Cancer Chemother Pharmacol (1987) 20: 213–218

Cancer Chemotherapy and Pharmacology

Towards an assessment of toxicity in the treatment of ovarian cancer*

L. Dones[1], E. Wiltshaw[2], B. G. Birkhead[1], and R. R. P. Jackson[1]

[1] Clinical Operational Research Unit, Department of Statistical Science, University College London
[2] Royal Marsden Hospital, London

Summary. Using data from a randomised clinical trial of two platinum drugs at The Royal Marsden Hospital, London, a descriptive model of toxicity has been developed and tested. Toxicity manifests itself in reducing the effectiveness of different body processes; five were selected in this study as being most critical. The model summarises and combines data from these five sites in terms of a clinician's assessment of the associated risk to the patient. It is hoped that the approach will help clarify toxicity information for use in patient management decision-making and in the reporting of clinical trial results.

Introduction

The treatment of ovarian cancer with platinum-based drugs often leads to situations in which a tumour responds well to treatment but the patient suffers unpleasant and possibly life-threatening toxicity. Difficult decisions about whether to continue or change treatment then have to be made. A method of combining the levels of toxicity in various sites of the body into one holistic score of toxicity would make this task easier.

A randomised trial between cisplatin and JM8 (carboplatin) accumulated 117 patients over the period August, 1981 to April, 1984. The major difference expected between the two drugs was in their toxic effects. This provided an opportunity to compare groups of patients on the basis of the toxic effects of the drugs alone.

Patients and methods

The patients in the trial were all primary ovarian cancer patients with FIGO stage III or IV disease. Patients could be crossed over from one arm of treatment to another because of toxicity, lack of response or both. As a consequence, patients received differing amounts of the drugs. The intention was to give each patient ten courses of treatment. All those who were clinically disease-free at the end of five courses were then surgically assessed. Patients who had not responded after two or three courses were usually switched to the other treatment arm; if response was still not observed, other therapy was given or treatment was discontinued. Eight patients were still undergoing treatment at the time of analysis.

Figure 1 shows the sequences in which drugs were given and the number of patients in each group.

For each patient in the trial, toxicity at each of five sites was to be evaluated with reference to the following scales:

1. Bone marrow: the absolute white blood cell count ($0-3000 \times 10^6/l$).
2. Kidneys (RENAL): the per cent reduction from the pretreatment level of glomerula filtration rate.
3. Inner ear (AUDITORY): the per cent high tone hearing loss, based on audiometry.
4. Gastro-intestinal tract (GIT): the frequency and duration of nausea, vomiting and/or diarrhoea, where particular points on the toxicity scale are described as being equivalent to a certain percentage toxicity score.
5. Peripheral nervous system (PNS): the degree of peripheral neuropathy (from paraesthesia to paralysis of hands and/or feet), measured subjectively on a scale of 0 to 1.

Utility assessment

The objective of the research was to establish a relationship between toxicity assessments and the level of utility which the clinician associated with them. "Utility" reflects the clinician's perception of the importance of the toxic effects in terms such as the risk and discomfort to the patient. Utilities were measured on a scale from 0 (maximum risk/discomfort/etc.) to 1 (miniumum risk/discomfort/etc.).

For selected points on each of the toxicity scales defined in the previous section, the clinician was asked to assess the corresponding utility to the patient ("u score"), and a smooth curve of utility versus toxicity was drawn.

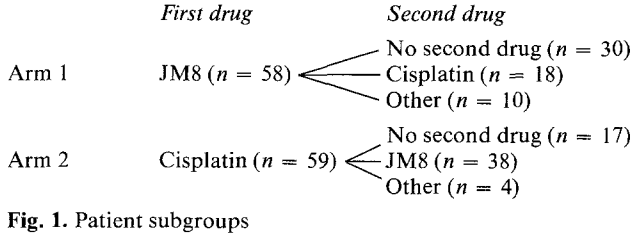


Fig. 1. Patient subgroups

* This work was funded by the Department of Health and Social Security and the Imperial Cancer Research Fund

Offprint requests to: L. Dones, Clinical Operational Research Unit, Department of Statistical Science, University College London, Gower Street, London WC1E, 6BT, England

Details of the questioning procedure used to elicit utility values are given in Appendix A, and the curves themselves are shown in Fig. 2. By interpolation, the curves can be used to read off the utility corresponding to any toxicity level in the measured range.

To test the reproducibility of the curves, the process was repeated some months later. The results are shown alongside the original curves in Fig. 2. For each site, the two curves were close together and of the same underlying shape.

Using a similar questioning procedure and the methods developed by Keeney and Raiffa [1], an overall utility score, U (a composite of the utility scores in individual sites), was developed. The mathematical expression used for combining scores is chosen according to the degree of dependence found between them. In this case, a multiplicative model for combining scores was found to be appropriate (see Appendix B). The strength of the overall U score is that it provides a means of comparing patients who have different levels of toxicity across the five sites in

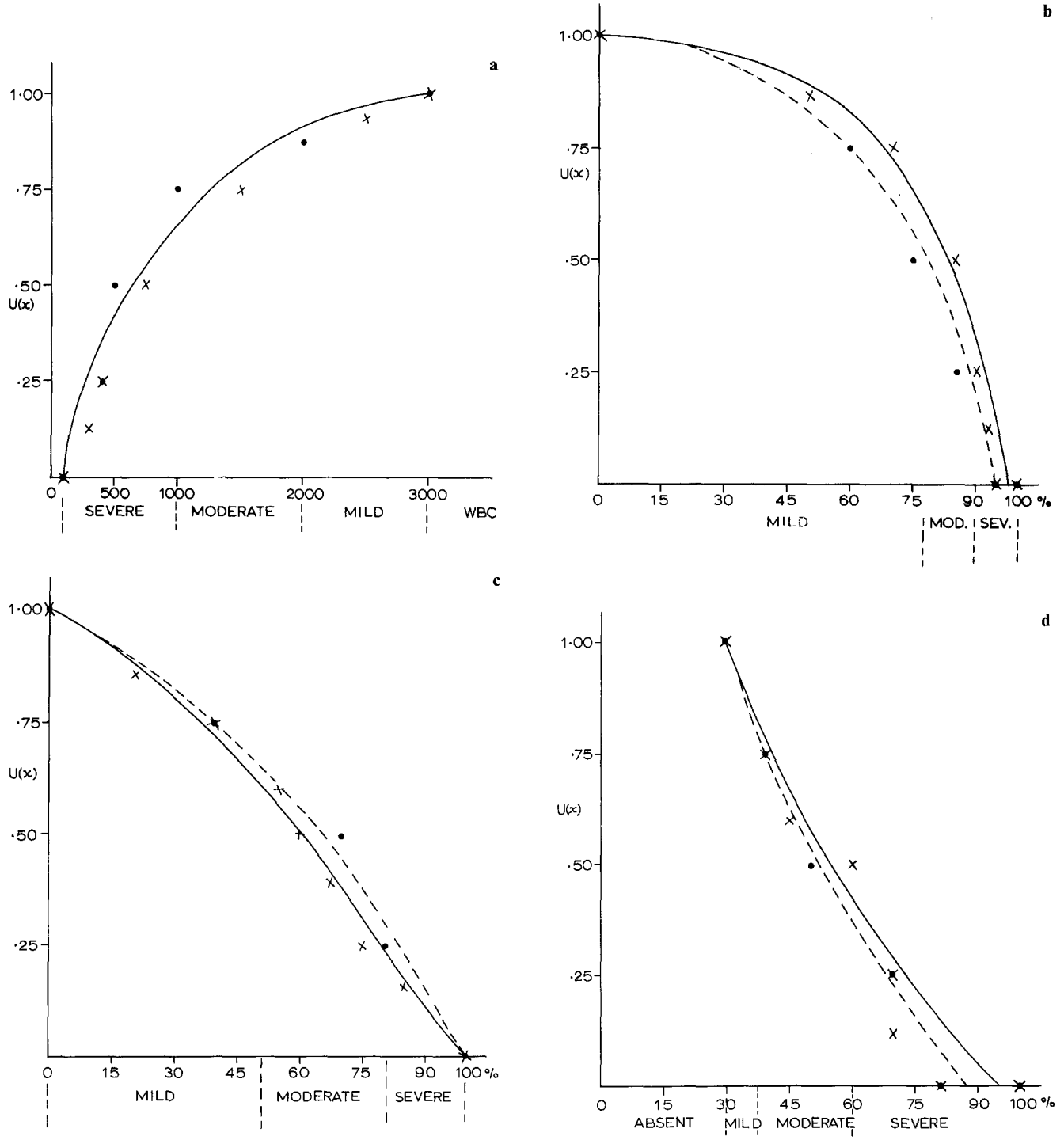
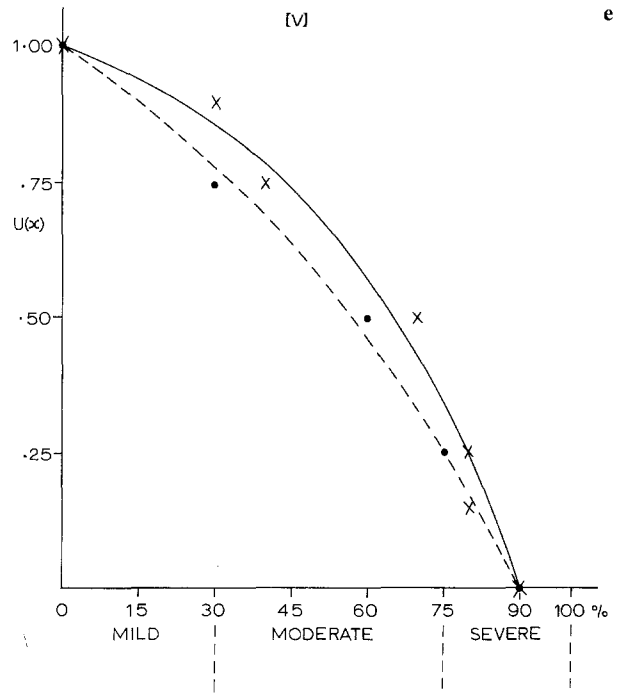


Fig. 2a–e. Utility curves. **a** Bone Marrow toxicity x-axis = white blood cell count (original and repeat drawn as one curve). **b** PNS toxicity x-axis = % PNS toxicity. **c** Auditory toxicity x-axis = % high-tone hearing loss. **d** Renal toxicity x-axis = % reduction in glomerular filtration rate from pre-trial level. **e** GIT toxicity x-axis = % GIT toxicity. x, original analysis; ●, repeated analysis



a way which takes account of the importance the clinician attaches to each of them.

Results of trial evaluation

The evaluation of the trial began with a survival analysis which has been reported more extensively by Wiltshaw [3]. The survival curve shown in Figure 3 indicates no significant difference between patients receiving JM8 or cisplatin as their first drug. This finding, which was not unexpected, placed more importance on a comparison of the toxicities of the two treatments.

For this trial, toxicity was abstracted from case-notes and recorded in one of only four categories (absent, mild, moderate, severe). The point scores used for the level of toxicity experienced by patients in a particular category are shown in Appendix C. A more sensitive analysis of utility than that described below could be performed if more precise toxicity data were used.

Fifty-eight patients received JM8 and 59 received cisplatin as their first treatment (Fig. 1). Patients in each arm received variable total amounts of drug because of the trial's quasi-crossover protocol. Table 1 shows the total numbers of courses received by patients before switching or discontinuing treatment.

In the cisplatin group, fewer patients received a large number of courses (greater than seven), either because of lack of response or because they experienced toxic effects.

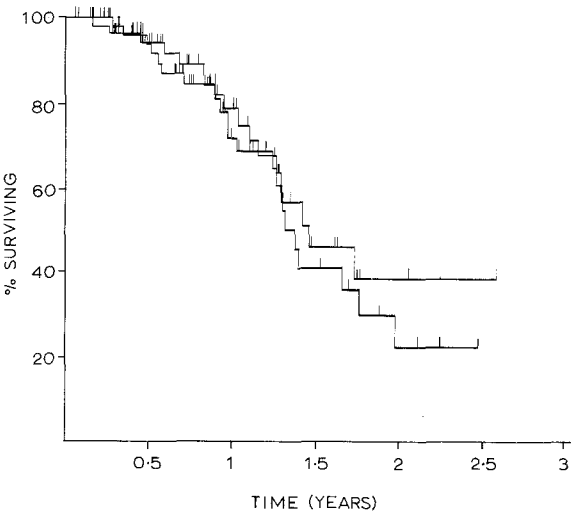


Fig. 3. Overall survival of patients receiving JM8 as first drug vs cisplatin as first drug. *Upper curve,* cisplatin (n = 59); *lower curve,* JM8 (n = 58). Log rank test results: X^2 = 0.19, df = 1, P = 0.66

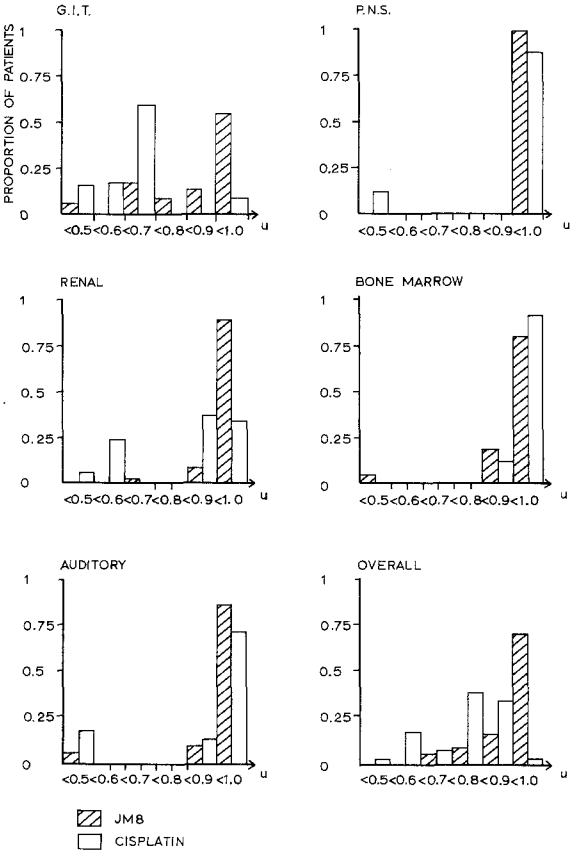


Fig. 4. The u score distributions for the five different sites and overall

Table 1. Numbers of patients receiving different number of courses of first-line drug

No. of courses	1	2	3	4	5	6	7	8	9	10	Un-known	Total
Cisplatin	5	14	7	7	8	11	3	0	3	1	0	59
JM8	8	7	5	9	10	0	0	4	6	5	4	58

Figure 4 shows the distributions of u scores in each site, and overall, for the two groups of patients. These were obtained by reading off from the curves of Fig. 2 the utility assessments corresponding to the toxicity levels recorded after the *last* course (of the first drug) each patient received. The overall U score was computed by means of the multiplicative model using the individual u scores. Mean scores for the two groups are shown in Table 2.

It can be seen that despite receiving fewer courses on average, patients given cisplatin experienced significantly more severe (less acceptable) toxicity in the cases of RENAL, PNS and GIT ($P < 0.05$; Wilcoxon rank sum test, two-sided). It was also found that several patients experienced more extreme ($u < 0.50$) AUDITORY toxicity (10 patients compared with three for JM8). Overall, the cisplatin patients had a significantly lower mean U score than the JM8 group ($P < 0.05$; Wilcoxon rank sum test, two-sided).

Of the 58 patients for whom JM8 was the first drug used, 18 were switched to receive cisplatin; 38 of 59 crossed over in the other direction. The individual overall U scores for these two groups of patients, after the last course received of both first-line and second-line drug, are listed in Appendix D.

Of the 18 patients switching to cisplatin, only two had severe overall U scores (arbitrarily defined as values less than or equal to 0.70) just before the change of treatment; both were improved after the one or two courses of cisplatin they received. Of the 38 patients switching from cisplatin to JM8, 11 had U scores of less than 0.70 at the time, and all but one of these showed improvement to more acceptable levels after the one to eight courses of JM8 they subsequently received. The mean score for these patients increased from 0.56 to 0.93.

For patients with more acceptable levels of overall toxicity ($U > 0.70$) prior to switching from cisplatin to JM8, U scores showed a significant improvement after crossover ($P < 0.01$; Wilcoxon matched pair test, two-sided). However, for the group of patients switching from JM8 to cisplatin, a significant decrease in U score was found ($P < 0.004$; Wilcoxon matched pair test, two-sided).

Complete response rates for the two drugs given as first treatment were statistically indistinguishable (Z test on proportions, two-sided; 38% for JM8 and 44% for cisplatin), and only two (of 18) and three (of 38) respectively responded *after* changing to the other drug.

Discussion

A means of combining the effects of toxicity experienced in different sites of the body into one single assessment has been presented. The resulting composite U score has been used to compare data on patients being treated with cisplatin and/or JM8.

It was found that more patients in this trial experienced severe overall toxicity when given cisplatin first than when given JM8 first. However, most patients in both "severe reaction" groups (JM8 and cisplatin) fared better when switched to the other drug. In general, the overall toxicity of JM8 was found to be more acceptable to this clinician than that of cisplatin, and given the similarity of the survival and response data, this suggests that JM8 might be preferred for first line treatment.

The model which has been developed and described here is specific to the clinician concerned. If the process were to be repeated with another clinician, the results might well be different, depending on that clinician's own attitude to the effects of toxicity.

It has been assumed throughout that the clinician is the major decision-maker with regard to treatment, but that patient preferences will implicitly influence any decisions. The model could, however, be equally well applied to any other major decision-maker, including the patient, if appropriate.

Administering the questioning procedures leading to the utility curves and the composite U score is a time-consuming process. In all, a total of 8 h of the clinician's time was required. This included a period during which the doctor familiarised herself with the type of question fundamental to the method, as well as several sessions for constructing the curves, for investigating the level of dependence between toxicities in different sites and for testing the reproducibility of the results.

The data used in the given example came from a trial which allowed patients to change treatment at times not specified in the protocol. Although this was considered appropriate for the patient group being treated, it does not provide an ideal context in which to test the method (there being many well documented [2] problems associated with the analysis of such data). It does, however, serve to demonstrate how the model may be used: both in providing a quantitative way of comparing overall toxicity between groups of patients, and in establishing a basis for monitoring individuals.

It is hoped that models of this kind, which are designed to reflect the clinician's assessment of the effects of toxicity, will eventually provide a useful means of summarising and reporting this important aspect of clinical trials.

In this study, no significant difference was found between the two treatment arms in terms of either survival or tumour response. Generally, however, such differences will be found. The extension of the model to allow toxicity and survival data to be considered together in a fuller general evaluatiion of treatments is another area for future research.

Appendix A

Constructing toxicity-utility curves for individual sites

The u scores for the separate toxicity sites were obtained as follows (each area of toxicity being examined as if it were the only one).

Table 2. Mean utility scores

First drug	Renal	GIT	Bone marrow	PNS	Auditory	Overall
Cisplatin	0.788	0.587	0.965	0.904	0.899	0.720
JM8	0.989	0.789	0.979	0.996	0.924	0.910

For example, BONE MARROW: A white blood cell (WBC) count of 3000 was regarded by the clinician as acceptably normal. The u values corresponding to 3000 and 0 were therefore set to 1 and 0 respectively, where u represents the clinician's assessment of risk and discomfort to the patient.

The following questioning procedure was employed to elicit values of u corresponding to intermediate levels of toxicity:

"Suppose 100 patients are to be treated with either drug A or drug B. If they are treated with A, 50 patients will have a WBC count of 3000 and 50 patients a count of 0. If they are treated with drug B, which has the same therapeutic effect as A, then all 100 patients will have the following toxicity:

WBC = x

What value of x would make it impossible for you to choose between A and B (i.e. would make you regard them as equivalent)?"

In this case the reply is $x = 500$.

Because the two drugs are regarded as equivalent at this point, we can say:

$$0.5\,(u(3000)) + 0.5\,(u(0) = l(u(500))$$
$$0.5\,(l) \quad + 0.5\,(0) \quad = u(500)$$

Thus the point at which u = 0.5 corresponds to a WBC of 500.

To find the point at which u = 0.75, a similar question was posed to the decision-maker, where drug A resulted in 50 patients having a WBC count of 3000 and 50 having a WBC count of 500. The clinician was then asked to provide the level of *bone marrow* toxicity (for all 100 patients) on drug B which would make the two drugs seem equivalent.

u(1000) = 0.75.

By repeating the same procedure the following points on the u scale were found:

u(400) = 0.25

u(2000) = 0.875.

A curve fitted by eye to these points is shown in Fig. 2a.

Appendix B

The aggregate utility score U

The way in which the individual u scores (corresponding to each of the five sites) are combined is dictated by the level of independence that exists between them. To determine this, a technique similar to that shown in Appendix A was used to discover the extent to which the clinician's attitude to toxicity in one site is affected by the level of toxicity in another.

In this case, sufficient dependence was found to render an additive model (in which overall utility can be regarded as the weighted sum of individual site utilities) inappropriate, and so a multiplicative model [1] was used, the general form of which is:

$$U = \left[\frac{1}{k} \left(\prod_{i=1}^{5} (1 + k k_i u_i) - 1 \right) \right]$$

where u_i is the utility in the i-th site, and k, k_i (i = 1, ... 5) are constant numbers. The magnitudes of the k_i values indicate the relative importance to the clinician of the five different sites of toxicity. These were estimated in this study using another similar questioning procedure [1] comparing particular combinations of toxicities in different sites, and the results are shown below:

	k_i
GIT	0.05
PNS	0.15
RENAL	0.10
BONE MARROW	0.05
AUDITORY	0.05

This indicates that PNS is regarded as the site where toxicity will be regarded with most concern. The factor k is a normalising factor (to fix U in the range 0–1) and was found by a simple iterative procedure described by Kenney and Raiffa [1] to take the value 6.40.

Appendix C: Toxicity levels used to represent categories in this study

	Bone marrow	Renal	Auditory	PNS	GIT
Toxicity category	White blood cell [0–3000]	% Reduction of EDTA clearance [0–100]	% High-tone hearing loss [0–100]	% Paralysis [0–100]	% Frequency of vomiting/ diarrhoea [0–100]
Absent	3000	30	0	10	0
Mild	2500	35	25	50	15
Moderate	1500	50	65	85	50
Severe	500	80	90	95	85

Appendix D: Patients crossed over to second drug

Cisplatin as first drug U (x)		JM8 as first drug U (x)	
After last course of cisplatin	After last course of JM8	After last course of JM8	After last course of cisplatin
0.73	0.76	0.98	0.77
0.70	0.67	0.94	0.83
0.71	0.34	0.75	0.90
0.81	0.98	0.94	0.79
0.72	0.97	0.94	0.75
0.88	0.95	0.70	0.83
0.61	0.93	0.94	0.90
0.70	0.94	0.98	0.55
0.94	0.98	0.94	0.70
0.81	0.97	0.97	0.80
0.60	0.97	0.94	0.80
0.83	0.94	0.93	0.77
0.82	0.91	0.90	0.73
0.86	0.98	0.65	0.83
0.86	0.98	0.97	0.98
0.73	0.79	0.93	0.48
0.79	0.79	0.98	0.86
0.75	0.79	0.97	0.90
0.74	0.60		
0.90	0.58		
0.83	0.98		
0.52	0.97		
0.76	0.97		
0.72	0.67		
0.75	0.95		

(continued)

Appendix D (continued)

Cisplatin as first drug U (x)		JM8 as first drug U (x)	
After last course of cisplatin	After last course of JM8	After last course of JM8	After last course of cisplatin
0.74	0.72		
0.71	0.95		
0.58	0.74		
0.58	0.94		
0.56	0.91		
0.34	0.67		
0.80	0.94		
0.80	0.76		
0.71	0.86		
0.58	0.90		
0.61	0.76		
0.58	0.90		
0.71	0.94		

References

1. Keeney RL, Raiffa H (1976) Decisions with multiple objectives: preferences and value trade-offs. Wiley, New York
2. Pocock SJ (1983) Clinical trials. A practical approach. Wiley, New York
3. Wiltshaw E (1985) Ovarian trials at the Royal Marsden Hospital. Cancer Treat Rev 12 (Suppl A) 67–71

Received November 13, 1986/Accepted July 13, 1987