STATISTICS IN MEDICINE
Statist. Med. **18**, 1627–1640 (1999)

MONTE CARLO ESTIMATION OF EXTRAPOLATION OF QUALITY-ADJUSTED SURVIVAL FOR FOLLOW-UP STUDIES

JING-SHIANG HWANG[1]* AND JUNG-DER WANG[2]

[1] *Institute of Statistical Science, Academia Sinica, Taipei 115, Taiwan*
[2] *Center for Research of Environmental and Occupational Disease, Institute of Occupational Medicine and Industrial Hygiene, College of Public Health, National Taiwan University, No. 1, Sec. 1, Jen-Ai Rd, Taipei 100, Taiwan*

SUMMARY

The expected quality-adjusted survival (QAS) for an index population with a specific disease can be estimated by summing the product of the survival function and the mean quality of life function of the population. In many follow-up studies with heavy censoring, the expected QAS may not be well estimated due to the lack of data beyond the close of follow-up. In this paper, we first created a reference population from the life tables of the general population according to the Monte Carlo method. Secondly, we fitted a simple linear regression line to the logit of the ratio of quality-adjusted survival functions for the index and reference populations up to the end of follow-up. Finally, combining information on the reference population with the fitted line, we predicted the expected quality-adjusted survival curve beyond the follow-up period for the index population. Simulation studies have shown that the simple Monte Carlo estimation procedure is a potential approach for estimating expected QAS and the survival function beyond the follow-up with a certain degree of accuracy. Copyright ©1999 John Wiley & Sons, Ltd.

1. INTRODUCTION

Quality of life (QOL) and quality-adjusted life year (QALY), as measures of outcome evaluation in health contexts, have been increasingly used in the analysis of health-related studies, especially cost-effectiveness analysis, in recent years.[1,2] The QOL is often measured by utility scale or health profile and then summarized to reference points between 0 and 1; perfect health is assigned a weight of 1, and a state equivalent to being dead a weight of 0.[3] An individual's quality-adjusted survival (QAS) is given as the integration of the patient's utility through his/her survival duration. Hwang *et al.* treated a patient's QOL as a stochastic process $q(t|T, x)$, $t \in [0, T]$, where T is the survival time and x is a covariate vector representing a specific cohort or population with the index disease.[4] The QAS time of the patient, measured in quality-adjusted survival years or

* Correspondence to: Jing-Shiang Hwang, Institute of Statistical Science, Academia Sinica, Taipei 115, Taiwan. E-mail: jshwang@stat.sinica.edu.tw

Contract/grant sponsor: National Science Council of ROC, Taiwan

CCC 0277–6715/99/131627–14$17.50

Received January 1998
Accepted October 1998

months, is therefore represented by

$$\mathrm{QAS}(x) = \int_0^T q(t|T, x)\,\mathrm{d}t. \tag{1}$$

If patients are assumed to experience finite k health states which differ in their QOL, such as in Glasziou *et al.*, the QAS is given as

$$\mathrm{QAS}(x) = \sum_{i=1}^{k} q_i s_i$$

where $q_1, \ldots, q_k$ are the utilities assigned to each of k health states and $s_1, \ldots, s_k$ are the times spent in each of the states.[5]

To estimate the expected QAS for a specific disease population or subpopulation with covariate vector x, Hwang *et al.* derived a simple approximation

$$E[\mathrm{QAS}(x)] \approx \int_0^\infty E[q(t|x)]S(t|x)\,\mathrm{d}t$$

where $E[q(t|x)]$ is the mean QOL at time t after onset for the disease subpopulation with covariate vector x, and $S(t|x)$ is the survival function of that subpopulation.[4] The estimated survival function, denoted by $\hat{S}(t|x)$, can be obtained by applying the commonly used methods such as life table, Kaplan–Meier methods and parametric models to available survival data. Patients' QOL utilities at some time points can be obtained through a cross-sectional survey. The estimated $E[q(t|x)]$, denoted by $\hat{q}(t|x)$, is calculated using kernel smoothing techniques or fitting a non-linear curve to the QOL survey data. In a simulation study, Hwang *et al.* demonstrated that the area under the *quality-adjusted survival curve*, defined by $\widehat{\mathrm{qasc}}(t|\mathrm{x}) = \hat{q}(t|x)\hat{S}(t|x)$, is a good estimate of the expected QAS for a specific population.[4] However, this estimator of expected QAS is limited only to acquisition of QOL survey data and a complete follow-up for survival. In many chronic diseases commonly encountered, the life expectancy may be very long. For example, patients with diabetes mellitus,[6] hypertension[7] or papillary thyroid cancer[8] frequently survive for more than 15–20 years if early recognized and carefully treated. Thus, one may be faced with sets of survival data with very high censoring rates, say more than 80 per cent. Then, the QAS estimation can only be computed up to the end of follow-up, say 5 years, instead of the whole life span.

Because of lack of data on both survival and QOL after censoring, the QAS method cannot be applied for clinical and public health decisions directly. While the mean QOL for the whole time span might be extrapolated by assigning some constant utility for the time beyond the follow-up, accurate survival estimates from heavy censoring data are not easily obtained. Parametric extrapolation of survival estimates beyond the follow-up limits is a common approach. Gelber *et al.* proposed an estimator that is a composite of the Kaplan–Meier product limit estimate and a parametric estimate of the tail of the survival function.[9] The estimator is especially useful whenever a parametric model could be feasibly fitted to the tail rather than to the entire survival curve. Mark *et al.* proposed another composite modelling approach for estimating survival rates after the end of follow-up for the Global Utilization of Streptokinase and Tissue Plasminogen Activator for Occluded Coronary Arteries (GUSTO) study, which extended 1-year survival data by an additional 14 years from the Duke Cardiovascular Disease Database and used a Gompertz function to extrapolate the tail of the survival curve.[10] The direct substitution of the

survival may produce accurate results to some degree. However, it works only when the relevant database is available.

In this paper, we propose a more feasible approach to project quality-adjusted survival estimates beyond the follow-up, especially when the data have been heavily censored. The main idea of this approach is to borrow information from a reference population, of which the survival function is easily obtained from some available life table data such as a table of vital statistics. This approach consists of roughly three phases. First we create a reference population with survival function estimated according to the Monte Carlo method from a population with known hazard functions. Second, we fit a simple linear regression to the logit transform of the ratio of QAS curves for both the index and the reference populations up to the end of follow-up. Finally, the estimated regression line and survival curve of the reference populations are used to estimate the entire quality-adjusted survival curve, and therefore the projected long-term QAS beyond the follow-up limit.

The general data structure for these studies is given in Section 2. Section 3 describes detailed modelling and estimating procedures. In Section 4, simulation studies are designed to mimic practical follow-up studies. The performance of the proposed approach is evaluated through these simulation studies. The potential applications of the proposed approaches and their limitations are discussed in the final section.

2. DATA STRUCTURE

2.1. Follow-up data for an index population

The methodology has been developed to estimate the expected QAS beyond the follow-up period for an index population with a disease or an injury based on a typical medical follow-up study that allows for arbitrary censorship and opportunity for quality of life interviews. Suppose that by the end of the study N patients have been recruited to form a sample of the index population. For the ith patient in the case sample, let Y_i denote the duration time since onset to the current study time or end of the follow-up study. If the patient is still alive, we assign the censor status variable $\delta_i = 0$; otherwise, Y_i represents complete survival time. Usually we have already collected covariate vector z_i describing the patient's characteristics such as sex, onset age, race, family history and social status etc. We have also assumed that a random cross-sectional survey has been conducted on patients still alive, to collect a sample of V_i and q_i, where V_i is the duration time since onset to the interview, and q_i is the measured utility of the patient's quality of life at time V_i. The estimated survival function can be estimated using the Kaplan–Meier method and denoted by $\hat{S}(t|\text{index})$ for $t \geqslant 0$. The mean quality of life function can also be estimated using the kernel smoothing method, as described in Hwang *et al.*, and denoted by $\hat{q}(t|\text{index})$ for $t \geqslant 0$.[4] Note that the estimates of both survival and mean QOL are usually reliable only up to the end of follow-up.

2.2. Simulated data for a reference population

The first phase of our estimation procedures is to create an informative reference population. The selection of reference subjects is similar to a cohort study, where one usually chooses a group according to comparability of effects, contrasted populations and information.[11] For feasibility considerations, we suggest at least matching the gender and onset age. This is because vital statistics of a nation's general population are more readily accessible. If certain sources of

reference subjects are available for better comparability of potential confounders, we may consider those alternatives instead.

The idea of borrowing information from the general population to improve estimation of expected life in survival studies with incomplete follow-up data was proposed, for example, by Hakama and Hakulinen.[12] We have integrated this idea with the Monte Carlo techniques to establish a base for the first phase of estimation procedures for expected quality-adjusted survival. The detailed procedures of this phase for constructing a reference population from the general population are described as follows.

For the ith patient in the sample of the index population, we choose a person with the same covariate z_i, mainly onset age and gender, from the general population with known hazard function in vital statistics to form a reference sample. The survival time of the selected individual in the reference population is then generated according to the Monte Carlo method based on the hazard function of the individual's matched gender and onset age. For example, a survival time of a reference subject corresponding to a male patient of age x may be generated as follows. From the life table of the general population, we first find out p_{x+k}^{x+k+1}, the proportion of male persons alive at the beginning of age interval $(x+k, x+k+1)$ but dying during the interval for $k \geqslant 0$.[13] The conditional survival function of the male general population who have survived to age x is given by $S(t|x) = \prod_{k=0}^{t}(1 - p_{x+k}^{x+k+1})$, for $t > 0$, and $S(0|x) = 1$. Secondly, a uniform random number within zero and one is generated. The time t_x such that $S(t_x|x)$ equals the uniform random number is a survival time for the reference population.

The survival curve, denoted by $\hat{S}(t|\text{ref})$, of the reference population is then obtained by applying life table or product limit methods to the simulated survival times. In practical application, $\hat{S}(t|\text{ref})$ should be greater than $\hat{S}(t|\text{index})$. If there are quality of life data available for the general population, we may use similar techniques to obtain an estimated mean quality of life function, denoted by $\hat{q}(t|\text{ref})$. Usually we would simply assign $\hat{q}(t|\text{ref})$ a proper constant close to 1 to ensure the quality-adjusted survival curve for reference population $\widehat{\text{qasc}}(t|\text{ref}) = \hat{q}(t|\text{ref})\hat{S}(t|\text{ref})$ is above the estimated curve $\widehat{\text{qasc}}(t|\text{index})$ for the index population.

3. MODEL AND ESTIMATION

Since both index and reference populations have the same distribution for the covariate vector, which includes the factors affecting people's survival and QOL, the ratio of $\widehat{\text{qasc}}(t|\text{index})$ and $\widehat{\text{qasc}}(t|\text{ref})$, denoted by $W(t)$, is assumed to behave in a stable manner after a period of stage of which the disease was first noticed and more invasive diagnostic and/or therapeutic procedures were carried out to study the extent of the disease or to surgically remove the lesion, and patients were usually under psychological stress and physical discomfort as well as higher risk of mortality because of these invasive procedures. In most practical situations, especially in populations with chronic diseases, $W(t)$ will slowly decrease or remain constant after t is larger than this critical or unstable time point T_s. Meanwhile the hazard and quality of life are usually worse in the index population if compared with the reference population, therefore, we may assume that $W(t)$ is between 1 and 0.

In a follow-up study, we can confidently estimate $W(t)$ only up to some specific time T_f, which is usually the end of the follow-up. Assuming that $W(t)$ is stabilized after the unstable stage, we may use the predicted $W(t)$ and estimated $\widehat{\text{qasc}}(t|\text{ref})$ to estimate $\widehat{\text{qasc}}(t|\text{index})$ for $t > T_f$. More precisely, we apply a logit transform to $W(t)$ to extend the range from negative infinity to positive infinity, which approximates a straight line, except for the two tail ends. The left extreme tail

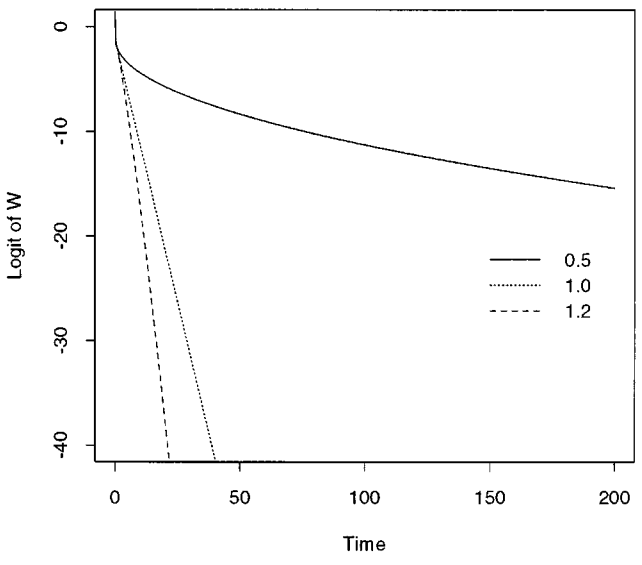

Figure 1. Plots of logit of $W(t)$ for the proportional hazard examples with survival time in the reference populations having Weibull distributions with scale value one and shape values 0·5, 1·0 and 1·2 and $\exp(Z\theta) = 2$

corresponds to the period between onset and some unstable stage to behave more stably at time T_s, indicating the necessary time for stabilization. The right extreme tail corresponds to the time of near death, when the survival function is near zero. The central part of the logit curve is close to a straight line.

To illustrate the linearity of the logit transform of the ratio of these two quality adjusted survival functions, we assume that the hazard function for index population is proportional to the hazard function for the reference population. For example, we may have $S(t|\text{index}) = S(t|\text{ref})^{\exp(Z\theta)}$, where Z is a covariate vector and θ is the model parameter vector. Also assuming that the ratio of mean quality of life functions between index and reference populations is a slowly decreasing function with time, for example $0{\cdot}8\exp(-t^{0{\cdot}01})$, then we have $W(t) = 0{\cdot}8\exp(-t^{0{\cdot}01})$ $S(t|\text{ref})^{\exp(Z\theta)-1}$. Suppose that the survival times of the reference population follow the popular Weibull distributions with a scale parameter value of one and shape parameter values $\gamma = 0{\cdot}5, 1{\cdot}0, 1{\cdot}2$ indicating decreasing, constant and increasing hazard rates, respectively. In Figure 1 we can see clear linearity of the logit of $W(t)$ for $t > 0$, expect for a short period at the beginning.

The linearity property provides us with an alternative and easy way for projecting the survival and quality-adjusted survival estimates beyond the follow-up period. Therefore, we propose fitting a simple linear regression to logit of $W(t)$ for $t \in [T_s, T_f]$, that is

$$\log\left(\frac{W(t)}{1-W(t)}\right) = \alpha + \beta t + N_t, \quad \text{for } T_s \leqslant t \leqslant T_f \tag{2}$$

where the noise term N_t is independently and normally distributed with mean 0 and variance σ^2.

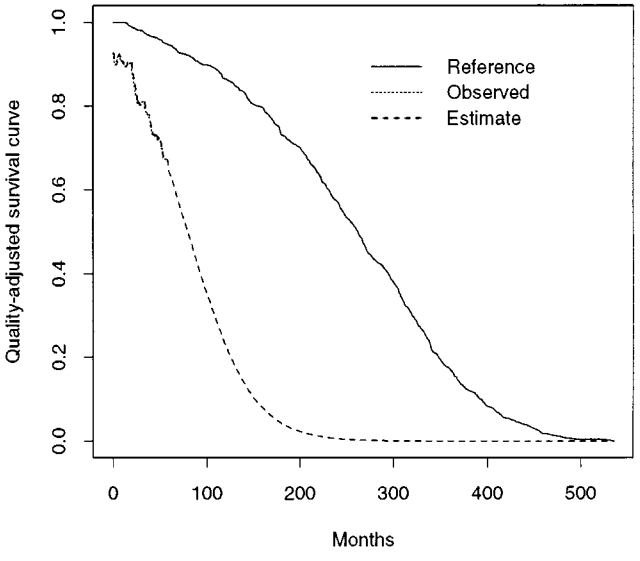

Figure 2. Quality-adjusted survival curves for index sample and reference populations, and the estimated whole curve for the index population in a simulated 60-month follow-up study

Given the least squares estimates of the two parameters, $\hat{\alpha}$ and $\hat{\beta}$, the new estimate $\widehat{\text{qasc}}(t|\text{index})$ for $t > T_f$ is given by

$$\widehat{\text{qasc}}(t|\text{index}) = \widehat{\text{qasc}}(t|\text{ref}) \frac{\exp(\hat{\alpha} + \hat{\beta}t)}{1 + \exp(\hat{\alpha} + \hat{\beta}t)}. \tag{3}$$

In order to gain an insight into estimation procedures, we calculated an example of $\widehat{\text{qasc}}(t|\text{index})$ from 60-month follow-up data and plotted the estimated quality-adjusted survival curve for index with the reference one in Figure 2.

Once we have obtained a better estimate of the whole curve of $\widehat{\text{qasc}}(t|\text{index})$, the area under the curve from 0 to t_0 is an estimate of the expected quality-adjusted survival restricted to t_0 for the specific disease population, denoted by $\widehat{\text{QAS}}_{t_0}(T_s)$. Note that the estimate of the expected $\widehat{\text{QAS}}_{t_0}(T_s)$ is affected by the choice of the necessary stabilization time T_s. If we choose a smaller T_s, then we may confront the problem of a lack of fit. If T_s is too close to T_f, fewer data are available for fitting a significant line. In substantive terms, we might make our choice of T_s according to the characteristics of the disease. For instance, it may take up to 1–3 months for a complete work-up for hypertension to rule out secondary causes of hypertension and another 1–3 months to adjust the treatment regimen. Take another example; it may take 1–3 months for a patient to recover from thyroid surgery for thyroid cancer. In practice, the choice of T_s can be similar to the procedures of Gelber *et al.*[9]. That is, use the helpful plot of logit of $W(t)$ to determine whether a simple linear regression fits the tail well and what value of T_s is appropriate. The plot of logit of $W(t)$ may provide several possible T_s values for modelling, thus we may calculate the $\widehat{\text{QAS}}_{t_0}(T_s)$ for each possible T_s and save the slope estimate, denoted as $\hat{\beta}_{T_s}$, only when the estimated intercept and slope parameters are all significant. In general, the estimated

slopes should be very close for each T_s, but it is possible to have an inaccurate estimate of slope due to outliers or influential points in the selected time interval, especially when the sample size is not large enough. In order to have a stable estimate, we suggest that the expected quality-adjusted survival time, $\widehat{\mathrm{QAS}}_{t_o}$, be given by $\widehat{\mathrm{QAS}}_{t_o}(T_s^*)$, where T_s^* is the value T_s value such that $\hat{\beta}_{T_s^*}$ is the median of all the saved $\hat{\beta}_{T_s}$.

The standard error of $\widehat{\mathrm{QAS}}_{t_o}$ can also be estimated by using resampling techniques similar to the bootstrap method.[14] The bth bootstrap data set of size N for the index population is sampled with replacement from the original data set $\Omega = \{Y_i, \delta_i, z_i, V_i, q_i\}_{i=1}^N$, denoted by Ω^b. Follow the entire Monte Carlo estimation procedure to create a new reference sample based on the bootstrap data Ω^b. The modelling and estimation procedures are then applied to the bootstrap data set and corresponding reference sample to produce a bootstrap estimate of expected QAS, denoted by $\widehat{\mathrm{QAS}}_{t_o}^b$. We may repeat the bootstrap procedure B times to collect a sample of B bootstrap estimates of expected QAS. The standard error of $\widehat{\mathrm{QAS}}_{t_o}$ is therefore given by the sample standard deviation of $\{\widehat{\mathrm{QAS}}_{t_o}^b\}_{b=1}^B$.

4. SIMULATION STUDY

4.1. Hypothetical disease populations

Three hypothetical populations of size 50,000 with specific diseases representing moderate, longer and shorter mean survival times were generated for the performance evaluation of our Monte Carlo estimation procedures. Patient's gender was generated according to a Bernoulli distribution with probability 0·5 for the three populations. In populations I and III, the onset ages were generated from gamma distributions with means of 55 and 60 years old for men and women, respectively. The standard deviation of the onset were 12 and 6 years for both men and women in these two populations. In population II, the onset ages were also generated from gamma distributions with smaller means of 37 and 42 years old for men and women, respectively. A larger standard deviation of 42 years was set for these onset ages.

Each patient's hazards after the onset age are assumed to be proportional to the hazards in the general population with the same gender and age for populations I and II. The hazard function of the general population is based on the 1993 vital statistics of Taiwan. The exact survival times T of patients in populations I and II with gender z_1 and onset age z_2 are generated from the following two hazard functions, respectively:

$$h_c(t|z_1, z_2) = h_g(t|z_1, z_2)\,(1 + A \times (1 + z_1) + \tfrac{1}{2} B \log(z_2)) \tag{4}$$

and

$$h_c(t|z_1, z_2) = h_g(t|z_1, z_2)\,(1 + A \times (1 + z_1) + B \log(z_2)) \tag{5}$$

where $h_g(t|z_1, z_2)$ is the hazard function of the general population for the age z_2, and gender $z_1 = 1$ for male and 0 for female. The random variable A is uniformly distributed in (0,1), and B is beta distributed with both of the parameters 0·5. These two populations were constructed to have the patient's hazard function worse than the general population with same age and gender. The degree of worsening conditions is partially contributed by the personal unknown random factor A and patient's onset age with a scale random factor B. Males in these two populations have a highest risk.

The survival time is much longer and more variant in population II than in population I, while the survival times in population III were constructed independently, not related to the hazard function $h_g(t|z_1, z_2)$ used in the other two populations, from a gamma distribution with a shorter mean of 8 years and a standard deviation of 4 years. The survival curves given in Figure 3 depict the survival differences among these three hypothetical index populations.

The quality of life function of the ith patient with simulated survival time T_i among these three hypothetical index populations is determined by the same function as in Hwang *et al.*,[4] that is

$$q_i(t|p, \eta, \gamma, \delta) = p(1 - t/T_i)^{\eta} + \delta(1 - p)\sin^2(\gamma t\pi/T_i) \tag{6}$$

where p, η, γ and δ are uniformly distributed in (0·8, 1), (0·01, 0·5), (0,4) and (0,1), respectively. Let $\mathrm{QAS}_i(t_0)$ be the ith patient's quality-adjusted survival since onset to time t_0, that is

$$\begin{aligned}\mathrm{QAS}_i(t_0) &= \int_0^{t_0} q_i(t|p, \eta, \gamma, \delta)\,\mathrm{d}t \\ &= \frac{pT_i}{1+\eta}\left(1 - \left(1 - \frac{t_0}{T_i}\right)^{1+\eta}\right) + \frac{\delta(1-p)t_0}{2}\left(1 - \frac{T_i}{2\gamma\pi t_0}\sin\frac{2\gamma\pi t_0}{T_i}\right).\end{aligned} \tag{7}$$

With the above equation we can easily obtain the population mean QAS up to any time t_0 by averaging these 50,000 $\mathrm{QAS}_i(t_0)$, which is denoted by $E(\mathrm{QAS}_{t_0})$. The true cumulative distributions of mean quality-adjusted survival times, generated from (7), for these three hypothetical index populations are plotted in Figure 4. The mean QASs restricted to 60 months are very close, which are 48·6, 52·4 and 47·2 months for these three populations, respectively. As time expands to whole life, these three mean QASs grow with different speeds to 131·9, 242·4 and 74·4 months. In the following steps of the simulation studies, we use k-month data sampled from each of these three populations to project results beyond k months which will be compared with the above true mean QAS for evaluating the degree of accuracy of the Monte Carlo approach.

4.2. Sample from the index populations

A hypothetical k-months follow-up study was designed to have patients entering the sample every year to accumulate a final sample of size N by the end of the follow-up. In the simulation, we selected a random sample of N patients uniformly from the above hypothetical index population. For the ith patient we recorded his/her gender and onset age and compared survival time Y_i to the follow-up time U_i, which is k months times a random number generated from beta(9·5, 0·5). Note that we have used a random $U_i \leqslant k$ to make the sample heavily censored. If Y_i is smaller than U_i months, we treat Y_i as a complete survival time and assign the censor status variable $\delta_i = 1$. Otherwise we assign $\delta_i = 0$, indicating the case is still alive with right censoring time U_i. Meanwhile we also generate another uniform random number V_i from 0 to U_i to represent the time of quality of life interview. To allow sampling error, the patient's quality of life $q_i(V_i)$ at this time point is recorded as the patient's simulated quality of life values at a time point drawn uniformly between three months before and after V_i.

4.3. Simulation results

The accuracy and precision of projected estimates of quality-adjusted survival in a follow-up study are determined mostly by the underlying true survival curve, length of follow-up, time to

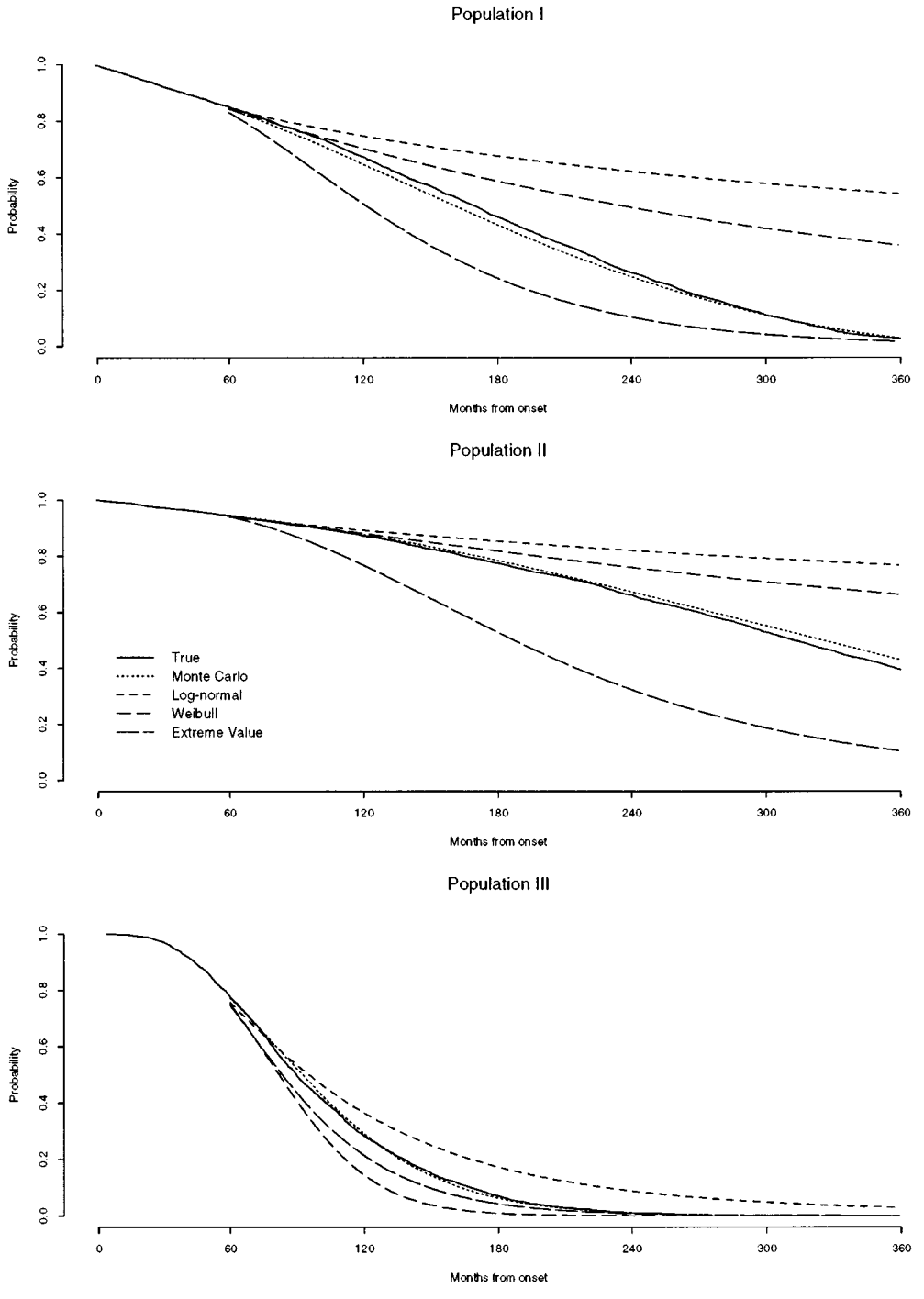

Figure 3. The true and projected survival functions based on Monte Carlo and parametric model approaches for the three hypothetical index populations

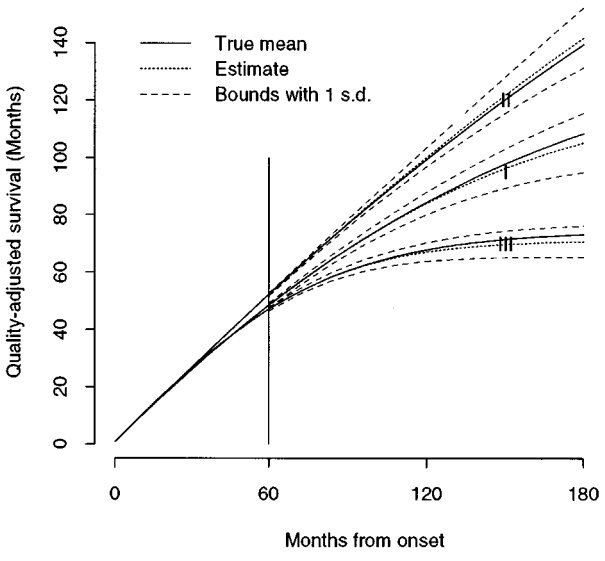

Figure 4. The true and projected quality-adjusted survival estimates with bounds of one standard deviation for the three hypothetical index populations

extrapolation and sample size. For demonstration, we used 60-month follow-up data with sample sizes $N = 100$ and 300 from these three hypothetical data sets to estimate and project the expected quality-adjusted survival time restricted to several limits beyond the follow-up period. The short follow-up data end up with censoring rates of 0·85, 0·95 and 0·79 for the three index populations.

For each setup we repeated the sampling and estimation procedures 300 times to obtain 300 projected estimates of quality adjusted survival time and 300 bootstrap standard errors. The average of these 300 QAS estimates and standard errors are denoted as $\widehat{\mathrm{QAS}}_{t_p}$ and $\mathrm{SE}(\widehat{\mathrm{QAS}}_{t_p})$ for projection time t_p, respectively. We then assessed the accuracy of the Monte Carlo estimator by the relative deviation of $\widehat{\mathrm{QAS}}_{t_p}$ from the true mean QAS of the index population, $E(\mathrm{QAS}_{t_p})$. The simulation results in Tables I, II and III and Figure 4 show that projected short-term results are quite good in terms of relative biases. The relative biases, as expected, tend to increase as extrapolations extend. However, small relative biases of 5–7 per cent are still found for the estimates even restricted to a long projection of 300 months.

In population I with sample size 100, projected QAS tends for a longer time to have a large underestimate. To further study the bias of the projected estimates, we may extend the follow-up period to a longer time. Suppose that we have a longer follow-up of 240 months and have computed the logit of W such as depicted in Figure 5. We see the slope of logit of W in population I has a slight upward change around the 140th month. This is why the projected estimates have produced underestimated results for time beyond that point in population I. Therefore, seeking a more appropriate reference population and continuing follow-up are necessary when a high accuracy is needed for inference.

The true standard error of the estimator is estimated by the mean squares of these 300 QAS estimate deviations from the true $E(\mathrm{QAS}_{t_p})$. Comparing the average of the 300 bootstrap standard

Table I. Simulation results for population I based on 60-month follow-up sample with censoring rate 0·85

Sample size (N)	Projected months (t_p)	True $E(\mathrm{QAS}_{t_p})$	Estimated $E(\mathrm{QAS}_{t_p})$	Relative bias	True SE	Estimated SE
100	0	48·64	48·13	− 0·011	1·46	1·99
100	60	84·47	82·86	− 0·019	6·63	10·12
100	120	108·45	103·90	− 0·042	15·81	20·94
100	180	122·50	115·16	− 0·060	24·33	30·53
100	240	129·19	120·44	− 0·068	30·06	37·36
100	300	131·45	122·30	− 0·070	32·65	40·79
300	0	48·64	48·45	− 0·004	0·85	0·82
300	60	84·47	84·50	− 0·000	3·64	3·79
300	120	108·45	106·92	− 0·014	9·02	9·06
300	180	122·50	118·63	− 0·032	14·64	14·01
300	240	129·19	123·69	− 0·043	18·50	17·28
300	300	131·45	125·26	− 0·047	20·17	18·74

Table II. Simulation results for population II based on 60-month follow-up sample with censoring rate 0·95

Sample size (N)	Projected months (t_p)	True $E(\mathrm{QAS}_{t_p})$	Estimated $E(\mathrm{QAS}_{t_p})$	Relative bias	True SE	Estimated SE
100	0	52·42	52·09	− 0·006	1·17	2·93
100	60	99·42	98·53	− 0·009	6·64	14·63
100	120	139·45	136·91	− 0·018	18·67	31·15
100	180	172·37	167·61	− 0·028	34·46	49·15
100	240	198·08	191·89	− 0·031	51·37	67·04
100	300	216·84	210·82	− 0·028	67·52	83·72
300	0	52·42	52·26	− 0·003	0·56	0·83
300	60	99·42	100·08	0·007	3·10	6·19
300	120	139·45	141·61	0·016	9·52	16·84
300	180	172·37	176·06	0·021	18·83	30·35
300	240	198·08	203·34	0·027	29·75	44·75
300	300	216·84	223·98	0·033	40·98	58·68

errors, $\mathrm{SE}(\widehat{\mathrm{QAS}}_{t_p})$, with the estimated true standard error, we see that the bootstrap approach of standard error estimation works well in populations I and III. For the 95 per cent censoring rate example of population II, the bootstrap standard error estimation tends to be too conservative. We also found that the main contribution of a larger sample size is in reducing standard error and of limited help on bias.

Although the methodology was developed for expected quality-adjusted survival estimation, it can be applied directly to survival function estimation simply by assuming that the quality of life function is a constant of one in the index population. Parametric model approaches for extrapolating survival function in follow-up studies are popular and also available in statistical

Table III. Simulation results for population III based on 60-month follow-up sample with censoring rate 0·79

Sample size	Projected months (t_p)	True (N) $E(\mathrm{QAS}_{t_p})$	Estimated $E(\mathrm{QAS}_{t_p})$	Relative bias	True SE	Estimated SE
100	0	47·19	47·11	− 0·002	1·02	1·21
100	60	67·91	68·50	− 0·009	5·48	5·76
100	120	73·23	74·19	− 0·013	9·38	9·55
100	180	74·23	75·53	− 0·017	11·10	11·42
300	0	47·19	47·19	0·000	0·63	0·65
300	60	67·91	69·10	0·018	3·40	3·24
300	120	73·23	74·77	0·021	5·60	5·37
300	180	74·23	75·88	0·022	6·37	6·19

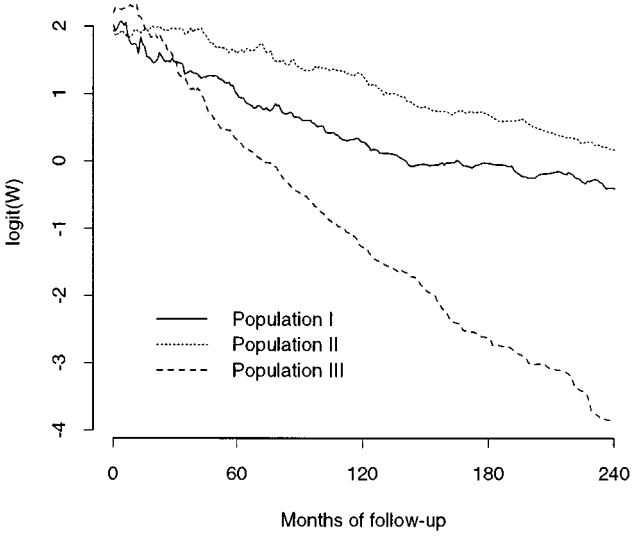

Figure 5. Plots of logit of $W(t)$ calculated from samples (size = 100) of the three hypothetical populations

software such as SAS and IMSL. For demonstration, we compared the Monte Carlo approach with three models of log-normal, Weibull and extreme value in the FORTRAN library IMSL.[15] Survival data in the above three hypothetical samples of size 300 were used for these approaches to project survival functions. In the simulation study, 100 repetitions were implemented and the averages of the 100 estimated survival functions obtained from each of the two approaches are plotted with the true ones in Figure 3. It is very clear that the Monte Carlo approach has a quite good performance of long-term projection in these three hypothetical populations. It also does not seem easy to identify proper parametric models to have such accurate projections, although we have not tried all available models.

5. DISCUSSION

Survival data with heavy censoring are often encountered in follow-up studies of cohorts with long-term survival. Without other auxiliary information, there is probably little chance for complex techniques to produce any convincing result beyond the end of follow-up from heavily censored data. In this paper we proposed to borrow information from easily accessible data, such as vital statistics, to match the index population, apply simple procedures including a logit transform of the ratio of quality-adjusted survivals for reference and index populations, and use the linearity of the transformed curve to make inferences. We successfully applied the methodology to data from three different types of hypothetical disease populations. In simulation, we predicted from 12-month to life-long results using a sample of 60-month follow-up with heavy censoring and compared these predictions to the true ones. Given a sample of 100 or 300, the relative biases of the projected estimates are within 5–7 per cent even for a long projection of 300 months. The true expected QAS beyond any time point of follow-up is also within one standard error of the projected estimate.

The methodology can be directly applied to the whole survival function estimation in follow-up studies. Simulation studies showed quite convincing results on three hypothetical examples, compared with approaches by popular parametric models. In fact, the usual parametric models may be suitable only for short-term projection as shown on Figure 3 or extrapolations based on better fit of tail part of observed survival proposed by Gelber *et al*. Figure 3 also indicates that these approaches may not produce accurate enough estimates for a longer time projection.

The performance of the proposed methodology is mainly determined by an available proper reference database, but for most follow-up studies of chronic disease, using the life table of general population as the reference is probably good enough. Besides, if there is a more valid reference population available, it can still be used directly in our method, conceptually similar to the approach proposed by Mark *et al.*[10] Therefore, our simple methodology is more feasible and may be a universal solution for practical applications.

The accuracy and precision of the Monte Carlo estimator for the expected QAS are also affected by the length of follow-up, the behaviour of the two tails underlying the true curve of logit W, and, of course, sample size. Among these factors, we note that the non-linear left tail, reflecting the patient's unstable quality of life in the beginning stage, is usually not uniform and must be adjusted for different diseases in terms of magnitude and extent. For diseases which require both surgery and chemotherapy, say, breast cancer, this unstable stage may last up to 1 or 1·5 years, depending on the time required for the comprehensive diagnosis and treatment. For a chronic disease which only needs initial diagnostic work-up and selection of suitable therapeutic regimen, such as diabetes mellitus, this early unstable period may only take 2–3 months. The performance of the Monte Carlo estimator applied to a study with too short a follow-up period and too small a sample size is therefore not guaranteed for a long-term projection. In order to judge whether the Monte Carlo estimator of expected QAS is appropriate for a follow-up study, we suggest a diagnostic tool, which checks the linearity of the logit of W, such as that depicted in Figure 5. If the linearity is significant, our approach can be applied to produce an accurate estimate of the expected QAS beyond the follow-up. On the other hand, the life-long projected QAS might be slightly underestimated when the true curvature of the right tail of logit of W is large. This may happen when an inappropriate reference population is used and cannot be fully evaluated based only on the observed data. Although a larger sample size will increase the precision, the gain in reducing bias may be limited. Therefore, it may be necessary to continue the follow-up to ensure

the accuracy. However, the estimator and its confidence limits still provide a ballpark idea of prognosis and may be useful for policy and resource planning.

ACKNOWLEDGEMENTS

The authors are particularly grateful to the referees and the editor who provided helpful comments and suggestions on this paper. This study was supported in part by the National Science Council of ROC, Taiwan.

REFERENCES

1. Gold, M. R., Siegel, J. E., Russel, L. B. and Weinstein, M. C. *Cost-effectiveness in Health and Medicine*, Oxford University Press, New York, 1996.
2. Patrick, D. L. and Erickson, P. *Health Status and Health Policy*: *Allocating Resources to Health Care*, Oxford University Press, New York, 1993.
3. Testa, M. A. and Simonson, D. 'Assessment of quality-of-life outcomes', *New England Journal of Medicine*, **334**, 835–840 (1996).
4. Hwang, J. S., Tsauo, J. Y. and Wang, J. D. 'Estimation of expected quality-adjusted survival by cross-sectional survey', *Statistics in Medicine*, **15**, 93–102 (1996).
5. Glasziou, P. P., Simes, R. J. and Gelber, R. D. 'Quality adjusted survival analysis', *Statistics in Medicine*, **9**, 1259–1276 (1990).
6. Shaper, A. G., Wannamethee, S. G. and Walker, M. 'Body weight: implication for the prevention of coronary heart disease, stroke, and diabetes mellitus in a cohort study of middle aged men', *British Medical Journal*, **314**, 1311–1317 (1997).
7. Sytkowski, P. A., D'Agostino R. B., Belanger, A. J. and Kannel, W. B. 'Secular trends in long-term sustained hypertension, long-term treatment, and cardiovascular mortality: The Framingham heart study 1950 to 1990', *Circulation*, **93**, 697–703 (1996).
8. Mazzaferri, E. L. and Jhiang, S. M. 'Long-term impact of initial surgical and medical therapy on papillary and follicular thyroid cancer', *American Journal of Medicine*, **97**, 418–428 (1994).
9. Gelber, R. D., Goldhirsch, A. and Cole, B. F. 'Parametric extrapolation of survival estimates with applications to quality of life evaluation of treatments', *Controlled Clinical Trials*, **14**, 485–499 (1993).
10. Mark, D. B., Halatky, M. A., Califf, R. M. *et al.* 'Cost effectiveness of thrombolytic therapy with tissue plasminogen activator as compared with streptokinase for acute myocardial infarction', *New England Journal of Medicine*, **332**, 1418–1424 (1995).
11. Wang, J. D. and Miettinen, O. S. 'Occupational mortality studies: principles of validity', *Scandinavian Journal of Work and Environmental Health*, **8**, 153–158 (1982).
12. Hakama, M. and Hakulinen, T. 'Estimating the expectation of life in cancer survival studies with incomplete follow-up information', *Journal of Chronic Diseases*, **30**, 585–597 (1977).
13. Lee, E. T. *Statistical Methods for Survival Data Analysis*, 2nd edn, Wiley, New York, 1992.
14. Efron, B. and Tibshirani, R. J. *An Introduction to the Bootstrap*, Chapman and Hall, New York, 1993.
15. *IMSL Stat/Library*, Visual Numerics, Inc., 1997.

