JOHN P. ANDERSON, ROBERT M. KAPLAN and CHRISTOPHER F. AKE

ARTHRITIS IMPACT ON U.S. LIFE QUALITY: MORBIDITY AND MORTALITY EFFECTS FROM NATIONAL HEALTH INTERVIEW SURVEY DATA 1986–1988 AND 1994 USING QWBX1 ESTIMATES OF WELL-BEING

(Accepted 29 August 2003)

ABSTRACT. *Objective*: To estimate the impact of arthritis using a general health index and National Health Interview Survey (NHIS) data. *Methods*: Morbidity data came from NHIS Public Use data, from the years 1986–1988 and 1994. The data are 423 400 cases, representing 975 421 153 person-years. Quality of Well-being Scale (QWB) morbidity scores were imputed from NHIS questions about health conditions and limitations in functioning. Both the QWB and multiple linear regression were used to estimate the effects of arthritis with and without adjustments for co-morbidity. Mortality data for NHIS-sampled adults were drawn from the National Death Index by staff of the National Center for Health Statistics. *Results*: The mean QWB for those with self-reported arthritis was 0.608 on a scale ranging from 0.0 (for death) to 1.0 (for fully functional without symptoms or problems). This observed mean for arthritis is 39.2% below the 1.000 comparison standard. QWB morbidity scores for self-reported arthritis appear (0.701 – 0.608=) 9.3% more severe than mean effects of all other health conditions. Mortality adds an average 13.8% to the morbidity burden. Persons with arthritis constitute 4.7% of the population, but account for 9.6% of Quality-Adjusted Life Years (QALYs) lost to morbidity. *Conclusions*: Self-reported arthritis is associated with very significant losses in Quality-Adjusted Life in the US population.

KEY WORDS: arthritis, National Health Interview Survey, Quality of Well-being Scale, Quality of Life

INTRODUCTION

Musculoskeletal disorders are among the most common and most disabling medical conditions. Kelsey and her colleagues analyzed the Health and Nutrition Examination Survey (HANES) data from 1971–1975 in an effort to determine the impact of musculoskeletal disorders on quality of life and role function (Kelsey et al., 1979).

Social Indicators Research **69:** 67–91, 2004.

Musculoskeletal problems were found to rank first among conditions producing "impairment – a chronic and permanent defect representing a decrease or loss of ability to perform various functions". This was true of both men and women of all age groups (Lawrence et al., 1989). Furthermore, musculoskeletal impairments were the second most common reason for physician office visits, the third most common indication for surgery, and fourth most common diagnosis for hospital admissions. The category "arthritis and rheumatism" was found to be the most common cause of disability for subjects aged 18–64 years. The total economic cost of musculoskeletal conditions was estimated at 20 billion dollars per year.

The non-fatal nature of arthritis and related conditions makes them invisible to traditional public health indicators based on mortality. Recent improvement in health outcome measures has helped portray the impact of arthritis upon function, disability, and quality of life. Over 21% of the population reporting musculoskeletal symptoms experience moderate or severe restriction of activity. Eighteen percent of these patients changed job status because of their illness (Cunningham and Kelsey, 1984). In one study, 60% of the patients with rheumatoid arthritis (RA) were considered disabled (Yelin et al., 1980). Meenan and colleagues (Meenan et al., 1981) reported that persons with RA who are able to work earned only 50% of their expected income and 63% experienced major psychosocial changes as a result of their disease. Although RA is more costly to care for than osteoarthritis (OA), OA is more common and its societal costs may be seven times higher (Lanes et al., 1997).

Outcome measures in studies of rheumatology have been difficult to evaluate because they typically focus on clinical measures such as joint tenderness, grip strength, and joint circumference (Deyo, 1988). Clinical outcome measures are very difficulty to use for estimates of population health. A US Public Health Service expert panel on cost/effectiveness in medicine and healthcare suggested that outcomes be measured using Quality-Adjusted Life Years (QALYs), which are measures of life expectancy with adjustments for quality of life (Kaplan et al., 1997; Russell, 2001). If a woman dies of lupus at age 50 and one would have expected her to live to age 75, the

disease was associated with 25 lost life years. If 100 women died at age 50 (and also had life expectancies of 75 years) 2500 (100×25 years) life years would be lost.

Death is usually not the primary outcome of concern in lupus or other musculoskeletal diseases. Many adults continue to suffer from the disease, leaving them somewhat disabled over long periods of time. Although still alive, the quality of their lives has diminished. Quality-adjusted life years take into consideration consequences of these illnesses for quality of life. For example, a disease that reduces quality of life by one half will take away 0.5 QALYs over the course of one year. If it affects two people, it will take away one QALY (equal to 2×0.5) over a one-year period. A pharmaceutical treatment that improves quality of life by 0.2 for each of five individuals will result in the equivalent of one QALY if the benefit is maintained over a one-year period. The basic assumption is that life years can be adjusted for quality of life by multiplying the time in each health state by its quality of life preference weight in order to estimate QALYs. QALYs can be added together and estimated over multiple patients and multiple years.

This system has the advantage of considering both benefits and side effects of treatment programs in terms of the common QALY units. In their report, *Summarizing Population Health*, the Institute of Medicine (IOM) of the National Academy of Science recommended that population health metrics be used to evaluate public programs and to assist the decision making process (Field and Gold, 1998).

METHODS

Overview

The purpose of this study is to estimate the impact of self-reported arthritis on morbidity (all persons) and mortality (in adults) using a general summary index and data from the National Health Interview Survey (NHIS). We offer a novel method that aggregates data across years of the NHIS to produce a national sample size of 423 400 respondents, representing 975 421 153 person-years.

QWB Scale

Over the past 30 years, the UCSD Group has planned, developed and implemented methods for measuring Quality-Adjusted Life Years as tools for attacking the health resources allocation problem. In particular, the Quality of Well-being Scale is a widely applied measure in health outcomes research. It combines preference-weighted measures of functioning with symptoms and problems to generate a point-in-time measure of Well-being (W), on a metric running from 1.0 (for asymptomatic full function) to 0.0 (for death). Elements of the QWB include the Symptom/Problem Complexes (CPX) and their calculating weights, and Function Scale steps and calculating weights. For our purposes, **W** is a Health-Related Quality of Life measure composed of CPX and function limitations. The QWB methodology has been described or applied in more than 150 published papers (see, for examples, Anderson and Moser, 1985; Kaplan and Anderson, 1988; Kaplan et al., 1984, 1989, 1994, 1995).

Underlying the QWB is the concept that illnesses, conditions, and injuries are thought to cause Symptoms and Problems, which may lead to dysfunctions (DYS) – here meaning limitations in functioning. For purposes of this study, the various words commonly used to describe health-related effects such as "disability", "restricted activity days", "bed days", etc., may be described as sub-sets of dysfunction. Dysfunction is defined as being below the top step (Not Limited) on the Mobility (MOB), Physical Activity (PAC) and/or Social Activity (SAC) Scales without reference to duration. On a particular day, an individual's QWB score is therefore composed of One (or No) Symptom weight, and 0, 1, 2, or 3 dysfunction steps.

Preference weights empirically locate how "desirable" or "undesirable" a particular situation is in QWB terms and were empirically derived for these purposes from a probability sample from the general population (Kaplan and et al., 1976; Kaplan and Bush, 1982). Preference weights (for symptoms and for dysfunction) provide the metric underlying all QWB scores. The Quality-Adjusted Life Expectancy (QALE) integrates a mortality component with **W**.

National Health Interview Survey

Data from the study come from the National Health Interview Survey (NHIS). NHIS estimates health status for U.S. civilian non-institutionalized population. The study uses a probability sample of American households. Personal interviews are conducted by trained interviewers from the U.S. Bureau of the Census. Samples are drawn weekly and, depending on the year of the survey, the total sample includes 36 000 to 47 000 households, including 92 000 to 125 000 persons. Information is obtained on the number of restricted-activity days, bed days, work- or school-loss days, and all physician visits occurring during the 2-week period prior to the week of the interview. The questionnaire also asks about acute and chronic conditions that were responsible for disabilities, limitations, or health care visits. The NHIS includes sample weight to allow generalization of survey results to the entire US population. Typically, there are 686 ethnicity-unique cells and adjustments are made in order to properly weight data for representation in the population. These weights are available in Public Use Data Tapes.

QWBX1 Development

Although the QWB methodology has been applied widely, this paper describes a new methodology for estimating the QWB from NHIS data. Health effects of any NHIS condition may be analytically traced to the symptom weights and Dysfunction weights associated with it, and thence to QWB score itself. QWB scores are preference weights for symptoms, dysfunctions, or a combination of both.

For this study, NHIS data were used to create a QWBX1 index. In order to accomplish this, NHIS person and condition files were used to identify respondents with self-reported arthritis. The QWB Social Activity Scale was matched to items on activity limitations in the NIHS. Questions on both measures are very similar. The physical activity component for the QWB was matched to NHIS items on 2-week work loss days and school loss days. The self-care component of the QWB Social Activity Scale could be estimated from NHIS items on self-care, which were added in 1982.

One of the most difficult challenges was estimating the symptoms and problems needed to calculate the QWBX1. The QWB asks

about symptoms but not about medical conditions. Conversely, the NHIS asks about conditions, but not about symptoms. The QWBX1 method requires that symptoms and problems be estimated from conditions on the basis of expert medical judgment. NHIS allows respondents to report multiple conditions. In the QWBX1, the number of specific conditions was limited to the first six reported. These constituted over 99% of the people reporting conditions. Physicians estimated the likely QWB symptoms or problems associated with each NHIS condition. This allowed estimation of each QWB component for each respondent. These components were then incorporated on the 0.0 to 1.0 continuum using standardized weights and an algorithm selecting the single Symptom/Problem with the highest standardized preference weight. The original purpose of the NHIS-QWB estimation project was development of QALY information on persons in the NHIS having illness, meaning those reporting one or more NHIS Conditions.

One of the issues in developing these estimates is that QWB is generally more sensitive than the NHIS because it emphasizes symptoms rather than medical conditions. NHIS reports 40% of the population as having a Condition, while QWB probability samples report 75–80% of the population to have one or more Symptom/Problem Complexes. To fill out what would be the normally expected CPX among those without Conditions or any function limitations, a "hot-decking" (Ford, 1980) randomization procedure was employed. CPX were assigned to persons who reported no health conditions, based on an age-related, observed frequencies basis, drawn from among probability samples. The purpose of these procedures was to adjust the overall yearly QWB total population average score to empirically reasonable levels.

Analyses

Four years of NHIS data (1986–1988 and 1994) were selected for preliminary analyses. All persons self-reporting "Arthritis" were included in the analysis. All reported analyses are based on NHIS sampling-weighted numbers, and were developed using SUDAAN software for correlated data. The population was initially divided into three analytic groups: Group 1), persons with neither arthritis nor other health condition; Group 2), persons with one or more

NHIS health Conditions, but not arthritis; and Group 3), persons with arthritis, whether or not they also have any other NHIS health Condition present.

Two approaches to the analysis were applied. First, the QWB General Health Policy Model (GHPM) was used to separate persons with co-morbidities into cases attributable to arthritis and those attributable to other NHIS conditions. Second, multiple linear regression was employed to estimate the unique association between demographic and arthritis variables and QWB.

Among persons with arthritis and co-morbidities, if the heaviest-weighted Symptom (Most Undesirable CPX, or MUCPX) present is due to arthritis, QWB-GHPM identifies QALY loss with that category. If the heaviest-weighted symptom present is due to another NHIS condition, QWB identifies QALY loss to the appropriate group. Adding QALYs from those with arthritis alone to the co-morbidity cases produces an overall estimate of QALY losses due to arthritis. Morbidity analyses in Table I compare of QWBX1 mean scores, the Symptom/Problem Complex weights and dysfunction weights, and the GHPM QALY losses (# cases – ((QWB $*$ # cases)) = QALYs Lost Overall) for each analytic Group.

Table II reports on results of dividing Group 3 into cases attributable to arthritis alone (Group 4), cases with arthritis plus other condition(s), where the heaviest-weighted Symptom is attributable to arthritis (Group 5), and cases with arthritis plus other condition(s), where the heaviest-weighted Symptom is attributable to another condition (Group 6). The combined Group 4 and Group 5 results will be presented as Group 7.

NHIS divides arthritis into several ICD-9 codes. Table III presents an estimate, using multiple linear regression procedures, of contributions by demographic characteristics along with several of the highest-frequency ICD-9 arthritis codes to QWB, symptom, dysfunction and QALY loss. Regression beta-weights and QALYs derived from them represent deviations from the mean QWB in the overall population. By contrast, the GHPM QALYs represent deviations from the 1.0 top of the QWB scale.

In regression analysis, where the variable is continuous, the standardized regression coefficient (beta) represents gain or loss on the mean for each unit increment (e.g., year) for all persons in

TABLE I

General health policy model symptom/problem complex, dysfunction and QWB means with QALY morbidity-only analyses of persons with arthritis versus other conditions for combined NHIS 1986–1988 and 1994 data

Groups	A QWB start	B Mean CPX weight	C CPX only QWB	D Mean DYS weight	E Total QWB mean	F QWB std err. of mean	G NHIS # cases (millions)	H QWB QALYs lost (millions)	I % of all % of all QALYs lost
(1) No conditions, no arthritis	1	0.113	0.887	0.000	0.887	0.00048	586.1	66.2	35.4%
(2) 1 or + conditions, no arthritis	1	0.252	0.748	0.046	0.701	0.00042	343.5	102.7	54.9%
(3) Arthritis, arth. + conditions	1	0.282	0.718	0.110	0.608	0.00109	45.8	18.0	9.6%
Mean/total	1	0.170	0.830	0.021	0.808	0.00055	975.4	186.9	100.0%

TABLE II

QALY analysis of arthritis subgroups for combined NHIS 1986–1988 and 1994 data

Analysis groups	A Total QWB mean	B NHIS # cases (millions)	C QWB QALYs lost (millions)	D % of all QALYs lost
(3) Arthritis, arth. + conditions	0.608	45.8	18.0	9.6%
(4) Arthritis, no conditions	0.767	11.3	2.6	1.4%
(5) Arthritis + conds, arthritis MUCPX	0.573	24.1	10.3	5.5%
(6) Arthritis + conds, condition MUCPX	0.516	10.4	5.0	2.7%
(7) Arthritis, arth. +, conds, arth MUCPX (Group 4 + Group 5)	0.642	35.4	12.7	6.8%

the analytic group. For the dummy (0/1) variables, the beta represents the model increment of switching from one variable value to another. Thus, converting QWB regression outcomes into QALY analyses are an application of the QWB-General Health Policy Model.

Table IV provides a comparison of GHPM and regression results, and Table V presents a t-test evaluation of GHPM changes in morbidity health effects (QWB, symptoms and dysfunction) for 1986–1988 versus 1994. Table VI compares the point-in-time QWBX1 with the Quality-Adjusted Life Expectancy (QALE), which is a combination of QWBX1 morbidity with mortality. Mortality among NHIS samples has been followed through matching with the National Death Index, a task performed by staff at the National Center for Health Statistics. This matching is complete up to December 31, 1997, and covers NHIS sampled persons aged 18 and over only.

TABLE III

Regression of independent variables with QWB, Most Undesirable Symptom/ Problem Complex (CPX) and Dysfunction (DYS) weights, NHIS years 1986–1988 and 1994 combined, with QALYs gained/lost

Independent variables	A QWB beta	B QWB SE beta	C CPX beta	D DYS beta	E # cases (millions)	F QALYs (+, –) (millions)
Intercept	91.075	0.086	9.852	–0.927		
Age	–0.254	0.001	0.194	0.060	824.4	–2.09
Education groups:						
LT high sch.	–1.264	0.093	0.258	1.006	153.9	–1.91
HS diploma	–0.117*	0.072	0.059*	0.059	310.3	–0.36
Some college	–0.715	0.071	0.524	0.191	180.0	–1.29
Race	1.333	0.071	–1.116	–0.218	127.6	+1.70
Female	–0.506	0.051	0.408	0.098	423.6	–2.14
Income groups:						
$17 999 or —	–2.744	0.091	1.104	1.640	241.6	–6.63
$18 000–$29 999	–0.850	0.078	0.323	0.527	201.8	–1.72
$30 000–$49 999	–0.360	0.072	0.136	0.224	230.0	–0.83
Hispanic	1.710	0.127	–1.159	–0.550	69.7	+1.19
Rheum arth.	–14.548	0.347	6.473	8.074	2.2	–0.32
Osteo arth.	–12.408	0.402	5.092	7.316	2.3	–0.29
Arthropathy	–7.435	0.159	3.213	4.222	19.8	–1.47
Spondylosis	–12.078	0.449	5.252	6.828	2.1	–0.25

*T-test of beta = 0 has p-value < 0.05.

RESULTS

Overall, the mean raw QWB for arthritis was 0.608, which is 39.2% below the 1.000 comparison standard. Column A of Table I shows the beginning QWBX1 score (1.0), with Column B showing the mean CPX weight for each group. Column C lists the "CPX Only" QWB scores (A – B = D), which range from 0.887 to 0.718.

TABLE IV

General health policy model and multiple linear regression model results for arthritis, weighted combined NHIS data 1986–1988 and 1994

	Deviation from 1.000 standard				
	GHPM QWB	Lower 95% CI	Upper 95% CI	Number of cases (M)	QALY loss (M)
Arth w/ co-morb.	0.608	0.606	0.610	45.8	18.0
Arth w/o co-morb.	0.642	0.640	0.644	35.4	12.8
	Regression QWB	Lower 95% CI	Upper 95% CI	Number of cases (M)	QALY loss (M)
Arth w/ co-morb.	0.675	0.673	0.678	45.8	14.9
Arth w/o co-morb.	0.718	0.715	0.721	35.4	10.0

	Deviation from 0.808 standard		
	Regression QWB beta	Number of cases (M)	QALY loss (M)
Arth w/ co-morb.	–0.13349	45.8	6.1
Arth w/o co-morb.	–0.08973	35.4	3.2

Column D reports the mean Dysfunction (DYS) weight, ranging from 0.000 to 0.110, and Column E has the total QWBX1 mean for each analytic group ((A – (B + D) = E), varying from 0.887 to 0.608. All group scores (Symptom, Dysfunction, QWB) are significantly different from one another, with the arthritis group 9.3% (–3.0% on symptoms, –6.4% on dysfunction) below the 1 or More Conditions group. Column F shows group standard error of the mean to be relatively small, and Column G has the number of cases. Column H has total QALYs (G – ((G ∗ E)) = 186.9 million) lost overall, and lost by each group. These figures integrate each Group QWBX1 mean with the number of cases, varying from 18.0 to 107.2 million QALYs lost. Column I reports the preliminary percentage of all QALYs lost by each analytic Group.

TABLE V

T-test analyses of arthritis QWB, CPX and DYS differences over time, 1986–1988 vs. 1994 NHIS years, by analysis groups, using NHIS sampling weights

Analysis groups	86–88 QWB	1994 QWB	QWB Delta	p. =	CPX Delta	p. =	DYS Delta	p. =
(2) Condit. only	0.7038	0.6945	–0.00932	0.000	0.00301	0.000	0.00631	0.000
(6) Arth + cond, cond MUCPX	0.5165	0.5132	–0.00330	0.000	0.00115	0.000	0.00214	0.000
(7) Arth, arth. + cond arth. MUCPX	0.6365	0.6323	–0.00443	0.000	0.00036	0.000	0.00407	0.000

TABLE VI

Morbidity and morbidity plus mortality comparison of analysis subgroups for combined NHIS 1986–1988 and NHIS 1994 data, adults age 18 and over

Analysis groups	A Total QWBX1 mean	B Total QALE mean	C % Dif.	D NHIS # cases represented	E Morbidity QWBX1 QALYs lost	F Morbidity & Mortality QALYs lost
(1) No conditions	0.856	0.817	3.9%	380.4	54.8	69.6
(2) Conditions, no arthritis	0.692	0.611	8.1%	268.8	82.8	104.6
(4) Arthritis, no other conditions	0.672	0.554	11.8%	10.9	3.6	4.9
(5) Arthritis + conds, arthritis MUCPX	0.573	0.421	15.2%	23.6	10.1	13.7
(6) Arthritis + conds, condition MUCPX	0.515	0.389	12.6%	10.1	4.9	6.2
(7) Arthritis, arth. + conds, arthritis MUCPX (Group 4 + Group 5)	0.605	0.463	14.2%	34.5	13.6	18.5

In Table II, Group 3 was subdivided by QWB attribution of CPX to either other condition or arthritis. We estimated that 2.6 million QALYs were lost due to arthritis alone (Group 4), while a further 10.3 million QALYs (Group 5) were lost by those with arthritis plus other conditions, where CPX was attributable to arthritis. For Group 6, with arthritis and other conditions, where CPX was attributed to another condition, 5.0 million QALYs were lost. All told, Group 7 (Groups 4 and 5 combined) showed 12.9 million QALYS were lost due to arthritis, or (12.9/186.9 =) 6.9% of all QALYs lost over the four years analyzed.

Table III displays the contribution to overall QWB scores, Symptom and Dysfunction weights, made by a range of independent variables, including having at least one of the arthritis ICD-9 recode variables. Overall, the largest effects were negative, and the strongest effects came from Rheumatoid Arthritis (–14.5%), followed by Osteoarthritis (–12.4%), and Spondylosis (–12.1%). Arthropathy, with the lowest QWB beta, had the largest overall effect, given the higher number of people affected. In arthritis, the beta for dysfunction was larger than for symptoms. The effect for demographic variables, whether positive or negative, was less than 3% for any individual characteristic.

In the upper panel of Table IV, QALY losses from the 1.000 standard estimated by GHPM for the arthritis groups were 18.0M (Group 3) and 12.8M (Group 7). In the middle panel, regression results have also been treated as deviations from the 1.000 standard, and QALY losses were 14.9M (Group 3) and 10.0M (Group 7). In the lower panel, QALY losses calculated from the overall population QWB mean comparison standard (0.808), were 6.1M and 3.2M for the two relevant patient Groups.

From t-test analysis of changes in QWB scores for the Groups over time, Table V shows QWB decreased (got worse) by 0.00932 (9 tenths of one percent) for Group 2 (conditions only, no arthritis) between 1986–1988 and 1994, with 0.00301 of the loss due to symptoms and 0.00631 of the loss due to dysfunction. For cases with arthritis plus other condition(s), where symptom was attributable to another condition (Group 6), the QWB loss was 0.00330 (3 tenths of one percent) with 0.00115 due to symptoms and 0.00214 due to dysfunction. For all QWB decreases due to arthritis (Group 7), the

loss was 0.00443 (over 4 tenths of one percent), with 0.00036 due to symptoms and 0.00407 due to dysfunction.

Table VI displays the additional health effect of including mortality with morbidity. Subtracting Column B from Column A produces an estimate of the additional percentage effect of mortality on morbidity (Column C). For the various arthritis Groups, the additional effect varies between 11.8% and 15.2%, with an average of 13.8%.

DISCUSSION

Progression of QWBX1 Development

Individual health status indicators (e.g., self-rated health, Restricted Activity Days) normally gathered by the National Health Interview Survey (NHIS) have several weaknesses. For NHIS self-rated health, about 90% of U.S. population describe themselves as being in Excellent, Very Good or Good Health. For the NHIS functional status items, over 85% generally report no Activity Limitations and over 95% report no Restricted Activity Days. These data suggest the misleading conclusion that ill health is characteristic of only a small part of the population. We believe that NHIS indicators are insensitive to many important aspects of population health status. Population findings where composite measures have been used and 75–80% of persons were reported below the top level in health status (Kaplan et al., 1976) support this position.

Principles for developing QWB-like (or other) composite measures from NHIS data were reviewed and employed (Erickson et al., 1988, 1989) in funded research in the 1980s. However, the early QWB-like composite reported an overall population health score about 10% higher than experience suggested was reasonable.

Efforts to produce composite national health estimates continued during the 1990s. One approach developed Years of Healthy Life (Erickson et al., 1995), a preference-weighted composite of NHIS self-rated health and chronic functional limitations. This measure is potentially available for the years 1984 and beyond. The major report produced by this approach (renamed Health and Limitations Index or HALex) used NHIS 1987–1992 data to estimate average weights for a large set of NHIS conditions (Gold et al.,

1998). These figures were intended to serve as national standards for aiding estimation in studies of program, policy or treatment cost-effectiveness. Another general departure led to QWBX1, using NHIS conditions to estimate QWB Symptom/Problem Complexes (CPX), combined with NHIS acute and Activity Limitations. QWBX1 has currently been calculated for NHIS years 1979–1996 (18 years, 1.9+ million persons), may be extended to earlier NHIS years, and will be calculated for later years as data become available. This paper represents one of the first applications of QWBX1 methodology.

QALY and DALY

Quality Adjusted Life Years (QALY) and Disability-Adjusted Life Years (DALY) are both attempts to characterize the combined morbidity and mortality impact of diseases on populations, but they employ fundamentally different methods.

The QALY approach, which is usually interview-based, uses a formal model of health status that can be applied across all diseases and health conditions. It is based on a set of defined health states (in the case of the QWB, a state defined by one CPX, and one step each on the Mobility, Physical Activity and Social Activity Scales). Each state is valued for its relative "desirability" or "undesirability" by a probability sample of adults, resulting in a preference weight for each state, anchored by "0" for dead and "1" for asymptomatic full function.

The QWB mean score for a disease is empirically measured as the mean among all persons with the disease at the time of interview. Among people with the disease, as the number of people in the various health states varies over time, so the mean for the disease may vary over time. Further, this method can also be used to estimate the effect of co-morbidities, as will be done in this paper. Mortality will be empirically measured in the study groups over time.

DALY, developed in the Global Burden of Disease Study (Mathers et al., 2001, 2002; Murray and Lopez, 1997), begins with specific diseases from an International Classification of Diseases list, and uses expert judgment as part of deriving consensus estimates of the burden (morbidity and mortality) of specific diseases

on given populations. If evidence exists that the burden of a specific disease varies over time (getting better or worse), a new set of expert judgments is used to take account of the new situation. The DALY method has not yet been employed in the attempt to estimate co-morbidity effects.

Many investigators present analyses using QALY outcomes without a formal QALY model and without interview-based data on persons with specific diseases (e.g., see The CDC Cost-effectiveness Group, 2002). Instead, they employ (their own) expert judgment to weight the disease conditions, and then follow normal QALY calculations to produce results. We criticize these *ad hoc* methods as producing results that are not necessarily stable – a different set of investigators, using a different *ad hoc* set of weights, can produce a totally different set of conclusions. Anchoring analyses with a stable formal model allows meaningful comparison of results from different studies of the same disease or health condition.

Other QOL Analyses of Diseases

There has been a myriad of specific-disease QOL studies, and many studies of arthritis (Bombardier et al., 1986; Tugwell et al., 2000; Fries, 1999; Wolfe and Hawley, 1997; Hurst et al., 1997; Borstlap et al., 1995; Soderman et al., 1995; March et al., 1999). It is beyond the scope of this report to review them all. However, at least three attempts at making generalizations about arthritis are worth comparing with the present effort.

In 2000, Sprangers et al., reported on a comparison of results using the SF-36 (and its variants) across a large number of studies involving multiple diseases (including rheumatoid arthritis). SF-36 is a morbidity-only measure, not a combined index of morbidity and mortality. Additionally, SF-36 cannot be combined over dimensions to produce a single expression of point-in-time health status. Thus, they conclude: "On the basis of rank-ordering among the QOL dimensions, three broad categories could be distinguished" (895), with musculoskeletal diseases in the lowest (least healthy) group. Our tools allow finer discrimination than a rank-ordering among groups of diseases.

The impact of arthritis was also compared to that of other diseases in the Beaver Dam Health Outcomes study (Fryback et

al., 1993). This population study in Beaver Dam, Wisconsin administered the QWB to 1356 adults (average age 64 years). Arthritis was the most prevalent condition affecting 598 of the participants. The mean QWB score for arthritis (0.69) was lower than that for most other chronic conditions including hypertension, hyperlipidemia, cataracts, and non-insulin dependent diabetes. However, the QWB score for those with self-reported arthritis was higher than that for congestive heart failure (0.63), myocardial infarction (0.64), macular degeneration (0.67), and depression (0.65). Although the Beaver Dam study was population based, it still had a relatively small sample.

In 1998, Gold and colleagues (Gold et al., 1998) used NHIS data for the years 1987–1992 and the Health and Limitations Index (HALex) to estimate scores for a long list of NHIS conditions. Their intent was to create national, multiple-year criterion values against which values from smaller studies of these diseases could be compared. The aim was to aid cost-effectiveness studies where investigators were unable to collect primary data. Undoubtedly this has its value, but our aim is to study one disease in more detail, characterizing contributions from CPX and dysfunction, as well as co-morbidities. Additionally, we want to know if the overall (population mean) QWB score for arthritis changes over time, and, if so, which QWB elements contribute to the change.

Interpreting the Importance of Group Differences

All of the reported QWBX1 mean differences between analytic groups were statistically significant. However, tests of statistical significance may not necessarily be particularly meaningful substantively when using very large sample sizes. With the large sample, a difference on the QWB of 0.001 (one tenth of one percent) will most likely be statistically significant. With a very large sample size, statistical significance may not be a proper guide for establishing the substantive importance of between-group or between-year differences.

Given the number of health state elements in QWB, it would be virtually impossible to specify all combinations that would produce an effect of a given size. However, it is possible to use examples to make effect sizes comprehensible. To illustrate the substantive

effect of QWB differences, let us assume two populations of 100 persons each, and each in the same QWBX1 state of 0.800, with the following exception: One group has 33 persons in Social Activity Scale step 2 (Did not perform any Major Activity, but did perform self-care). This would produce a 2% QWBX1 difference between the groups (0.800 versus 0.780). Were the number of persons limited in this way increased to 67, the QWBX1 difference would be 4% (0.800 versus 0.760), and were all persons so limited, the difference would be about 6% (0.800 versus 0.740). Preference studies show that community members can detect that states differing by about 0.03 QWB units are different from one another (Kaplan et al., 1982). States that differ by only 0.01 units are not perceived as distinct and we question whether differences of 0.02 or less should be considered meaningful.

In the case of the mortality analysis (Table VI), the additional mortality effect is by these standards substantively important across the board. In the Group with no conditions reported, the additional mortality effect was 3.9%, while for the arthritis groups it was three or four times that. Further, the fact that there were substantively important differences in mortality effect among most of the arthritis Groups argues that the formal model employed to separate them is apparently sensitive to future health possibilities. Results of these mortality analyses do not mean that these persons died of arthritis, but rather that people with self-reported arthritis died, and did so in different numbers according to our prospective classification.

General Health Policy Model Analyses and Regression Analyses

The QWB-GHPM analyses employed here (Tables I, II, IV–VI) involves preferences for specific disease states, based on the preference weight and frequency of symptoms and dysfunction experienced by people with a specific disease. The question addressed in GHPM analyses concerns how “good” or “bad” a disease is, as expressed by average preference for one disease state versus any other. Regression analysis (Table III), showed the unique association between independent variables and a dependent variable. The nexus of each question differs. An important feature of the GHPM system is that QALYs can be estimated through either GHPM analyses or employment of regression beta-weights.

However, if conventional rules for interpretation of regression results are followed, as in the lowest panel of Table IV, QALY results will always be less than would be suggested by GHPM. This is simply because GHPM results are usually interpreted as deviations from the 1.000 standard, though they can also be interpreted from the overall group mean. The group mean standard is not usually employed in GHPM analysis, since it would leave the gap between 1.000 and the group mean unexplained, as it appears to be in regression analysis.

Symptom/Problem Complexes and Functional Limitations

Though QWB Symptom/Problem Complexes and functional limitations are asked independently, they are typically related. A person with severe knee pain, for example, will typically have functional limitations as well. When the QWB preference weights were developed, a regression model entered symptoms and problems first. As a result, symptoms have a large impact on the score. One general pattern with QWB analyses is for symptoms to take the "first bite" into Well-being, which will produce mean scores, averaged across symptoms, down to about 0.700. Dysfunction weight is required to produce scores below that. Thus, unlike some other health measures, QWB is able to specify the source of Well-being losses, into Symptoms generally, and distributions of individual symptoms in analytic groups, as well as dysfunction generally, and distribution of individual Scale-steps in analytic groups.

Sources of Arthritis Effects

A major part of arthritis effects comes from large average symptom weight (–0.282) associated with arthritis, as compared with the median symptom weight (–0.252) for all non-arthritis conditions combined (where –0.257 is the median symptom weight). However, dysfunction weight is comparatively heavy with arthritis, and this starts at a very early age. One way of interpreting dysfunction scores is that an average of –0.060 indicates most people in the group dysfunctional on one QWB scale, an average of –0.120 indicates most people in the group dysfunctional on two scales, and –0.180 indicates dysfunction on three scales. By this comparative standard,

persons with arthritis at age 5–9 (–0.070) are about as dysfunctional as are persons age 55–59 (–0.063) who have conditions other than arthritis. It is fortunate that comparatively few persons of early age are so afflicted.

Early versions of a QWB-like outcome measure have been used to estimate the influence of demographic and other variables on health outcomes, as in arthritis (Kaplan et al., 1996). This type of analysis has also been employed to estimate gender effects on health status, using Health Utilities Index scores imputed for National Health and Nutrition Examination Survey I Epidemiologic Follow-Up Study (NHEFS) data (Kaplan and Erickson, 2000). An article employing QWBX1 has been published, presenting evidence that an apparent sudden increase in Well-being among NHIS respondents at age 70 and later is an artifact of NHIS procedures (Anderson, 2001). Another, analyzing morbidity and mortality health differences between males and females in the US over 1979–1996, has also been published (Kaplan et al., 2001).

One of the important limitations is that these analyses are based on self-reported arthritis. Nothing in the NHIS data set allows confirmation of a diagnosis. False positive and false negative cases of arthritis are likely and we do not know the proportion of mislabeled cases. In addition, the NHIS data do not allow the separation of different types of musculoskeletal diseases. Another important limitation is that the QWBX1 is an imputed index. Although the measure corresponds to the QWB, the exact correlation between the measures has not been evaluated for NHIS respondents.

CONCLUSIONS

Our purpose here has not been to characterize exhaustively all paths through which arthritis might have an impact on US health status. Rather, we have quantified the overall effect on morbidity in the U.S. though two somewhat different analytic approaches. None of these findings contradict those drawn from individual indicators about the seriousness of arthritis as a health problem, but they do summarize and quantify effects on a common scale. Quality-Adjusted Life

Expectancy adjusts life expectancy for losses associated with health-related quality of life and mortality. On average, those with self-reported arthritis morbidity have Health-Related Quality of Life that is 39.2% less than optimum, and mortality subtracts an additional average of 13.8%, making their Quality-Adjusted Life Expectancy 53% less than optimum.

REFERENCES

Anderson, J.P. and R.J. Moser: 1985, 'Parasite screening and treatment among Indochinese refugees: Cost-benefit/utility and the General Health Policy Model', JAMA 253, pp. 2229–2236.

Anderson, J.P.: 2001, 'Activity limitations reported by the National Health Interview Survey: An anomaly and its effect on QWBX1 estimates of national well-being' [research letter], AJPH 91, pp. 1135–1136.

Bombardier, C., J. Ware, I.J. Russell, M. Larson, A. Chalmers and J.L. Read: 1986, 'Auranofin therapy and quality of life in patients with rheumatoid arthritis. Results of a multicenter trial', American Journal of Medicine 81(4), pp. 565–578.

Borstlap, M., J.L. Zant, R.M. van Soesbergen and J.K. van der Korst: 1995, 'Quality of life assessment: A comparison of four questionnaires: for measuring improvements after total hip replacement', Clinical Rheumatology 14(1), pp. 15–20.

The CDC Diabetes Cost-effectiveness Group: 2002, 'Cost-effectiveness of intensive glycemic control, intensified hypertension control, and serum cholesterol level reduction for type 2 diabetes', JAMA 287(19), pp. 2542-2551.

Cunningham, L.S. and J.L. Kelsey: 1984, 'Epidemiology of musculoskeletal impairments and associated disability', American Journal of Public Health 74, pp. 574–579.

Deyo, R.A.: 1988, 'Measuring the functional status of patients with low back pain', Archives of Physical Medicine and Rehabilitation 69, pp. 1044–1053.

Erickson, P., J.P. Anderson, E.A. Kendall, R.M. Kaplan and T.G. Ganiats: 1988, 'Using Retrospective Data for Measuring Quality of Life: National Health Interview Survey Data and the Quality of Well-being Scale', Quality of Life and Cardiovascular Care 4, pp. 179–184.

Erickson, P., E.A. Kendall, J.P. Anderson and R.M. Kaplan: 1989, 'Using composite health status measures to assess the nation's health', Medical Care 27(Suppl), pp. S66–S76.

Erickson, P., R. Wilson and I. Shannon: 1995, Years of Healthy Life. Healthy People 2000 Statistical Notes Number 7. USDHHS, Public Health Service, Centers for Disease Control and Prevention, NCHS.

Field, M.J. and M.R. Gold: 1998, Summarizing Population Health (Institute of Medicine, National Academy Press, Washington, DC).

Ford, B.: 1980, 'An overview of hot-deck procedures', in W.G. Madow, I. Olkin and D.B. Rubin (eds.), Incomplete Data in Sample Surveys, Vol. 2.(Academic Press, New York), pp. 185–206.

Fries, J.F.: 1999, 'How may quality of life for rheumatoid arthritis patients be enhanced by current and future treatments?' *Rheumatology* 38 Suppl 2(1), pp. 35–40.

Fryback, D.G., E.J. Dasbach, R. Klein et al.: 1993, 'The Beaver Dam health outcomes study: Initial catalog of health-state quality factors', Medical Decision-Making 13(2), pp. 565–578.

Gold, M.R., P. Franks, K.I. McCoy and D.G. Fryback: 1998, 'Toward consistency in cost-utility analyses: Using national measures to create condition-specific values', Medical Care 36: 778–792.

Hurst, N.P., P. Kind, D. Ruta, M. Hunter and A. Stubbings: 1997, 'Measuring health-related quality of life in rheumatoid arthritis: Validity, responsiveness and reliability of EuroQol (EQ-5D)', British Journal of Rheumatology 36(5), pp. 551–559.

Kaplan, R., J.W. Bush and C.C. Berry: 1976, 'Health status: Types of validity and the index of well-being', Health Serv. Res. 11, pp. 478–507.

Kaplan, R.M. and J.W. Bush: 1982, 'Health related quality of life measurement for evaluation research and policy analysis', Health Psychol., pp. 1–61.

Kaplan, R.M., J. Atkins and R.M. Timms: 1984, 'Validity of a Quality of Well-being Scale as an outcome measure in chronic obstructive pulmonary disease', J Chron Dis 37, pp. 85–95.

Kaplan, R.M. and J.P. Anderson: 1988. 'A general health policy model: Update and applications', Health Services Research 23, pp. 203–235.

Kaplan, R.M., J.P. Anderson, A. Wu, W.C. Matthews, F. Kozin and D. Ornstein: 1989, 'The Quality of Well-being Scale: Applications in AIDS, cystic fibrosis and arthritis', Med Care 27(Suppl), pp. S27–S43.

Kaplan, R.M., J.A. McCutchan, A.M. Navarro and J.P. Anderson: 1994, 'Quality adjusted survival analysis: A neglected application of the Quality of Well-being Scale', Psychology and Health: An International Journal 9, pp. 131–141.

Kaplan, R.M., J.P. Anderson, T.L. Patterson, J.A. McCutchan J.D. Weinrich, R.H. Heaton, J.H. Atkinson, L. Thal, J. Chandler and I. Grant: 1995, 'Validity of the Quality of Well-being Scale for persons with HIV infection', Psychosomatic Medicine 57, pp. 138–147.

Kaplan, R.M., J.E. Alcaraz, J.P. Anderson and M. Weisman: 1996, 'Quality-adjusted life years lost to arthritis: Effects of fender, race, and social class', Arthritis Care and Research 9, pp. 473–482.

Kaplan, R.M., W.J. Sieber and T.G. Ganiats: 1997, 'The Quality of Well-being Scale: Comparison of the interviewer-administered version with a self-administered questionnaire', Psychology & Health 12, pp. 783–791.

Kaplan, R.M. and P. Erickson: 2000, 'Gender differences in quality-adjusted survival using a Health-Utilities Index', Am J Prev Med 18, pp. 77–86.

Kaplan, R.M., J.P. Anderson and C.F. Ake: 2001, 'Gender differences in Quality-Adjusted Life Expectancy: Results from the National Health Interview Survey', Clinical Journal of Women's Health 1, pp. 191–198.

Kelsey, J.L., A.A. White, H. Pastides and G.E. Bisbee, Jr.: 1979, 'The impact of musculoskeletal disorders on the population of the United States', Journal of Bone and Joint Surgery. American 61, pp. 959–964.

Lanes, S.F., L.L. Lanza, R.W. Radensky et al.: 1997, 'Resource utilization and cost of care for rheumatoid arthritis and osteoarthritis in a managed care setting: The importance of drug and surgery costs', Arthritis and Rheumatism 40, pp. 1475–1481.

Lawrence, R.C., M.C. Hochberg, J.L. Kelsey et al.: 1989, 'Estimates of the prevalence of selected arthritic and musculoskeletal diseases in the United States', Journal of Rheumatology 16, pp. 427–441.

March, L.M., M.J. Cross, H. Lapsley et al.: 1999, 'Outcomes after hip or knee replacement surgery for osteoarthritis. A prospective cohort study comparing patients' quality of life before and after surgery with age-related population norms [see comments]', *Medical Journal of Australia* 171(5), pp. 235–238.

Mathers, C.D., C.J. Murray, A.D. Lopez, R. Sadana and J.A. Salomon: 2002, 'Global patterns of healthy life expectancy for older women', J Women Aging 14(1–2), pp. 99–117.

Mathers, C.D., R. Sadana, J.A. Salomon, C.J. Murray and A.D. Lopez: 2001, 'Healthy life expectancy in 191 countries, 1999', Lancet 357(9269), pp. 1685–1691.

Meenan, R.F., E.H. Yelin, M. Nevitt and W.V. Epstein: 1981, 'The impact of chronic disease: A sociomedical profile of rheumatoid arthritis', Arthritis and Rheumatism 24, pp. 544–549.

Murray, C.J. and A.D. Lopez: 1997, 'Alternative projections of mortality and disability by cause 1990–2020: Global burden of disease study', Lancet 349(9064), pp. 1498–1504.

Russell, L.B.: 2001, 'The methodologic partnership of effectiveness reviews and cost-effectiveness analysis', Am J Prev Med. 20(3 Suppl), pp. 10–12.

Söderman, P. and H. Malchau: 2000, 'Validity and reliability of Swedish WOMAC osteoarthritis index: A self-administered disease-specific questionnaire (WOMAC) versus generic instruments (SF-36 and NHP)', Acta Orthopaedica Scandinavica 71(1), pp. 39–46.

Sprangers, M.A.J., E.B. de Regt, F. Andries et al.: 2000, 'Which chronic conditions are associated with better or poorer quality of life', J. Clin. Epi. 53, pp. 895–907.

Tugwell, P., G. Wells, V. Strand et al.: 2000, 'Clinical improvement as reflected in measures of function and health-related quality of life following treatment with leflunomide compared with methotrexate in patients with rheumatoid

arthritis: Sensitivity and relative efficiency to detect a treatment effect in a twelve-month, placebo-controlled trial. Leflunomide Rheumatoid Arthritis Investigators Group', Arthritis and Rheumatism 43(3), pp. 506–514.

Yelin, E., R. Meenan, M. Nevitt and W. Epstein: 1980, 'Work disability in rheumatoid arthritis: Effects of disease, social, and work Factors', Annals of Internal Medicine 93, pp. 551–556.

Division of Health Care Sciences John P. Anderson
Department of Family and Preventive Medicine Robert M. Kaplan
University of California at San Diego
La Jolla, CA 92093-0622
USA
E-mail: jpa@ucsd.edu

Veterans' Administration Medical Center Christopher F. Ake
La Jolla, CA
USA