

Title: Quality-Adjusted Survival as Endpoint in Breast Cancer Trials

Keywords: Quality-Adjusted Survival Analysis, QALY, Quality of Life, Breast Cancer

Word length: 5859

Abstract: Breast cancer treatment recommendations will often require an appraisal of likely benefits in relation to likely side-effects on survival and quality of life (QoL) endpoints, and possibly also an evaluation of the size of the anticipated net clinical benefit against financial costs. Quality-adjusted survival (QAS) analysis methods provide a formal approach for deriving an estimate of net clinical benefit to facilitate this appraisal process. QAS analysis methods have been applied in trials with breast cancer patients of adjuvant therapies as well as treatments for advanced/metastatic disease. QAS analyses based solely on trial data may fail to capture plausible longer-term benefits; thus methods to explore the possible outcomes of treatment beyond the limits of trial data have been developed. These modelling approaches can help researchers gain insights and identify future research priorities, but do not replace the need for long-term evidence from randomised trials.

Text Box/Executive Summary: See text after bibliography

1. Introduction

The relevance of quality of life (QoL) outcomes in breast cancer care is widely recognised, with both disease symptoms and treatment side-effects having a substantial impact on patient well being[1]. Clinical trials of breast cancer therapies therefore often include QoL as an endpoint; however, integrating QoL data with information on other important endpoints such as survival time or even economic costs to determine the overall net-benefit and cost-effectiveness of a treatment poses a challenge.

Ideally a treatment should enhance quality of life, prolong survival time and reduce health care costs, but trade-offs between these desired outcomes commonly arise. In some cases a treatment may provide quality of life benefits but raise questions about compromising longer-term survival. For example, sentinel node biopsy provides a method for staging early breast cancer that reduces the risk of arm morbidity, compared to axillary lymph node dissection, but it may fail to detect cancer spread in a small proportion of cases leading to suboptimal treatment, raising the possibility of an increase in the likelihood of cancer recurrence, and poorer longer-term outcomes.[2]

In other cases, a treatment may provide gains in (progression-free) survival at the expense of worse QoL in the short-term due to side-effects. For example, adjuvant chemotherapy may produce modest survival gains[3] whilst inducing a number of unpleasant side-effects that will influence patients perception of the value of treatment[4-5]. Furthermore, dose-intensive adjuvant chemotherapy may provide an additional modest gain in disease-free survival, compared to standard dose chemotherapy, but with a deleterious impact on QoL during the treatment period[6]. In metastatic breast cancer, combination therapy may be superior to monotherapy in terms of progression-free survival, but worse with regards to side-effects.[7-8]

In each of these cases, the impact of trade-offs between quality of life and survival needs to be evaluated in order to assess the net value of each treatment. Such an evaluation should aim to

incorporate the subjective importance patients place on quality of life outcomes relative to survival outcomes. Most QoL instruments do not provide an assessment on this scale; a *utility-based* QoL assessment, as described in subsequent sections of this article, is required.

There may be a net clinical benefit for one treatment over the other for the individual patient after consideration of the above trade-offs, however difficult decisions may yet still need to be made as to whether the net health gains are sufficient to justify a treatment's financial cost in order to make best use of limited health care resources. If the incremental health gain is small but very expensive to achieve, then alternative uses of the funds may provide greater health benefits to the community.

The financial cost associated with oncology care has increased considerably with the introduction of new diagnostic technologies, drugs, and radiotherapy treatments[9]. For example, modern aromatase inhibitors appear to have greater efficacy than established adjuvant endocrine therapy with tamoxifen for oestrogen receptor positive breast cancers[10], but are many times more costly[11]. Adjuvant therapy with trastuzumab is effective for HER2 positive cancers, but is expensive with a 12 month treatment course costing around AU\$50,000. A new combination agent, trastuzumab emtansine, has the potential to be more effective than trastuzumab alone[12] but is likely to be even more costly.

The additional financial cost of the additional health gain associated with a new treatment may be expressed in terms of a single index known as an incremental cost-effectiveness ratio (ICER – i.e. the ratio of the difference in cost, between the new treatment and the comparator treatment, relative to the difference in their effectiveness)[13]. An ICER will not in itself provide an answer as to whether or not a new treatment affords good value, however collectively ICERs may be used to rank health initiatives and treatments to assist with funding decisions. Preferentially allocating funds to better value treatments (with lower ICERS) over those providing less value (with higher ICERS) provides a method for distributing finite resources efficiently. Yet, comparisons between ICERS for

different treatments are only possible when effectiveness is expressed on a common scale that ideally captures the overall net impact of treatment.

A quality-adjusted survival index that summarises the impact of a treatment on both QoL and survival simultaneously is therefore not only useful for evaluating trade-offs, but also as a common measure of net effectiveness in economic evaluations. This article provides an introduction to methods for deriving such a quality-adjusted survival time index, including an overview of approaches for obtaining QoL data on the required *utility* scale, as well as analysis techniques for evaluating treatment alternatives in terms of quality-adjusted survival. This material is presented with a focus on breast cancer care but is broadly applicable to many other therapeutic areas. Examples of published studies in breast cancer that have undertaken quality-adjusted survival are presented, as are general recommendations for the applicability and future implementation of such analyses.

2. Quality-Adjusted Survival and Expected Utility Theory

An evaluation of the net-benefit of a treatment will require consideration of survival outcomes, QoL outcomes, and some approach for weighting these components to form an overall assessment. Methods for measuring the effect of disease and its treatment on a single index capturing both QoL and survival outcomes first arose in the early 1970s[14-15]. It was proposed that the desirability of a health outcome could be represented by adjusting the observed survival period by a weight that reflected the relative quality of the health state experienced relative to full health, which is given a weighing of 1, and death, which is given a weighting of 0. According to the theory, a 12-month survival period spent in an intermediate health state with a quality weighting of 0.5 would be valued the same as a 6-month survival period spend in full health. Various names have been applied to the concept of adjusting survival time by a weight reflecting the desirability of the health state experienced but the expression 'quality-adjusted life years' (QALYs) has tended to dominate. An

important step in the evolution of the QALY was the establishment of a theoretical link with a formal method of decision making based on expected utility theory[16].

In the context of expected utility theory, a utility is essentially a measure of preference for a given outcome that is assessed on a probability scale. The criterion measurement approach for obtaining utilities is known as the *standard gamble* and involves having an individual compare a given outcome against a gamble (or lottery) between the best and the worst possible outcomes. For example an individual's strength-of-preference for a given intermediate health state (e.g. moderate symptoms of pain, fatigue, nausea, vomiting) can be evaluated against a hypothetical gamble that offers a p probability of full health and a $1-p$ probability of immediate death. The gamble between full health and death can be made to look more or less attractive by varying the probability p , and through an iterative process, a value for p can be found that makes the gamble appear equally preferable to the intermediate health state. At this point of indifference the gamble, with an expected pay-off of $(p \times 1) + (1-p \times 0) = p$, is said to be the *probability equivalent* of the intermediate health state – the utility of the intermediate health state is thus estimated as p . A utility function for survival time may likewise be modelled using information from standard gamble tasks that estimate the relative utility of a particular survival duration against a gamble between worst (e.g. immediate death) and best case (e.g. 20 years) scenarios.

Utility functions for multi-attribute outcomes, such as health outcomes expressed on a QALY scale, may also be constructed. The QALYs accrued by an individual may be estimated by multiplying (a function of) the time spent in various discrete health states by the utility of those states and summing the products. If we assume that quality of life varies with time in a continuous rather than discrete fashion, then the rate of gain in QALYs with respect to time may be represented by the equation below:

$$dQALY/dt = u_i(t) f'(t)$$

Equation 1: QALY model with changes in health state and utility function for time

where u_i represents the utility of health state i at time t and $f(t)$ is the utility function for time. Figure 1 illustrates this for hypothetical breast cancer patient as the area under the curve defined by the QoL experienced over time.

In addition to the difficulties in ensuring the validity of the assumptions underlying the QALY model[17], there are various methodological challenges in using quality-adjusted survival as a clinical trial endpoint. One of these is how best to obtain utilities for health states applicable to patients participating in a clinical trial. Another is how to perform an analysis of quality-adjusted survival to compare treatments. We explore each of these two tasks in the subsequent sections.

3. Utility-Based QoL Assessment

Utility assessment represents a distinct approach to the evaluation of QoL that is different to typical QoL instruments that express scores across separate aspects of QoL on arbitrary scales (e.g. the EORTC QLQ-C30 instrument [18]). We differentiate utility-based QoL scores hereafter by referring to them as *uQoL* scores. The criterion method of measuring *uQoL* scores is the standard gamble approach described above[19]. The time trade-off method has been used as a conceptually simpler alternative to the standard gamble and involves a hypothetical comparison between a fixed survival period spent in a given health state versus a shorter period of survival time spent in full health[14]. Simpler again are numerical category scaling or visual analogue scale tasks, however whilst these provide information on the magnitude of QoL impairments, they generally do not yield scores with true utility properties. Standard gamble, time trade-off, and category scaling methods typically produce systematically different estimates and approaches to correct for the measurement bias inherent with category scaling and time trade-off tasks have been demonstrated[20-21]

Self-administered questionnaires provide a very practical method for collecting information across the multiple aspects of quality of life in clinical trials however, as with the numerical category scaling or visual analogue scale tasks mentioned above, they generally provide information on QoL

impairments on an arbitrary scale that do not have utility properties. Only a subset of quality of life questionnaires have scoring systems designed to estimate uQoL. Underpinning such instruments is some form of classification system, that categorises an individual's health state according to his/her responses to the questionnaire items, and a method for assigning a uQoL score to the health state defined by the classification system. A variety of statistical approaches have been used to estimate uQoL scores from questionnaires with applicability to breast cancer patients.[8, 22-24]

An intriguing question is whose perspective and preferences should be reflected by uQoL assessments of health states. Quality of life is clearly a subjective concept and patients' self-ratings have indeed been shown to be more reliable than those made by clinicians on behalf of patients[25]. There may nevertheless be instances when adopting the perspective of individuals other than patients could be appropriate. In the context of a cost-effectiveness analysis for example, one might argue that the value of alternative treatment options should be appraised from the societal perspective because it is society's resources that are to be allocated across treatment alternatives[26-27]. In the context of reaching treatment decisions for individual patients, some patients themselves may feel that an experienced physician is in a better position to appraise the health outcomes associated with various treatments than they themselves, or there may be pragmatic grounds for employing uQoL estimates from breast cancer experts[28]. Differences have been demonstrated between patients, health professionals and healthy volunteers in terms of health state utility estimates as well as treatment priorities and preferences[29-31]. Such differences have in turn been shown to have the potential to appreciably alter estimates of the net-effectiveness and cost-effectiveness of treatments in various therapeutic areas. [32-33]

Combining the uQoL and survival data obtained within a clinical trial to undertake an analysis of QALY data appears conceptually straightforward but actually presents a number of technical challenges. A comparison between the treatment arms in a clinical trial based on a simple summation of the QALYs accrued by individual patients will be inappropriate if patients are followed-

up for different periods of time and there is censoring. Substituting QALYs for survival time and applying a log-rank test or Cox regression model to QALYs is also inappropriate because the assumption of non-informative censoring is likely to be violated.[34]

4. Quality-Adjusted Time Without Symptoms or Toxicity (Q-TWiST)

One way of calculating quality-adjusted survival that accommodates censoring and differentials in observation time appropriately is the quality-adjusted time without symptoms and toxicity (Q-TWiST) method[34]. The Q-TWiST method partitions the standard Kaplan-Meier survival curve for a treatment group into time spent in discrete health states – this is the defining characteristic of *health state-based* methods for estimating quality-adjusted survival. Figure 2 presents an example of this approach using data from a hypothetical study of adjuvant chemotherapy in breast cancer.

Commonly, three states are distinguished: one for time spent experiencing the short-term side-effects of treatment (toxicity—TOX), one for time spent without symptoms or toxicity (TWiST), and one for time spend after relapse (REL) or disease progression (PD). The Kaplan-Meier method is used to estimate time to event curves for the states, and each state is assigned a uQoL score. The sum of the uQoL-weighted area under each curve provides an estimate of quality-adjusted survival.

Depending on the maturity of the data, and the censoring patterns, a restricted estimate of quality-adjusted survival may be appropriate where the area under each curve is calculated up to a defined truncation point (e.g. median follow-up time).

Resampling methods (e.g. bootstrap) are typically used to obtain a measure of variance for these estimates allowing confidence intervals and p-values to be calculated. Commonly, the uQoL weights are assumed to be estimated without measurement error and the results of Q-TWiST analyses are presented as a two-way threshold analysis where the conclusions associated with the full range of possible utility values TOX and REL are presented graphically. This affords the opportunity to judge the net clinical benefit of a therapy from the perspective of individual patients with different

attitudes towards time spend in TOX and REL. Semi-parametric and fully parametric generalisations of the Q-TWiST model have been developed that allow for the addition of covariates to explore how quality-adjusted survival time may differ relative to various prognostic factors[35-36]

Q-TWiST analyses have been undertaken in breast cancer with data from trials of adjuvant therapies as well as treatments for advanced/metastatic disease. Examples of these are presented in the next sections.

5. Q-TWiST Evaluations of Adjuvant Chemotherapy

In early breast cancer, the trade-off between short-term treatment toxicity versus improved survival time associated with adjuvant chemotherapy was evaluated by applying the Q-TWiST approach to data from 47 trials involving over 17,000 women[37]. Actual treatment duration data was not consistently available across the included studies, thus a fixed duration of 6 months of TOX was applied to each trial as this was the standard duration of adjuvant therapy at the time. The average time spent in TWiST and REL was estimated as the area under standard Kaplan-Meier curves restricted to each trial's respective median follow-up duration. The estimates of (restricted) average time spend in TWiST and REL from each individual trial were then pooled in a meta-regression and used to estimate treatment effects over a 10-year period. Within-trial QoL data were unavailable to derive uQoL weights and so Q-TWiST results were presented as a two-way threshold analysis in which uQoL weights for TOX and REL were varied between 0 and 1. For women aged under 50, adjuvant chemotherapy produced greater quality-adjusted survival over a relatively wide range of uQoL weights. This equated to an additional 4.8 months of quality-adjusted survival when the uQoL of TOX and REL were assumed to be 0.5. The results for older women followed a comparable pattern but the benefits were somewhat smaller in magnitude.

The net benefit of adjuvant chemotherapy in older women (with node negative disease) was further investigated by performing a Q-TWiST analysis on the IBCSG-9 trial[38]. The actual duration of

adjuvant chemotherapy treatment was used to define the time spent in TOX, and uQoL weights were approximated by applying a bias correction to QoL VAS scores collected longitudinally from trial participants. The median QoL scores at 3 months, those between treatment completion and 24 months, and those during the first 6 months after relapse, were used to derive uQoL weights for TOX (0.89), TWiST (0.91), and REL (0.71). The estimated quality-adjusted survival, restricted to the median follow-up of 71 months, was superior by a very modest degree for adjuvant therapy (62.5 versus 61.4; $p=0.03$) and was supported by the findings of a two-way threshold analysis varying utility weights for TOX and REL.

Dose-intensive adjuvant chemotherapy regimens have been speculated to provide further gains over standard regimens in terms of disease free survival but cause more intense short-term toxicities. A Q-TWiST appraisal of this trade-off was undertaken using data from the IBCSG-15 trial comparing a dose-intensive regimen of epirubicin and cyclophosphamide to standard-dose anthracycline-based chemotherapy.[6] In that analysis, each episode of grade 3-4 toxicity was counted as contributing between 1 to 3 months (depending on type and timing of AE) towards each patient's time spent in TOX. uQoL weights for TOX and REL were estimated using data obtained from patients completing longitudinal QoL self-evaluations. The median QoL value obtained during each state (TOX =0.77, TWiST=0.91, and REL=0.77) was used as the base-case uQoL weights. A non-significant trend towards improved disease free survival for dose-intensive chemotherapy but worse QoL during the treatment period was observed in IBCSG-15 trial[6]. In terms of quality-adjusted survival, this translated into a small, and non-statistically significant, benefit for dose-intensive therapy. This did not reach statistical significance for any values tested in the two-way threshold analysis on the utility estimates.

6. Q-TWiST Evaluations in Advanced Breast Cancer

Trials of chemotherapy for advanced and metastatic breast cancer have also been analysed using the Q-TWiST approach as well as variants in which quality-adjusted time is assessed only up to the point

of disease progression[39]. Corey-Lisel et al. assessed the trade-off between improved progression-free survival but increased treatment toxicity observed for combination therapy with ixabepilone plus capecitabine compared to capecitabine alone in metastatic breast cancer. [7] The TOX state was defined as the total time spent with grade 3-4 toxicities before disease progression. Compared to patients in the monotherapy group, those randomised to combination therapy experienced a longer duration in TOX, a non-statistically significant longer duration in TWiST, and a comparable duration following disease progression (PD). No estimates of the QoL experienced by patients participating in the trial were used in the Q-TWiST analysis and results were presented as a two-way threshold analysis. Under the base assumption that TOX and PD both had a uQoL weight of 0.5, the quality-adjusted survival with combination therapy was statistically significantly higher than that for monotherapy (42 versus 38 weeks, $p=0.02$). The threshold analysis indicated that combination therapy was preferred across a range of plausible utility values for TOX and PD.

Various other studies in advanced and metastatic disease have used quality-adjusted survival time as an endpoint to simply provide a concise summary of the treatment effect despite there being no clear trade-off between QoL and survival to evaluate. For example, a Q-TWiST analysis was undertaken using data from a trial that concluded combination therapy using lapatinib with capecitabine improved disease free survival with no clear increase in toxicity, compared to capecitabine alone, in women with advanced or metastatic HER2 breast cancer who had previously progressed on established therapy[40]. For that analysis, each patient's time spent in TOX was estimated by summing the total time they spent experiencing a grade 3-4 toxicity. Although a utility-based QoL instrument was used in the trial, estimates of uQoL weights for TOX and REL were set to 0.5 in a base-case analysis and varied between 0 and 1 in a threshold analysis. A statistically significant benefit for combination therapy was found under all instances where the uQoL for REL was less than that for TOX – a result illustrating that threshold analyses may give a clinically useful result that does not require detailed uQoL assessment of multiple health states.

In advanced metastatic cancer, the question of the optimal duration of chemotherapy (with CMF) for palliation has been investigated with a Q-TWiST approach using data from an EORTC trial that compared the addition of continuous chemotherapy (until progression) following an initial 3-month treatment period versus no further chemotherapy[41]. Averaged QoL values from VAS measures performed by patients were used to approximate uQoL weights for TOX (0.54), TWiST (0.73) and REL (0.29). The Q-TWiST analyses (restricted to the median follow-up time) found a non-statistically significant difference favouring no treatment despite the fact that progression free survival was significantly longer in the continuous chemotherapy arm (5.2 versus 3.5 months; $p=0.01$). A two-way threshold analysis found no combination of the utilities evaluated made continuous chemotherapy statistically significantly better than stopping after the initial 3 months of treatment, however the range of utilities included in that analysis was constrained to values less than 0.73 and these results are not consistent with a related older study showing improvements in QoL up to the time of disease progression for continuous versus intermittent CMFP chemotherapy. [42]

7. Other Health State-Based Approaches for Estimating Quality Adjusted Survival Time

The Q-TWiST method is well suited to situations where patients pass through clinically distinct states with different uQoL weights, but may not otherwise reflect the treatment experience well. Cole et al.[43] proposed a variation on the conventional Q-TWiST model that distinguished the time prior to disease progression on the basis of whether the QoL experienced was 'good' or 'poor' rather than a clinical finding (e.g. discovery of tumour on CT scan). An arbitrary cut-point that delineates 'good' QoL from 'poor' QoL is applied to measurements obtained from patients repeatedly over time. Patients may transition between these states multiple occasions, and the Kaplan-Meier method is used to estimate the time spent in each state for each transition. These estimates are then summed over the total number of transitions applicable to yield a total duration and the analysis proceeds in

much the same way as a conventional Q-TWiST analysis with uQoL weights being applied to the 'good' and 'poor' QoL states.

8. Quality-Adjusted Survival Analysis with Repeated Measures

An alternative to Q-TWiST that affords even greater flexibility to accommodate changes in uQoL over time is the quality-adjusted survival analysis with repeated measures (QASA) method[17]. This requires an estimate for each treatment group of the survival function, and the function describing uQoL over time. The product of these functions integrated up to time t provides an estimate of quality-adjusted survival accrued up to time t (see Equation 1). The standard Kaplan-Meier method may be used to estimate a survival function and a variety of options are available to estimate a function for uQoL. One would be a step function that interpolates between mean uQoL estimates obtained at the fixed time points. In this case the QASA method would be analogous to partitioning the survival curve for a given treatment arm into discrete time periods during which a uQoL assessment was performed (Figure 3), as opposed to partitioning the survival curve into health states as in the Q-TWiST method. An alternative method would be to interpolate between the fixed uQoL assessment time points to generate a function for each patient, and then average the individual functions to estimate a group uQoL function for each treatment arm. Mixed linear models for repeated measures data provide another approach for estimating a uQoL function overtime time that yields unbiased estimates under less restrictive assumptions than those required when a simple summary of observed uQoL data is used[44].

9. Extrapolation of Outcomes Beyond Trial Data

Ideally a quality-adjusted survival analysis would capture the lifetime net effects of a treatment, however analyses based on trial data alone are likely to have a limited time horizon due to the finite follow-up duration of the study. This is a particularly relevant issue for trials in early breast cancer where long-term survival rates are relatively high. For instance, the Q-TWiST meta-analysis of

adjuvant chemotherapy in early breast cancer estimated treatment effects to 10 years, at which point the overall survival probability was still over 50%. Thus subsequent survival benefits beyond this point are effectively ignored yet the full impact of the short-term side-effects were counted in the Q-TWiST analysis.

Methods for exploring the net benefits of treatment beyond the limits of trial follow-up time have been developed and are presented in the next sections. These methods have included composite approaches that model the tails of Kaplan-Meier curves beyond a chosen truncation point (e.g. median follow-up time), and approaches that model the entire experience of patients using, for example, a Markov-type process.

9.1. Extrapolation of the Kaplan-Meier Curves

Gelber et al.[45] applied a composite approach to evaluate the net-benefit of long-duration versus short-duration adjuvant chemotherapy beyond the initial published Q-TWiST analysis that was restricted to 5 years. Parametric log-normal models were fitted to project out the TWiST and REL curves from 5 out to 10 years follow-up. The estimates of quality-adjusted survival from the standard Q-TWiST (to 5 years follow-up) and projected curves (from 5 to 10 years follow-up) were combined and re-sampling methods used to undertake statistical inferences. Extrapolating beyond a truncated Q-TWiST analysis even further into the future will generally require a parametric model that accounts for the increasing hazard of death with age such as the Gompertz function employed by Trippoli et al[46] to estimate the lifetime benefits of adjuvant chemotherapy in node-positive breast cancer.

9.2. Extrapolation Using Decision Analytic Modelling

Decision analytic modelling incorporating Markov processes has been a popular approach for extrapolating beyond the randomised trial evidence to estimate quality-adjusted survival (and undertaking cost-effectiveness analyses). The Markov processes used in this setting generally model

the experience of patients as a progression through a series of discrete health states over a time horizon divided into periods, or cycles, of equal duration. Transition probabilities determine the chance of moving from one health state to another at the conclusion of each cycle. Each state is associated with a uQoL (and potentially a financial cost) and the cumulative quality-adjusted survival (and the financial costs) accrued over the entire process is estimated by summing the individual uQoL gains (and financial costs) yielded at the conclusion of each cycle. The effect of varying the parameter estimates and assumptions used may be explored via sensitivity analyses. A one-way sensitivity analysis involves changing just one aspect of the model (e.g. a uQoL value for a particular health state), re-running the decision model, and examining the extent to which estimates change and whether conclusions are appreciably affected. Generally a range of plausible values will be specified for each parameter built into a decision model, and values over the entire range will be tested in a sensitivity analysis. Multi-way sensitivity analyses involving varying more than one parameter simultaneously (e.g. uQoL for TOX and uQoL for REL). This can be done by simply specifying the parameter combinations of clinical interest or by employing a probabilistic approach where parameter estimates are sampled at random from probability distributions specified by the decision analyst.

Verry et al. used decision analytic modelling to explore the impact of the remaining uncertainties for sentinel node biopsy (SNB) over axillary node dissection (AND) for staging early breast cancers[2]. Randomised trials have shown SNB reduces the risk of arm morbidity and have as yet produced no convincing evidence of an increase in recurrence in the short term; however the short-term QoL benefits of SNB have the potential to be offset in the longer-term depending on the incidence of false-positive findings and the outcomes experienced by these patients. The Markov process used to evaluate this trade-off simulated a cohort of patients transitioning through a set of health states representing: full health (i.e. disease free), disease recurrence (local, axillary or distant), death due to breast cancer, or death from other causes. Transition probabilities used to govern the likelihood of moving from one state to another were informed by a review of available randomised trial

evidence for SNB and published population statistics, costs associated with each health state were estimated from the perspective of the health care system, and uQoL weights for health states were informed by published studies (e.g. breast cancer patients who answered standard gamble questions) and clinical opinion. Over a 20-year time horizon, the model found SNB was associated with a small excess in axillary recurrence and mortality, but this was outweighed by the QoL benefits associated with the less invasive nature of SNB. While such excess in recurrence of this size has not as yet been demonstrated by the available trial data, the result supports the need for longer-term follow-up.

Decision analytic modelling not only provides researchers with a framework for extrapolating trial evidence beyond the duration of a study, but also helps to apply this evidence to different subgroups of patients, and identifying priority areas for future research. The conclusions of the SNB study, for instance, were sensitive to the likelihood, and uQoL, of lymphoedema – obtaining more precise estimates of these parameters would therefore be warranted. The model was also used to explore the value of SNB in women at elevated risk of nodal involvement – a group for which limited trial evidence exists. Sensitivity analyses represent a fundamental aspect of such studies and provide valuable insights. The cost-effectiveness of endocrine therapy with tamoxifen in women at high risk of cancer, for example, was found to be particularly sensitive to the assumed duration of the breast cancer risk reduction[47]. A modelling study of trastuzumab for early breast cancer likewise found that the favourable ICER was sensitive to the duration of the treatment benefit[48]. Such findings provide a compelling case for investing research efforts in the long-term follow-up of clinical trial participants well beyond the point corresponding to the planned primary analysis.

10. Future Perspective

Quality-adjusted survival time is a valuable evaluation endpoint when trade-offs are suspected between treatment side-effects and benefits on QoL and/or survival outcomes, and between net treatment benefits versus additional financial costs. Such trade-offs have had relevance to the

assessment of treatment strategies for breast cancer, but also have applicability to a broader range of therapeutic areas.

The Q-TWiST model has proven to be a practical and informative method for analysing quality-adjusted survival; however few, if any, of the trials cited in this article appear to have been designed using sample size calculations treating quality-adjusted survival as the primary endpoint. This is likely to be due in part to the difficulties of specifying the likely effect of treatment on each of the components comprising quality-adjusted survival and the interdependencies between the components[49]. Ultimately, study design planning for trial assessing quality-adjusted survival is non-trivial and we would recommend that a series of calculations and sensitivity analyses that accommodate a range of plausible scenarios be performed to comprehensively evaluate a planned trial's statistical power. Recommendations that a 10-15% improvement in quality-adjusted survival be regarded as the smallest worthwhile difference for the planning of clinical trials have been put forward[50], however the context and setting of the individual trial will play a key role in judging what might constitute a worthwhile net clinical benefit. For example, in the absence of an important financial cost trade-off, even a very small difference in net clinical benefit may be worthwhile to detect.

Portraying the results of a Q-TWiST analysis on a threshold plane can serve as a decision aid to facilitate discussion with subsequent patients on the relative merits of a treatment and provides the opportunity to tailor recommendations base on individual patient preferences. However, an estimate of the average net-benefit, undertaken from the perspective of patients participating in the trial, is of clear value for informing health policy and funding decisions. We advocate the measurement of quality of life outcomes in oncology trials via self-administered instruments with utility-based scoring systems calibrated against patient-preferences. The assessment schedule should be carefully constructed such that the experience of participants is adequately sampled over time without placing undue burden on patients and trial personnel. Efforts should also be taken to

minimise rates of missing uQoL data as the validity of analyses performed on incomplete data, even those that use sophisticated methodologies, rest on assumptions that may not be verifiable[51].

In many cases, breast cancer therapies that have been proven to provide a clear advantage in terms of prolonging (disease-free/progression-free) survival have demonstrated side-effect profiles that are clinically manageable. The trade-off between the toxicity and potential benefits of more modern biologic agents for breast cancer (e.g. trastuzumab) may well prove to be more favourable still. The primary value of applying quality-adjusted survival analysis methods to future trials of newer treatments will probably have less to do with evaluating net-treatment benefits in the presence of a concerning trade-off, but be motivated by a need to quantify the incremental net-benefit relative to increased financial cost. Applying quality adjustments to survival gains for the purposes of cost-effectiveness analyses can have somewhat surprising consequences however – particularly when the uQoL weights are based on the perspective of the general community rather than patients themselves. For example, a randomised trial of cetuximab in advanced colorectal cancer demonstrated a modest survival benefit and a modest improvement in quality of life from cetuximab compared with best supportive care. But the benefit was reduced by around one third when expressed in terms of quality-adjusted survival time[52] based the uQoL data that had been collected from patients via an instrument with a utility-scoring system that reflected the perspective of the general population. The cost-effectiveness of life prolonging therapies in patient populations with compromised QoL therefore has the potential to appear less attractive (i.e. larger cost-effectiveness ratios) when appraised in terms of quality-adjusted survival. In the case of cetuximab for advanced colorectal cancer, the cost-per-QALY was around 50% larger when compared to the cost-per-life year gained. The cost-per-QALY would likely be lower, and therefore more attractive, if uQoL weights calibrated against patients' perspectives were applied to estimate incremental QALYs as patients tend to appraise their health state more favourably than general population samples[29-30].

Randomised controlled trials provide the best evidence on which to base an evaluation of a treatment. Performing quality-adjusted survival analyses within the context of a randomised controlled trial is therefore recommended using the 'state-based' (e.g. Q-TWiST) or repeated measures (e.g. QASA) methods that incorporate uQoL estimates obtained via instruments that reflect the attitudes of trial participants themselves. These analyses may nevertheless need to be supplemented by modelled analyses that extrapolate beyond the observed data to provide a comprehensive assessment and incorporate uQoL estimates obtained from patients outside of the context of clinical trials. Decision analytic methods provide a framework for accomplishing this, and together with sensitive analyses, help researchers gain insights and identify future research priorities. They also provide an approach for synthesising the best available evidence to inform current decision making whilst waiting for longer-term following-up evidence to accrue. Modelled analyses of quality-adjusted survival outcomes therefore provide an important adjunct to evidence from randomised controlled trials, but do not replace the need for long-term randomised evidence.

11. References

1. Montazeri A: Health-related quality of life in breast cancer patients: a bibliographic review of the literature from 1974 to 2007. *J Exp Clin Cancer Res* 27, 32 (2008).
2. Verry H, Lord Sj, Martin A *et al.*: Effectiveness and cost-effectiveness of sentinel lymph node biopsy compared with axillary node dissection in patients with early-stage breast cancer: a decision model analysis. *Brit J Cancer* 106(6), 1045-1052 (2012).
3. Abe O, Abe R, Enomoto K *et al.*: Polychemotherapy for early breast cancer: an overview of the randomised trials. *Lancet* 352(9132), 930-942 (1998).
4. Simes Rj, Coates As: Patient preferences for adjuvant chemotherapy of early breast cancer: how much benefit is needed? *J Natl Cancer Inst Monogr* (30), 146-152 (2001).
5. Duric Vm, Stockler Mr, Heritier S *et al.*: Patients' preferences for adjuvant chemotherapy in early breast cancer: what makes AC and CMF worthwhile now? *Ann Oncol* 16(11), 1786-1794 (2005).
6. Bernhard J, Zahrieh D, Zhang Jj *et al.*: Quality of life and quality-adjusted survival (Q-TWiST) in patients receiving dose-intensive or standard dose chemotherapy for high-risk primary breast cancer. *Brit J Cancer* 98(1), 25-33 (2008).
7. Corey-Lisle Pk, Peck R, Mukhopadhyay P *et al.*: Q-TWiST analysis of ixabepilone in combination with capecitabine on quality of life in patients with metastatic breast cancer. *Cancer-Am Cancer Soc* 118(2), 461-468 (2012).
8. Lumley T, Simes Rj, GebSKI V, Hudson Hm: Combining components of quality of life to increase precision and evaluate trade-offs. *Stat Med* 20(21), 3231-3249 (2001).
9. Sullivan R, Peppercorn J, Sikora K *et al.*: Delivering affordable cancer care in high-income countries. *Lancet Oncol* 12(10), 933-980 (2011).
10. Dowsett M, Cuzick J, Ingle J *et al.*: Meta-Analysis of Breast Cancer Outcomes in Adjuvant Trials of Aromatase Inhibitors Versus Tamoxifen. *J Clin Oncol* 28(3), 509-518 (2010).
11. Karnon J: Aromatase inhibitors in breast cancer - A review of cost considerations and cost effectiveness. *Pharmacoeconomics* 24(3), 215-232 (2006).
12. Verma S, Miles D, Gianni L *et al.*: Trastuzumab Emtansine for HER2-Positive Advanced Breast Cancer. *New Engl J Med* 367(19), 1783-1791 (2012).
13. Drummond Mf, Drummond Mf: Methods for the economic evaluation of health care programmes. (3rd). Oxford University Press, Oxford ; New York. (2005).
14. Torrance Gw, Thomas Wh, Sackett Dl: A utility maximization model for evaluation of health care programs. *Health Serv Res* 7(2), 118-133 (1972).
15. Fanshel S, Bush Jw: Health-Status Index and Its Application to Health-Services Outcomes. *Oper Res* 18(6), 1021-& (1970).
16. Pliskin Js, Shepard Ds, Weinstein Mc: Utility-Functions for Life Years and Health-Status. *Oper Res* 28(1), 206-224 (1980).
17. **Glasziou Pp, Cole Bf, Gelber Rd, Hilden J, Simes Rj: Quality adjusted survival analysis with repeated quality of life measures. *Stat Med* 17(11), 1215-1229 (1998).
18. Aaronson Nk, Ahmedzai S, Bergman B *et al.*: The European-Organization-for-Research-and-Treatment-of-Cancer Qlq-C30 - a Quality-of-Life Instrument for Use in International Clinical-Trials in Oncology. *J Natl Cancer I* 85(5), 365-376 (1993).
19. Torrance Gw, Feeny D: Utilities and quality-adjusted life years. *Int J Technol Assess Health Care* 5(4), 559-575 (1989).
20. Martin Aj, Glasziou Pp, Simes Rj, Lumley T: A comparison of standard gamble, time trade-off, and adjusted time trade-off scores. *Int J Technol Assess Health Care* 16(1), 137-147 (2000).

21. Hurny C, Van Wegberg B, Bacchi M *et al.*: Subjective health estimations (SHE) in patients with advanced breast cancer: an adapted utility concept for clinical trials. *Br J Cancer* 77(6), 985-991 (1998).
22. Crott R, Briggs A: Mapping the QLQ-C30 quality of life cancer questionnaire to EQ-5D patient preferences. *Eur J Health Econ* 11(4), 427-434 (2010).
23. Kim Ej, Ko Sk, Kang Hy: Mapping the cancer-specific EORTC QLQ-C30 and EORTC QLQ-BR23 to the generic EQ-5D in metastatic breast cancer patients. *Qual Life Res* 21(7), 1193-1203 (2012).
24. Grimison Ps, Simes Rj, Hudson Hm, Stockler Mr: Preliminary validation of an optimally weighted patient-based utility index by application to randomized trials in breast cancer. *Value Health* 12(6), 967-976 (2009).
25. Slevin Ml, Plant H, Lynch D, Drinkwater J, Gregory Wm: Who Should Measure Quality of Life, the Doctor or the Patient. *Brit J Cancer* 57(1), 109-112 (1988).
26. Sullivan Pw, Slejko Jf, Sculpher Mj, Ghushchyan V: Catalogue of EQ-5D scores for the United Kingdom. *Med Decis Making* 31(6), 800-804 (2011).
27. Sullivan Pw, Lawrence Wf, Ghushchyan V: A national catalog of preference-based scores for chronic conditions in the United States. *Med Care* 43(7), 736-749 (2005).
28. Dehaes Jcjm, Dekoning Hj, Vanoortmarssen Gj, Vanagt Hme, Debruyn Ae, Vandermaas Pj: The Impact of a Breast-Cancer Screening-Program on Quality-Adjusted Life-Years. *Int J Cancer* 49(4), 538-544 (1991).
29. Rosser R, Kind P: Scale of Valuations of States of Illness - Is There a Social Consensus. *Int J Epidemiol* 7(4), 347-358 (1978).
30. Slevin Ml, Stubbs L, Plant Hj *et al.*: Attitudes to Chemotherapy - Comparing Views of Patients with Cancer with Those of Doctors, Nurses, and General Public. *Brit Med J* 300(6737), 1458-1460 (1990).
31. Schleinitz Md, Depalo D, Blume J, Stein M: Can differences in breast cancer utilities explain disparities in breast cancer care? *J Gen Intern Med* 21(12), 1253-1260 (2006).
32. Boyd Nf, Sutherland Hj, Heasman Kz, Tritchler Dl, Cummings Bj: Whose Utilities for Decision-Analysis. *Med Decis Making* 10(1), 58-67 (1990).
33. Gabriel Se, Kneeland Ts, Melton Lj, Moncur Mm, Ettinger B, Tosteson Ana: Health-related quality of life in economic evaluations for osteoporosis: Whose values should we use? *Med Decis Making* 19(2), 141-148 (1999).
34. **Glasziou Pp, Simes Rj, Gelber Rd: Quality Adjusted Survival Analysis. *Statistics in Medicine* 9(11), 1259-1276 (1990).
35. Cole Bf, Gelber Rd, Goldhirsch A: Cox regression models for quality adjusted survival analysis. *Stat Med* 12(10), 975-987 (1993).
36. Cole Bf, Gelber Rd, Anderson Km: Parametric approaches to quality-adjusted survival analysis. International Breast Cancer Study Group. *Biometrics* 50(3), 621-631 (1994).
37. Cole Bf, Gelber Rd, Gelber S, Coates As, Goldhirsch A: Polychemotherapy for early breast cancer: an overview of the randomised clinical trials with quality-adjusted survival analysis. *Lancet* 358(9278), 277-286 (2001).
38. Bernhard J, Zahrieh D, Coates As *et al.*: Quantifying trade-offs: quality of life and quality-adjusted survival in a randomised trial of chemotherapy in postmenopausal patients with lymph node-negative breast cancer. *Br J Cancer* 91(11), 1893-1901 (2004).
39. Stockler Mr, Harvey Vj, Francis Pa *et al.*: Capecitabine versus classical cyclophosphamide, methotrexate, and fluorouracil as first-line chemotherapy for advanced breast cancer. *J Clin Oncol* 29(34), 4498-4504 (2011).
40. Sherrill B, Amonkar Mm, Stein S, Walker M, Geyer C, Cameron D: Q-TWiST analysis of lapatinib combined with capecitabine for the treatment of metastatic breast cancer. *Br J Cancer* 99(5), 711-715 (2008).

41. Nooij Ma, De Haes Jc, Beex Lv *et al.*: Continuing chemotherapy or not after the induction treatment in advanced breast cancer patients. clinical outcomes and oncologists' preferences. *Eur J Cancer* 39(5), 614-621 (2003).
42. Coates A, GebSKI V, Bishop Jf *et al.*: Improving the quality of life during chemotherapy for advanced breast cancer. A comparison of intermittent and continuous treatment strategies. *N Engl J Med* 317(24), 1490-1495 (1987).
43. Cole Bf, Gelber Rd, Gelber S, Mukhopadhyay P: A quality-adjusted survival (Q-TWiST) model for evaluating treatments for advanced stage cancer. *J Biopharm Stat* 14(1), 111-124 (2004).
44. Mounier N, Flechtner H, Henry-Amar M, Lepage E: Model-based methodology for analyzing incomplete quality-of-life data and integrating them into the Q-TWiST framework. *Med Decis Making* 23(1), 54-66 (2003).
45. Gelber Rd, Goldhirsch A, Cole Bf: Parametric extrapolation of survival estimates with applications to quality of life evaluation of treatments. International Breast Cancer Study Group. *Control Clin Trials* 14(6), 485-499 (1993).
46. Trippoli S, Becagli P, Messori A: Adjuvant cyclophosphamide, methotrexate and fluorouracil for node-positive breast cancer: a lifetime cost-utility analysis based on a modified Q-TWiST method. *Eur J Clin Pharmacol* 53(3-4), 281-282 (1997).
47. Eckermann Sd, Martin Aj, Stockler Mr, Simes Rj: The benefits and costs of tamoxifen for breast cancer prevention. *Aust Nz J Publ Heal* 27(1), 34-40 (2003).
48. Hall Ps, Hulme C, McCabe C, Oluboyede Y, Round J, Cameron Da: Updated cost-effectiveness analysis of trastuzumab for early breast cancer: a UK perspective considering duration of benefit, long-term toxicity and pattern of recurrence. *Pharmacoeconomics* 29(5), 415-432 (2011).
49. Murray S, Cole B: Variance and sample size calculations in quality-of-life-adjusted survival analysis (Q-TWiST). *Biometrics* 56(1), 173-182 (2000).
50. Revicki Da, Feeny D, Hunt Tl, Cole Bf: Analyzing oncology clinical trial data using the Q-TWiST method: Clinical importance and sources for health state preference data. *Qual Life Res* 15(3), 411-423 (2006).
51. Liu M, Wei L, Zhang J: Review of guidelines and literature for handling missing data in longitudinal clinical trials with a case study. *Pharm Stat* 5(1), 7-18 (2006).
52. Mittmann N, Au Hj, Tu D *et al.*: Prospective cost-effectiveness analysis of cetuximab in metastatic colorectal cancer: evaluation of National Cancer Institute of Canada Clinical Trials Group CO.17 trial. *Journal of the National Cancer Institute* 101(17), 1182-1192 (2009).

References of considerable interest

- Glasziou PP, Simes RJ, Gelber RD. Quality Adjusted Survival Analysis. *Statistics in Medicine* 1990;9:1259-76.

Details the methods for undertaking *health state-based* analyses of quality-adjusted survival.

- Glasziou PP, Cole BF, Gelber RD, Hilden J, Simes RJ. Quality adjusted survival analysis with repeated quality of life measures. *Stat Med* 1998;17:1215-29.

Details a method for undertaking a quality-adjusted survival analysis that affords flexibility to accommodate fluctuations in QoL over time.

Text Box: Executive Summary

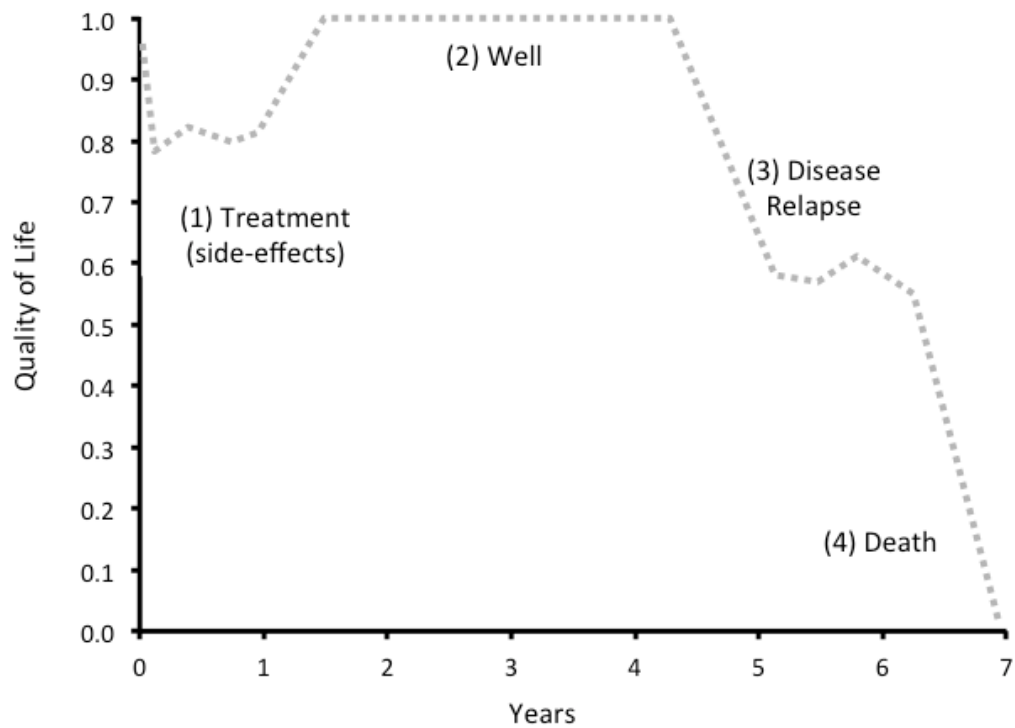
- **Introduction:** Breast cancer treatment recommendations will often require an appraisal of likely benefits in relation to likely side-effects on survival and quality of life (QoL) endpoints.
- **Quality-Adjusted Survival Analysis:** Trade-offs between QoL and survival outcomes may be formally evaluated using quality-adjusted survival (QAS) analysis methods to determine the net-benefit of a treatment.
- QAS also provides a common metric of treatment effectiveness for economic evaluations that facilitate comparisons to be made between alternative uses of health care resources.
- QAS approaches generally involve combining information from standard Kaplan-Meier survival curves with information on utility weights for QoL.
- **Utility-Based QoL Assessment:** Patient preferences and the assessment of the relative utility of the components comprising QAS are important in many instances. Most QoL instruments do not provide an assessment on the required *utility-scale* however methods are available to accomplish this either directly (e.g. using a standard-gamble question) or indirectly (e.g. using a self administered QoL questionnaire with a utility-based scoring system).
- **Application of Quality-Adjusted Survival Analysis:** QAS analysis methods have been applied in trials with breast cancer patients of adjuvant therapies as well as treatments for advanced/metastatic disease.
- Threshold analyses performed on the components of QAS may simplify the interpretation of results to facilitate treatment recommendations.
- Long-term benefits and costs may not be captured within a clinical trial. Modelled analyses that extrapolate results beyond the trial follow-up can be important complements to standard QAS analysis methods.

- **Future Perspective:** Quality-adjusted survival time is a valuable treatment evaluation endpoint when trade-offs are suspected between treatment side-effects and benefits on QoL and/or survival outcomes, and between net treatment benefits versus additional financial costs.
- In these cases we recommend QAS analysis be undertaken within the context of randomised controlled trials using practical self-administered instruments with utility-based scoring systems calibrated against patient-preferences.
- Within-trial analyses may nevertheless need to be supplemented by modelled analyses that extrapolate beyond the observed data to provide a comprehensive assessment.

Acknowledgements

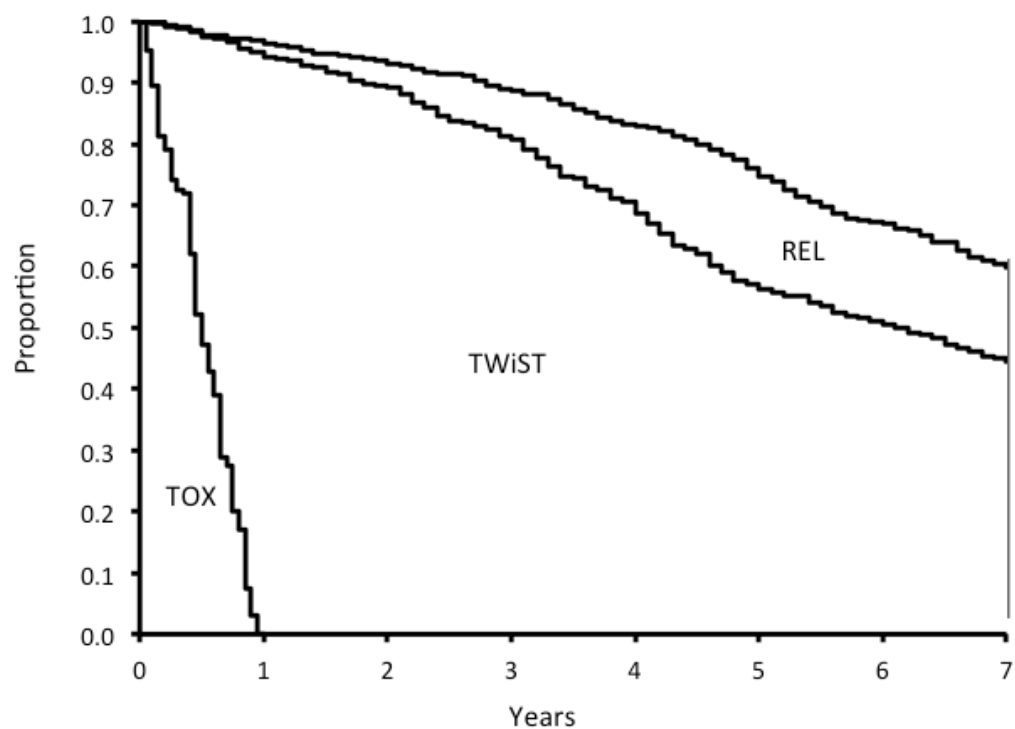
Funding assistance for this review was provided by the National Health and Medical Research Council.

Figure 1: Quality-Adjusted Life Years



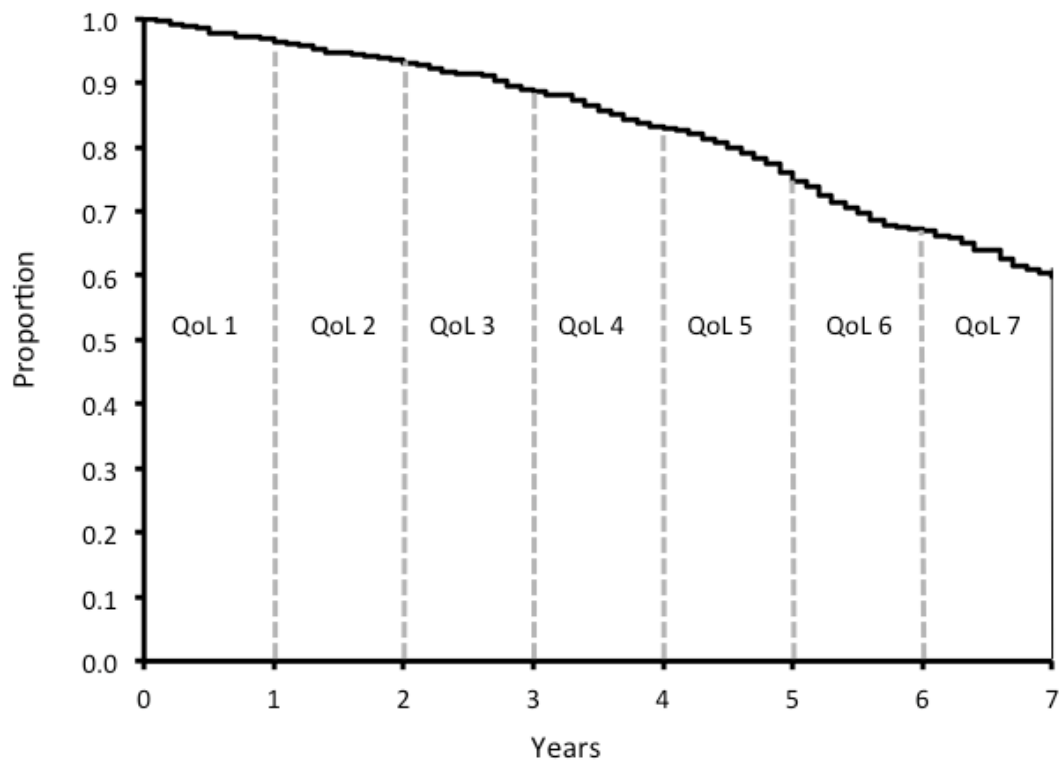
Legend: The area under the curve defined by the QoL experienced over time for a hypothetical patient represents the QALYs accrued following initial diagnosis of early breast cancer. QoL declines during the adjuvant treatment period due to side-effects and again following cancer recurrence due to disease symptoms.

Figure 2: Quality-Adjusted Time Without Symptoms or Toxicity (Q-TWiST)



Legend: The survival curve for a treatment group is partitioned into time spent in discrete health states that each has their own utility weighting for QoL. The sum of the utility-weighted area under each curve provides an estimate of quality-adjusted survival.

Figure 3: Quality-Adjusted Survival with Repeated QoL Measures



Legend: Interpolating between utility-based QoL estimates obtained at the fixed time points using a step function and combining this with an estimated survival function is analogous to partitioning the standard survival curve into discrete time periods – that each has its own utility weighing for QoL. The sum of the utility-weighted area under each curve provides an estimate of quality-adjusted survival.