Quality of Life Research, 5, pp. 162–164

Comparison of two methods of calculating Quality-adjusted Life Years

T. G. Ganiats,* D. K. Browner and R. M. Kaplan

University of California San Diego (UCSD) School of Medicine Department of Family and Preventive Medicine and the UCSD Health Outcomes Assessment Program, San Diego, CA, USA

This paper compares two methods for calculating QALYs using quality of life data from a clinical trial. The methods produced similar results in the population as a whole, but they gave different results in a large subset. Different methods for calculating QALYs may give different results, and care should be taken to select the correct method.

Key words: Quality-adjusted life years; quality of life; Quality of Well-being scale.

Introduction

Cost-effectiveness considerations are now common in evaluations of medical therapies, surgeries and new pharmaceutical products. The Province of Ontario and the Australian government have proposed using cost-effectiveness analysis to obtain better informed decisions in formulary purchases. At least one state (Oregon) proposed using cost-effectiveness analysis as a method for prioritizing services to be reimbursed under their Medicaid programme. The Ontario and Oregon proposals confronted significant obstacles, in part because the methodology for cost-effectiveness analysis is still emerging.

The quality-adjusted life year (QALY) has been proposed as a standard outcome measure for cost-effectiveness analysis.[1] The approach adjusts survival for the quality of life during years prior to death. Most studies using QALYs apply the measurement to a hypothetical group of subjects. For example, Stason and Weinstein use their judgment to assign QALYs to different health states in their classic article on the treatment of hypertension.[2] Others calculate QALYs by explicitly making both quality of life and time part of the utility judgement.[3,4] This paper focuses on a third method where quality of life is prospectively evaluated and combined with time to estimate QALYs.

One way to quantify quality of life is to classify individuals into observable levels of functioning, consider self-reported symptoms, and apply weights on a scale ranging from 0–1.0 to reflect the desirability of the observed states. The Quality of Well-being (QWB) scale is a general health status index that includes preference-weighted measures of symptoms and three levels of functioning. It is administered through a structured interview. The reliability and validity of the instrument, along with the preferences weights, have been reported previously.[5,6]

By integrating the results of repeated measurements of well-being, these serial measurements of quality of life over the course of time allow the estimation of duration of stay in states and ultimately can be used for the calculation of QALYs. There are at least two techniques for calculating the QALY output in a study where quality of life measurements are obtained prospectively. In this paper we use data from a prospective clinical trial of stroke prophylaxis to compare two different techniques for incorporating quality of life measurements in the calculation of QALYs.

Methods

The health-related quality of life of subjects participating in two large, multicentre trials was estimated by intermittent health status assessments. The National Institutes of Health-funded Stroke Preven-

Supported by grants RO1-HS-06098 from the Agency for Health Care Policy and Research, RO1-NS-24224 from the National Institute of Neurological Disorders and Stroke and from the UCSD Health Outcomes Assessment Program.

* To whom correspondence should be addressed at Division of Family Medicine-0807, UC San Diego School of Medicine, 9500 Gilman Drive, La Jolla, CA 92093-0807. Telephone: 619-543-6393.

tion of Atrial Fibrillation (SPAF II) study was a multicentre, randomized trial that addressed the relative effectiveness of stroke prophylaxis with aspirin or warfarin in patients with nonrheumatic atrial fibrillation.[7] The Cost/Utility of Stroke Prevention (CUSP) project was an Agency for Health Care Policy and Research-funded health outcome assessment of subjects participating in SPAF II. Health outcomes were estimated by intermittent QWB interviews.[8]

The QALY output for each subject was calculated using two different techniques. In the first, the subject's quality of life was assumed to change constantly in a linear fashion between interviews (Figure 1-A, 'Average'). In the second method (Figure 1-B, 'Early'), the subject was assumed to maintain the same QWB score from one interview until the subsequent interview.

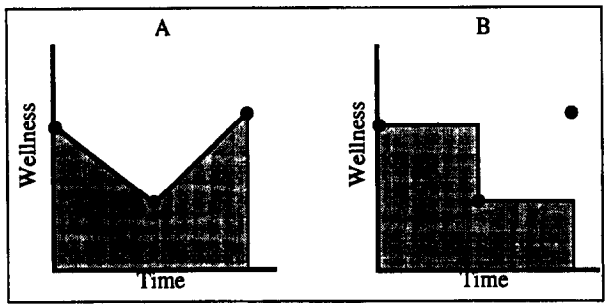
To evaluate whether the results were dependent on the degree of change in health status during the trial, the analysis was repeated on a subset of the subjects that experienced the greatest change in health status while in the study. After plotting the QWB score for each subject as a function of time (a time-wellness plot), we calculated the slope of the linear regression line for the time-wellness plot for each subject. The QALY output using the 'Average' and 'Early' methods was again calculated for the 5% sample with the greatest decrease in health status, as estimated by the linear regression, during the trial. A paired *t*-test was performed on the results from the two different techniques.

Results

A total of 956 subjects had 5,783 interviews over a period of up to 4.09 years. Twenty-seven subjects had only one interview before they either died or withdrew from the study, and these subjects were not included in the calculations. The average QALY output of the subjects was 1.74 and 1.76 using the 'Average' and 'Early' methodologies, respectively. This difference is not considered to be clinically significant.

Despite the similarity of the two methods in measuring the mean QALY output of the population, there were wide individual variations between methodologies. The difference between the scores calculated by the two methodologies varied from -10.3% to 34.3%. The subset of 47 subjects with the greatest decrease in the health status had 199 interviews over a period of up to 3.09 years. The average QALY output of these subjects using the 'Early' method was almost 10% higher than in the same subjects using the 'Average' method (0.943 QALY vs. 0.860 QALY, $p < 0.000$).

Figure 1. Two methods for calculating the area under points on the Time-Wellness plot. The 'Average' method assumes a constant change in quality of life over time. The 'Early' method assumes the subject's quality of life remains constant between assessments. • represents an individual's score at a given time.



Conclusions

Health-related quality of life is steady over the first few decades of life and slowly falls in the later decades. However, in prospective clinical trials subjects are likely to have greater changes in health status over time than would be expected in the general population. For example, if an intervention is successful, the subjects may experience improved health. Conversely, the subjects may be sicker than the general population and may experience greater morbidity or mortality. In this case the overall health status of the study population will decrease faster than that of the general population. Given these possible variations, the method of calculating QALYs may be important. For example, if health status improves, the 'Average' method may understate the magnitude of the improvement; if health status declines, the 'Early' method may understate the magnitude of the decline.

In addition, the method of sampling the subjects' well-being may be important. If the health status assessment occurs at regular intervals and the changes in health status are random, the 'Average' method may be the more appropriate technique. If health status assessment is keyed to known sudden changes in the health status of the subject (e.g., an interview occurs whenever the subject is hospitalized), then the 'Early' method may be more appropriate.

In the current study there was no significant difference between the methodologies when the entire subject population was evaluated. This is either because during the trial the health status of the

subjects remained unchanged or because those with improving health status balanced those who experience a deterioration of health. The fact that the different methodologies for calculating QALYs produced different results for those with declining quality of life scores indicates that the issue is of potential importance in any clinical trial.

The method of calculating QALY may be a key issue in selected cases. Until more work is done to classify which method is appropriate for different study populations, investigators should be aware of the potential importance of this factor. Investigators should either determine *a priori* which methodology is the most appropriate for their clinical trial or report the results of a sensitivity analysis that utilizes both methodologies.

Acknowledgement

The authors wish to thank Andrea Halverson for her assistance in evaluating the two different models.

References

1. Kaplan RM, Anderson JP, Ganiats TG. The Quality of Well-being scale: Rationale for a single quality of life index. In: Walker SR, Rosser RM, eds. *Quality of Life Assessment: Key Issues in the 1990s*. 2nd ed. London: MTM Press, 1993: 65–94.
2. Stasin WB, Weinstein MC. Public-health rounds at the Harvard School of Public Health. Allocation of resources to manage hypertension. *New Eng J Med* 1977; **296**: 732–739.
3. Mehrez A, Gafni A. The healthy-years equivalents: how to measure them using the standard gamble approach. *Med Dec Making* 1991; **11**: 140–146.
4. Torrance GW. Utility approach to measuring health-related quality of life. *J Chronic Dis*. 1987; **40**: 593–600.
5. Kaplan RM, Bush JW. Health-related quality of life measurements for evaluation research and policy analysis. *Health Psych* 1982; **1**: 61–80.
6. Kaplan RM, Bush JW, Berry CC. The reliability, stability and generalizability of a health status index. In: Society AS, ed. Proceedings of the Social Statistics Section. Washington, DC: American Statistical Society, 1978.
7. The Stroke Prevention in Atrial Fibrillation Investigators. Design of multicenter randomized trial for the stroke prevention in atrial fibrillation study. *Stroke* 1990; **21**: 538–545.
8. Ganiats T, Palinkas L, Kaplan R. Comparison of Quality of Well-being scale and Functional Status Index in patients with atrial fibrillation. *Med Care* 1992; **30(10)**: 958–964.

(Received 7 July 1995;
accepted 21 August 1995)