Multi-Criteria Decision Analysis for Benefit-Risk Analysis and Pharmaceutical Pricing

Kao-Tai Tsai and Bruce E. Dornseif

Celgene Corporation, Summit, NJ

Abstract

As a conventional practice, the National Institute for Health and Care Excellence (NICE) of the United Kingdom utilizes quality adjusted life years (QALY) for pharmaceutical pricing. Given the multivariate nature of benefit-risk of a drug, QALY is an effort to summarize the overall efficacy and safety effects using the estimated quality of life with somewhat subjective utility weighting. With the inherent weakness of quality of life data and subjectivity of the utility weighting, benefit and risk measure using QALY can be improved. Multi-Criteria Decision Analysis (MCDA) is a multivariate approach for decision analysis which can be implemented for benefit-risk analysis purposes. Even though MCDA takes into consideration selected efficacy and safety endpoints of interest, it primarily uses the presence or absence of the endpoint events and summarizes those for the study. Since the duration of an event has larger impact to a patient's well-being and societal cost, with respect to benefit-risk than just the status of presence, we propose an approach considering the duration, instead of only the status of events. For the endpoints in the MCDA, we estimate the mean durations and the covariance matrix for all endpoints so that statistical inference can be performed.

Keywords: Benefit Risk Analysis, MCDA, QALY, Clinical Trials.

1 Introduction

Pharmaceutical pricing is an important step in the long path of drug development and in the effective delivery of a new drug to patients. It is usually based on a multifaceted consideration which combines the various aspects of a new drug's benefit and risk and its societal impact. As a conventional practice, the NICE of UK utilizes QALY as the "currency" in health technology appraisals (HTA) [1]. A QALY is the amount of health represented by a year of life at full health. It provides an estimate of how much additional length of life and quality of life a

person might gain as a result of treatment. Health gains from different treatments can be expressed in terms of the number of QALY so that they can be compared. Using QALY, the pricing threshold would be expressed as a cost per QALY gained. Given the multivariate characteristics of benefit-risk of a drug, the consideration of QALY is an effort to summarize the overall drug efficacy and safety effects using the estimated quality of life along with certain subjective utility weighting. With the inherent weakness of quality of life data and subjectivity of the utility weighting, benefit and risk measurement using QALY can be improved to better reflect the merits included in the study data and other extraneous factors not captured therein.

Multi-Criteria Decision Analysis (MCDA) is a multivariate approach for decision analysis which can be applied for benefit-risk analysis, and hence pharmaceutical pricing. Selected multi-factors can be incorporated into the data analysis simultaneously and ultimately utilized for decision making. Many practices of MCDA take into consideration the selected efficacy and safety endpoints of interest; however, they primarily use the presence or absence status of the endpoint events and summarize those for the study.

Since the duration of an event has larger impact to a patient's well-being and societal cost than just the status of event presence, we propose an approach considering the event durations instead of the status of events only. For the endpoints included in the MCDA, we estimate the mean durations and the covariance matrix for all endpoints so that statistical inference can be performed. The method we propose can incorporate both study data and extraneous factors in statistical inference. We use data from a recent clinical trial to illustrate the implementation of our proposed method.

2 QALY

As a brief review of the concept of QALY, let h_i be the ith health state, where $i = 1, 2, \cdots, n$; $\mu(h)$ be the utility coefficient of health state h, that is the rate of utility accrual in that state relative to good health, and $S(y)$ be a function of life year, then QALY can be defined as

$$\text{QALY}(h, y) = \sum_{i=1}^{n} \mu(h_i)\big(S(y_i) - S(y_{i-1})\big)$$

where y_{i-1} to y_i is the time interval spent in health state i. If the health states change continuously, then QALY can be expressed as the following integral for some predefined time L

$$\text{QALY} = \int_0^L \mu(t)S(t)dt,$$

where $S(t)$ can be estimated using Kaplan-Meier's method. Figure 1 is a schematic graph for a patient's QALY, which can be estimated by the area under the curve.

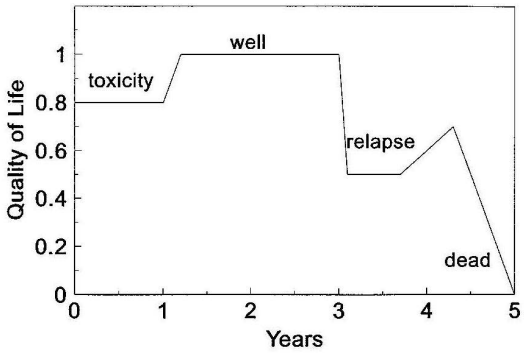

Figure 1: Schematic graph for a patient's QALY

In health technology appraisals, QALY is most appropriate for severe or terminal diseases. It is a proxy of overall drug effect for efficacy and safety. Even though it may consider the differential weight of various factors associated with efficacy and safety of a new drug implicitly, QALY cannot easily incorporate factors regarding the degree of "burden of illness," the degree of therapeutic innovation and improvements, and the wider societal benefits. In addition, QALY can be confounded by a subject's demographics and other concurrent health issues.

3 Value-Based Pharmaceutical Pricing

In recent years, advocates in the industry suggest using Value-Based Pharmaceutical pricing (VBP) to strike a balance between delivering reasonable prices for the health services and ensuring that the industry is still incentivized to undertake R&D, and market new and improved medicines [2, 3, 4]. Utilization of VBP should improve outcomes for patients through better access to effective medicines, stimulate innovation and development of high value treatments, improve the process for assessing new medicines, ensuring transparent, predictable and timely decision-making. Utilization of VBP should include a broad assessment, alongside clinical effectiveness, of the range of factors through which medicines deliver benefit for patients and society, and ensure the greatest value for money and best use of health service resources.

The U.S. and EU governments propose that the price threshold structure should consider the basic benefit of a new medicine, the degree of "burden of illness," the degree of therapeutic innovation and improvements, and the wider societal benefits. With a VBP system, the government could apply weightings to the benefits provided by new medicines, which would imply a range of price thresholds. These price thresholds would be explicitly adjusted to reflect a broader

range of relevant factors, so they could be used to calculate the full value of a new product.

Sine pharmaceutical pricing is essentially a decision based on the benefit-risk trade-off for a health technology, the methodologies to evaluate benefit-risk can be strategically applied to pharmaceutical pricing policies. In the context of benefit-risk analysis, researchers have suggested using MCDA, which includes the quality adjusted time without symptom and toxicity (Q-TWiST) statistic as one possible measure. In this research, we combine the principle of MCDA and the measure of Q-TWiST to propose a method for VBP.

4 The Principles of MCDA

MCDA is a method to consider multiple criteria simultaneously for decision making. The merits (usually expressed in a score) of each criterion is calculated based on some pre-defined benchmarks. A weighting scheme is then used to weight the scores and to obtain a total score of outcomes for MCDA. To utilize the MCDA procedure, this approach requires critical assumptions about benchmarks for scoring and the weights for each criterion. In general practice, one usually estimates the weighted scores but not the associated variance, which has become a common criticism in the implementation of the MCDA procedure. In this research, we estimate the Q-TWiST-type statistic with multiple efficacy and safety endpoints. We also derive its associated variance so that treatment effects can be compared. Since the weighting scheme can be subjective, we also perform sensitivity analyses by varying the weighting (or utility) functions so that the treatment effects can be compared under a variety of scenarios.

4.1 Historical Research on Q-TWiST

In the *International Breast Cancer Study Group* in the early 1990s, researchers conducted a treatment effect investigation under adjuvant treatment settings [6, 7, 8, 9, 10, 11]. The stages of treatment outcomes considered including toxicity from therapy, time without symptoms and toxicity, disease relapse, and death.

Under neo-adjuvant treatment settings, treatments can also lead to toxicity, response, or death. The order of responses and toxicities can have multiple combinations (except for death) and the same outcome may recur multiple times, especially in multi-cycle treatments.

Common methods utilized in published clinical research usually consider the frequencies of each type of outcome, and/or duration of each type of outcome, which do not always provide a detailed insight of interest.

5 Framework of Utility Adjusted Models

For patient i, let the total number of health outcome stages be r_i, and the outcome stages do not change continuously, the total amount of time spent in stage k be S_{k_i}, and the utility coefficient of stage k be μ_k. Then the utility adjusted score for each patient can be defined as

$$\omega_i = \sum_{k=1}^{r_i} \mu_k S_{k_i}.$$

5.1 Modeling without Censored Observations

Under the scenario of no censored data, the utility adjusted score for each patient can be estimated. Assume the scores for treatment groups A and B are as follows:

$$\Omega_A = \{\omega_{1_A}, \omega_{2_A} \cdots, \omega_{n_A}\}$$
$$\Omega_B = \{\omega_{1_B}, \omega_{2_B}, \cdots, \omega_{n_B}\},$$

then the difference in statistics derived from utility adjusted scores,

$$\Delta = \hat{\Omega}_A - \hat{\Omega}_B,$$

can be estimated and tested for a given significance level using either parametric or nonparametric procedures.

5.2 Modeling with Censored Data

When a study consists of censored data, let S_{k_1} be the function of time-to-onset of stage k, and S_{k_2} be the function of time-to-end of stage k, then the total duration of stage k is

$$\tau_k = \int_0^L \Big\{ S_{k_2}(t) - S_{k_1}(t) \Big\}\, dt. \tag{1}$$

The utility adjusted score for treatment group A can then be defined as

$$\Omega_A = \sum_{k=1}^{r} \mu_k \tau_k.$$

5.3 Q-TWiST Score in Adjuvant Treatment Settings

For simplicity, with possible censoring, let $S_1(t)$ be the time-to-end-of-Toxicity (or time-to-TWiST), $S_2(t)$ be the time-to-end-of-TWiST (or time-to-Relapse), and $S_3(t)$ be the time-to-end-of-Relapse (or time-to-death/lost-to-follow-up). Then the utility adjusted score can be expressed as

$$\begin{aligned}\Omega_{\text{qtwist}} = \mu_{\text{tox}} \int_0^L S_1(t)\, dt + \mu_{\text{qtwist}} \int_0^L \Big\{ S_2(t) - S_1(t) \Big\}\, dt \\ + \mu_{\text{rel}} \int_0^L \Big\{ S_3(t) - S_2(t) \Big\}\, dt,\end{aligned} \tag{2}$$

with respective utility coefficients $\mu_{\text{tox}}, \mu_{\text{qtwist}}$, and μ_{rel}.

As mentioned previously, the utility coefficient is a weight and its choice can be subjective. Alternatively, a patient specific *quality-of-life* score can be used as utility. In addition, in multiple-cycle therapies, the same outcomes may appear in successive cycles and multiple outcomes may occur simultaneously. In our proposed method, we aggregate the durations of multiple occurrences of the same outcome for each patient in order to estimate Ω_{qtwist}.

6 Estimation of Parameter and Variance

To estimate Ω_{qtwist} and variance, let S_i be the function of time-to-event of stage i, U_i be the censoring random variable independent of S_i, $\Delta_i = I(S_i < U_i)$ be the failure indicator, $X_i = \min(S_i, U_i)$ be the observable event time, and $\lambda_i(t)$ be the marginal hazards at time t for S_i. For example, if $i = 3$, the Ω_{qtwist} can be estimated empirically by

$$\begin{aligned}\hat{\Omega}_{\text{qtwist}} = \mu_{\text{tox}} \int_0^L \hat{S}_1(t)\ dt + \mu_{\text{qtwist}} \int_0^L \left\{\hat{S}_2(t) - \hat{S}_1(t)\right\} dt \\ + \mu_{\text{rel}} \int_0^L \left\{\hat{S}_3(t) - \hat{S}_2(t)\right\} dt = \sum_{k=1}^{3} \eta_k \int_0^L \hat{S}_k(t)\ dt. \end{aligned} \tag{3}$$

For the general case, the covariance matrix between $\hat{S}_i$ and $\hat{S}_j$ at time t_1, t_2 can be expressed as:

$$\begin{aligned}\sigma_{ij}(t_1, t_2) &= \text{cov}\Big[\sqrt{n}\{\hat{S}_i(t_1) - S_i(t_1)\}, \sqrt{n}\{\hat{S}_j(t_2) - S_j(t_2)\}\Big] \\ &= S_i(t_1)S_j(t_2) \int_0^{t_1} \int_0^{t_2} G_{ij}(u, v) du dv, \end{aligned} \tag{4}$$

where

$$\begin{aligned}G_{ij}(u, v) = \frac{P(X_i \geq u, X_j \geq v)}{P(X_i \geq u)P(X_j \geq v)} \\ \times \Big[\lambda_{ij}(u, v) - \lambda_{i|j}(u|v)\lambda_j(v) - \lambda_{j|i}(v|u)\lambda_i(u) + \lambda_i(u)\lambda_j(v)\Big], \end{aligned} \tag{5}$$

and the joint hazard of S_i and S_j at time u, v being

$$\begin{aligned}\lambda_{ij}(u, v) = \lim_{\Delta u, \Delta v \to 0} P(u \leq X_i < u + \Delta u, v \leq X_j < v + \Delta v, \\ \Delta_i = \Delta_j = 1 \mid X_i \geq u, X_j \geq v)/(\Delta u \Delta v), \end{aligned} \tag{6}$$

the conditional hazard of $S_i|S_j$ at time u, v being

$$\lambda_{i|j}(u|v) = \lim_{\Delta u \to 0} P(u \leq X_i < u + \Delta u, \Delta_i = 1 \mid X_i \geq u, X_j \geq v)/(\Delta u), \tag{7}$$

and the conditional hazard of $S_j|S_i$ at time v, u being

$$\lambda_{j|i}(v|u) = \lim_{\Delta v \to 0} P(v \leq X_j < v + \Delta v, \Delta_j = 1 \mid X_i \geq u, X_j \geq v)/(\Delta v). \quad (8)$$

The covariance between $\int_0^L \hat{S}_i(t)dt$ and $\int_0^L \hat{S}_j(t)dt$ can be estimated by

$$\hat{V}_{ij} = \frac{1}{n} \int_0^L \int_0^L \Big\{ \int_v^L \hat{S}_j(t)dt \Big\} \Big\{ \int_u^L \hat{S}_i(t)dt \Big\} \hat{G}_{ij}(u, v) \; dudv. \quad (9)$$

Hence, the estimated variance of $\hat{\Omega}_{\text{qtwist}}$ is

$$\text{Var}(\hat{\Omega}_{\text{qtwist}}) = \sum_{i=1}^{r} \sum_{j=1}^{r} \eta_i \eta_j \hat{V}_{ij}.$$

If the utility function is random,

$$\begin{aligned} \text{Var}(\hat{\Omega}_{\text{qtwist}}) &= \text{Var}\Big\{ E(\hat{\Omega}_{\text{qtwist}}|\hat{\eta}) \Big\} + E\Big\{ Var(\hat{\Omega}_{\text{qtwist}}|\hat{\eta}) \Big\} \\ &= \text{Var}\Big\{ \sum_{i=1}^{R^*} \hat{\eta} \int_0^L S_i(t)dt \Big\} + E\Big\{ \sum_{i=1}^{r} \sum_{j=1}^{r} \hat{\eta}_i \hat{\eta}_j V_{ij} \Big\}. \end{aligned} \quad (10)$$

The covariance matrix is derived, based on the asymptotic results proposed by Nelson-Alann using a counting process [5]. When the sample size or the number of events is not sufficiently large, the consistency of the estimated covariance matrix may need further investigation. Some authors suggest using either a permutation test or bootstrap method as an alternative approach to estimate the empirical sample variation.

7 Treatment Comparisons

Statistically, the treatments can be compared via the standardized $\hat{\Omega}_{\text{qtwist}}$

$$(\hat{\Omega}_1 - \hat{\Omega}_2)\text{Var}(\hat{\Omega}_{\text{qtwist}})^{-1}(\hat{\Omega}_1 - \hat{\Omega}_2)' \sim \chi^2(df), \quad (11)$$

where Ω_i denotes the Ω_{qtwist} for the ith group. However, in VBP setting, other non-statistical factors, such as R&D incentive, societal effect, degree of innovation, etc., are also considered. The "subjective" weights can be incorporated into this equation by

$$(\hat{\Omega}_1 - \hat{\Omega}_2)D^{1/2}\text{Var}(\hat{\Omega}_{\text{qtwist}})^{-1}D^{1/2}(\hat{\Omega}_1 - \hat{\Omega}_2)' \sim \chi^2(df), \quad (12)$$

where D is a diagonal matrix with factor weights for HTA considerations. Equation (12) may still follow a χ^2 distribution, however, the degrees of freedom may need to be adjusted accordingly.

8 Illustrative Example

Approximately 700 patients were randomized to a study with two treatment groups. Efficacy (i.e., benefit) endpoints to be considered are complete response, partial response, stable disease, and progression disease. Safety (i.e., risk) endpoints considered are infection and blood disorder. The duration of the endpoint is defined as the summation of individual differences between time-to-onset and time-to-the-end of the endpoint event at different cycle of the treatment period. Figure 2 to Figure 7 show the duration of each event for both treatment groups, and the duration for all the events in each group are shown respectively in Figure 8 and Figure 9.

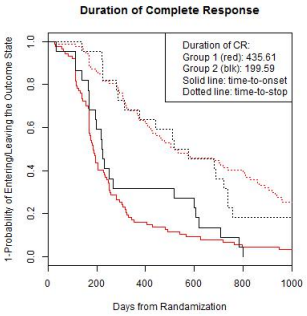

Figure 2: Complete Response

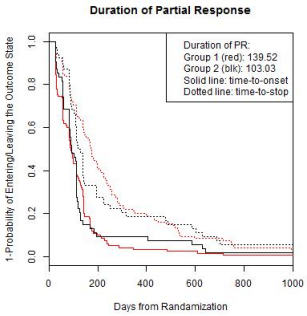

Figure 3: Partial Response

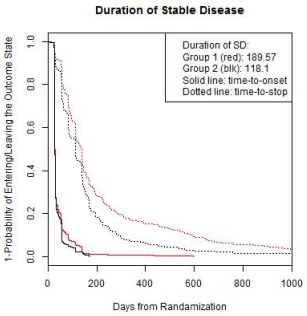

Figure 4: Stable Disease

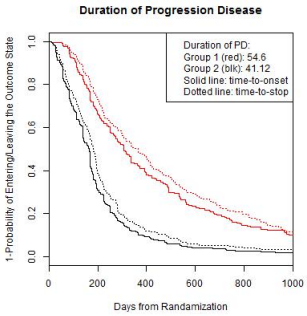

Figure 5: Progression Disease

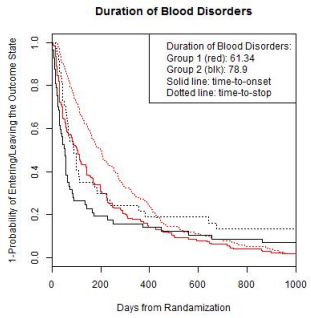

Figure 6: Blood Disorder

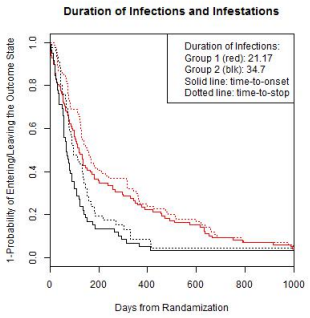

Figure 7: Infection

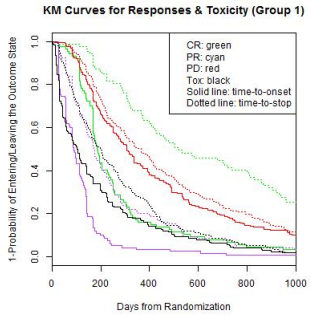

Figure 8: All Cretiria for Group 1

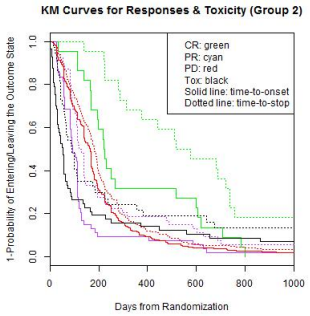

Figure 9: All Cretiria for Group 2

8.1 Estimated Means and Covariance Matrix

The estimated means of event onset and event end for group 1 are shown below

```
[1,] 280.28 718.14 119.38 263.26 44.37 236.15 455.52 510.22 193.06 274.87 258.85 296.48
```

and the estimated covariance matrix between all event onsets and ends for group 1 is also shown below:

```
       [,1]    [,2]   [,3]    [,4]   [,5]   [,6]    [,7]    [,8]   [,9]  [,10]   [,11]   [,12]
 [1,] 868.74  563.97 439.93  798.87  -4.61 531.55  614.18  570.37 -10.99  68.26  560.78  643.33
 [2,] 563.97 3045.16 426.68  962.63  -6.05 796.32 2888.11 2923.57 310.49 373.66 1097.41 1205.49
 [3,] 439.93  426.68 155.26  148.22  -3.18 115.74   40.18   52.44 131.49 240.36  167.10  116.49
 [4,] 798.87  962.63 148.22  682.79 -11.02 191.13  146.46  178.09 213.52 404.51 1200.13  928.77
 [5,]  -4.61   -6.05  -3.18  -11.02   8.34   8.93   -0.25   -0.37  -2.93  -3.25    9.58   11.79
 [6,] 531.55  796.32 115.74  191.13   8.93 318.00  187.73  199.01  59.12  90.00  578.22  697.03
 [7,] 614.18 2888.11  40.18  146.46  -0.25 187.73  967.22  885.78 254.45 468.10  427.87  632.80
 [8,] 570.37 2923.57  52.44  178.09  -0.37 199.01  885.78 1093.54 278.13 594.13  445.47  703.17
 [9,] -10.99  310.49 131.49  213.52  -2.93  59.12  254.45  278.13 323.95 330.84  214.30  272.18
[10,]  68.26  373.66 240.36  404.51  -3.25  90.00  468.10  594.13 330.84 414.54  370.61  394.90
[11,] 560.78 1097.41 167.10 1200.13   9.58 578.22  427.87  445.47 214.30 370.61 1029.73  974.80
[12,] 643.33 1205.49 116.49  928.77  11.79 697.03  632.80  703.17 272.18 394.90  974.80 1234.22
```

Similar statistics for group 2 are shown below:

```
[1,] 333.68 656.14 144.7 251 37.83 159.06 202.71 246.68 179.09 277.49 136.88 183.57
```

```
         [,1]     [,2]    [,3]    [,4]   [,5]    [,6]    [,7]    [,8]     [,9]    [,10]   [,11]   [,12]
 [1,]  2530.08  1622.47  644.45  564.58   6.17  672.20  -82.74 -168.14 -2966.52 -1822.01 3161.58 3036.31
 [2,]  1622.47  9025.72 3377.66 3040.68  -8.89 3113.98 1441.20 2906.42 -2995.37  -159.11 -308.65 -314.27
 [3,]   644.45  3377.66  722.10  594.99 -16.28  255.88  380.51  410.60  1890.34  6235.39  -44.90  -30.15
 [4,]   564.58  3040.68  594.99 1789.53 -12.31  318.00  978.33 1231.40  3263.35  8686.66 -180.25 -181.82
 [5,]     6.17    -8.89  -16.28  -12.31   2.02   -0.10   -0.40   -1.06     0.85     7.42   -0.84   -4.14
 [6,]   672.20  3113.98  255.88  318.00  -0.10  150.99  124.22  140.99   820.54  1481.34   59.74   61.16
 [7,]   -82.74  1441.20  380.51  978.33  -0.40  124.22  156.09  195.12   677.28  1135.70  204.09  246.47
```

```
 [8,]  -168.14  2906.42  410.60 1231.40  -1.06  140.99  195.12  339.74  1271.92  2562.45  215.96  257.40
 [9,] -2966.52 -2995.37 1890.34 3263.35   0.85  820.54  677.28 1271.92  1771.66  2684.34  948.71 1902.78
[10,] -1822.01  -159.11 6235.39 8686.66   7.42 1481.34 1135.70 2562.45  2684.34  4145.02  953.39 1454.01
[11,]  3161.58  -308.65  -44.90 -180.25  -0.84   59.74  204.09  215.96   948.71   953.39  991.01  746.23
[12,]  3036.31  -314.27  -30.15 -181.82  -4.14   61.16  246.47  257.40  1902.78  1454.01  746.23 1692.30
```

Similar statistics for groups 1 and 2 together are shown below:

```
[1,] 291.06 705.62 127.33 259.41 41.16 198.27 307.12 355.52 189.61 275.38 213.01 256.52

       [,1]    [,2]   [,3]   [,4]  [,5]   [,6]    [,7]    [,8]   [,9]  [,10]  [,11]  [,12]
 [1,] 660.73  455.81 544.38 733.36 -1.70 371.42  272.15  271.73 -52.52  62.30 670.18 746.54
 [2,] 455.81 2313.34 583.62 977.44 -3.93 736.46 1674.94 2052.44 204.32 350.67 711.99 714.87
 [3,] 544.38  583.62 145.05 168.14 -4.40 100.88   68.85   80.93 208.37 366.25 106.84  72.09
 [4,] 733.36  977.44 168.14 497.94 -8.57 139.57  209.09  280.82 337.66 590.31 789.87 584.02
 [5,]  -1.70   -3.93  -4.40  -8.57  2.66   3.20    0.46    0.22  -2.28  -2.19   5.68   5.59
 [6,] 371.42  736.46 100.88 139.57  3.20 121.28   96.18  106.12  89.06 127.90 288.47 359.54
 [7,] 272.15 1674.94  68.85 209.09  0.46  96.18  251.25  300.71 215.50 373.27 285.15 361.50
 [8,] 271.73 2052.44  80.93 280.82  0.22 106.12  300.71  338.88 293.97 562.42 312.23 401.93
 [9,] -52.52  204.32 208.37 337.66 -2.28  89.06  215.50  293.97 291.83 363.65 292.78 410.95
[10,]  62.30  350.67 366.25 590.31 -2.19 127.90  373.27  562.42 363.65 427.98 470.64 569.09
[11,] 670.18  711.99 106.84 789.87  5.68 288.47  285.15  312.23 292.78 470.64 563.39 622.69
[12,] 746.54  714.87  72.09 584.02  5.59 359.54  361.50  401.93 410.95 569.09 622.69 749.61
```

The differences of estimated means for each endpoint event between groups are shown below:

```
Difference of Mean-Time-to-CR                -53.405943
Difference of Mean-Time-to-End-CR             62.001565
Difference of Mean-Time-to-PR                -25.322331
Difference of Mean-Time-to-End-PR             12.262710
Difference of Mean-Time-to-SD                  6.534527
Difference of Mean-Time-to-End-SD             77.086107
Difference of Mean-Time-to-PD                252.812119
Difference of Mean-Time-to-End-PD            263.539708
Difference of Mean-Time-to-Blood-AE           13.969759
Difference of Mean-Time-to-End-Blood-AE       -2.618936
Difference of Mean-Time-to-Infection-AE      121.965578
Difference of Mean-Time-to-End-Infection-AE 112.910966
```

and the differences of mean duration of each event stage between groups are:

```
          CR    PR    SD    PD Blood-Disorder Infection
[1,] 115.41 37.59 70.55 10.73   -16.59        -9.05
```

9 Utility Coefficients for Sensitivity Analysis

To better investigate the subjectivity of utility coefficients, perturbations are considered for various scenarios of utility coefficient choices. The initial set of utility coefficients are shown in the second column of Table 1, and the coefficients used for the sensitivity analysis are shown in the last column of the table. The results are summarized in figures and shown in the next section.

Table 1: **Utility Coefficient Values and Perturbations**

Event Considered	Initial Weights	Perturbations
CR	1	1
PR	0.5	0.2 (0.1) 0.8
SD	0.0	-0.2 (0.1) 0.2
PD	-0.5	-0.7 (0.1) -0.3
Blood Disorder	-0.25	-0.45 (0.1) -0.05
Infection	-0.25	-0.45 (0.1) -0.05

10 Asymptotic Covariance Adjusted MCDA

Using the clinical trial data with the weights shown in Table 1, we estimate the MCDA score differences between the two treatment groups. The vertical solid line in Figure 10 is the estimate of MCDA (3.02) for treatment differences with initial weights and the dotted lines are the 5% and 95% quantiles of MCDA differences of the sensitivity analysis. The vertical solid line in Figure 11 is the estimated p-value of MCDA (0.00126) for treatment difference with initial weights, and the dotted lines are the 5% and 95% quantiles of p-values of MCDA differences for the sensitivity analysis.

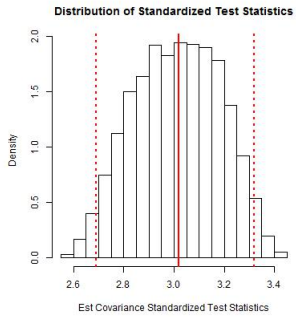

Figure 10: Distribution of MCDA statistics

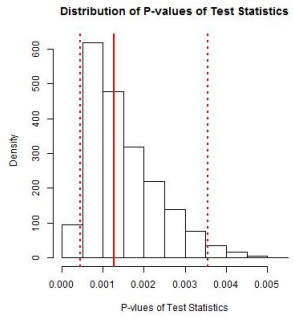

Figure 11: Distribution of p-values of MCDA statistics

The dotted lines in Figure 12 are the 5% and 95% quantiles of MCDA differences from bootstrap samples. The dotted lines in Figure 13 are the 5% and 95% quantiles of MCDA differences from permutation samples. The results show a highly significant difference of MCDA scores between the groups in p-value and an appreciable difference of MCDA of 3.02 with a 95% confidence level $(2.69, 3.32)$. The significance of a MCDA score of 3.02 will have to be carefully interpreted based on clinical considerations.

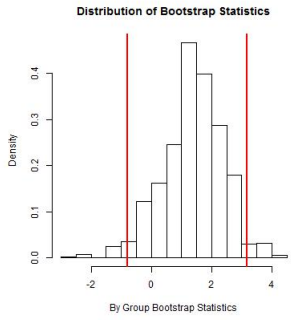

Figure 12: Distribution of bootstrap statistics

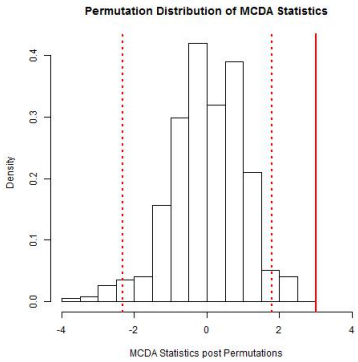

Figure 13: Distribution of permutation statistics

11 Summary

Benefit-Risk evaluation is an integral part of drug discovery and development. MCDA is a framework for a comprehensive analysis of benefit-risk evaluation among the methodologies utilized by pharmaceutical companies and regulatory agencies worldwide. In this article, we propose a method which combines the principle of MCDA and the concept of Q-TWiST to provide an estimated statistic and its corresponding variance. This proposed method extends the QALY method which is the primary methodology in conventional HTA for pharmaceutical pricing. This extension considers the duration of each criterion selected in the analysis, and it also provides the flexibility to incorporate extraneous factors which are usually not collected in clinical trials, therefore allows analysts to combine all relevant factors to properly set the policies for pharmaceutical pricing. Since the proposed method also provides the variance estimate of the difference of MCDA statistics, it can also be used to compare treatment effect and therefore allow the analysts to better set relative pricing among various kinds of health technologies.

References

[1] National Institute for Health and Clinical Excellence (NICE). Guide to the methods of technology appraisal. London: NICE, 2008.

[2] Drummond M. Value based pricing in the United Kingdom. Conference abstract ISPOR 2011. http://www.ispor.org/congresses/spain1111 /presentations/IP2drummondmichael.pdf.

[3] A new value-based approach to the pricing of branded medicines a consultation. Department of Health. December 2010.

[4] Value-based pricing: Key recommendations from ABPI. ABPI 2011. (http://www.abpi.org.uk/our-work/policyparliamentary/Pages/vbp.aspx)

[5] Fleming, T. R. and Harrington, D. P. (1991). *Counting Processes and Survival Analysis*. New York: Wiley.

[6] Gelber, R. D., Goldhirsch, A., and Cavalli, F. for the International Breast Cancer Study Group (IBCSG). (1991). Quality-of-life-adjusted evaluation of a randomized trial comparing adjuvant therapies for operable breast cancer. *Annals of Internal Medicine*, v 114, 621-628.

[7] Gelber, R. D., Goldhirsch, A., Hiirny, C., Bernhard, J., and Simes, R. J. for the International Breast Cancer Study Group (IBCSG). (1992a). Quality of life in clinical trials of adjuvant therapies. *Journal of the National Cancer Institute Monographs*, v 11, 127-135

[8] Gelber, R. D., Cole, B. F., Gelber, S., and Goldhirsch, A. (1995). Comparing treatments using quality-adjusted survival: The Q-TWiST method. *American Statistician*, v49, 161-169.

[9] Gelber, R. D., Goldhirsch, A,, Cole, B. F., Wieand, H. S., Schroeder, G., and Krook, J. E. (1996). A quality adjusted time without symptoms or toxicity (Q-TWiST) analysis of adjuvant radiation therapy and chemotherapy for resectable rectal cancer. *Journal of the National Cancer Institute*, v88, 1039-1045.

[10] Glasziou, P. P., Simes, R. J., and Gelber, R. D. (1990). Quality adjusted survival analysis. *Statistics in Medicine*, v9, 56-64.

[11] Murray, S., Cole, B. (2000). Variance and Sample Size Calculations in Quality-of-Life-Adjusted Survival Analysis (Q-TWiST). *Biometrics*, v56, 173-182.