.29	0.0202	22.78	0.76	352
	M	14.34	14.32	0.0145	18.69	0.77	240
61	F	16.52	16.49	0.0279	21.95	0.75	337
	M	13.68	13.66	0.0212	17.93	0.76	214
62	F	16.20	16.18	0.0232	21.17	0.76	364
	M	12.52	12.50	0.0175	17.24	0.73	232
63	F	15.23	15.21	0.0248	20.38	0.75	382
	M	12.12	12.11	0.0089	16.61	0.73	231
64	F	15.13	15.11	0.0179	19.65	0.77	360
	M	11.73	11.72	0.0145	15.90	0.74	250
65	F	14.27	14.25	0.0204	18.82	0.76	365
	M	11.63	11.62	0.0117	15.29	0.76	261
66	F	13.91	13.88	0.0220	18.10	0.77	406
	M	11.17	11.16	0.0094	14.68	0.76	232
67	F	13.46	13.45	0.0113	17.34	0.78	391
	M	10.74	10.74	0.0004	14.02	0.77	259
68	F	12.97	12.96	0.0115	16.64	0.78	387
	M	10.19	10.18	0.0117	13.40	0.76	253
69	F	11.92	11.91	0.0079	15.97	0.75	365
	M	9.89	9.89	0.0022	12.80	0.77	223
70	F	11.52	11.52	0.0064	15.18	0.76	383
	M	9.42	9.41	0.0095	12.20	0.77	232
71	F	11.09	11.09	0.0083	14.49	0.76	353
	M	8.54	8.54	0.0039	11.60	0.74	230
72	F	10.20	10.20	0.0015	13.89	0.73	359
	M	8.34	8.33	0.0092	11.03	0.76	210
73	F	9.74	9.73	0.0047	13.19	0.74	356
	M	7.99	7.98	0.0047	10.54	0.76	221
74	F	9.37	9.36	0.0077	12.51	0.75	344
	M	7.52	7.52	0.0012	9.96	0.76	232
75	F	8.70	8.69	0.0062	11.95	0.73	304
	M	7.02	7.02	0.0018	9.44	0.74	159
76	F	8.43	8.43	0.0025	11.25	0.75	282
	M	6.71	6.71	0.0017	8.97	0.75	174
77	F	7.46	7.46	0.0048	10.65	0.70	246
	M	5.96	5.96	0.0016	8.46	0.70	163
78	F	7.43	7.43	0.0019	10.07	0.74	271
	M	5.92	5.92	0.0019	7.98	0.74	122

Age	Sex	E[q×E[ℓ|A,S, Race]|A,S]	E[q|A,S]× E[ℓ|A,S]	cov(q,ℓ|A,S)	E[ℓ|A,S]	E[q|A,S]	N. Obs.
79	**F**	6.59	6.59	0.0016	9.45	0.70	250
	M	5.70	5.70	0.0006	7.48	0.76	142
80	**F**	6.44	6.44	0.0030	8.88	0.72	225
	M	4.94	4.94	0.0016	7.09	0.70	99
81	**F**	6.17	6.17	0.0009	8.38	0.74	221
	M	4.53	4.53	0.0022	6.69	0.68	95
82	**F**	5.40	5.40	0.0009	7.79	0.69	191
	M	4.24	4.24	-0.0009	6.19	0.68	106
83	**F**	4.74	4.74	0.0019	7.29	0.65	165
	M	4.28	4.28	-0.0001	5.89	0.73	81
84	**F**	4.44	4.44	0.0000	6.80	0.65	152
	M	3.82	3.82	0.0007	5.50	0.70	61
85	**F**	4.02	4.02	0.0009	6.39	0.63	124
	M	3.58	3.58	0.0027	5.19	0.69	63

Table 3

Simulation Results for YHL Example

		Sample Means		
Population ρ_{12}	Empirical $\hat{\rho}_{QL}$	Q_i	L_i	$YHLi=Q_i \times L_i$
(1)	(2)	(3)	(4)	(5)
-0.9	-0.56	0.50	6.05	1.3
-0.5	-0.35	0.51	6.05	2.2
-0.1	-0.07	0.51	6.05	2.9
0.0	.004	0.51	6.05	3.1
+0.1	0.07	0.51	6.05	3.2
+0.5	0.35	0.51	6.05	3.9
+0.9	0.56	0.51	6.05	4.8