Clinical and Epidemiologic Research

Devising Two-Stage and Multistage Phase II Studies on Systemic Adjuvant Therapy for Uveal Melanoma

John Whitehead,[1] *Svetlana Tishkovskaya,*[2] *Jemma O'Connor,*[3] *and Bertil Damato*[4]

Purpose. Almost all uveal melanomas showing chromosome 3 loss (i.e., monosomy 3) are fatal. Randomized clinical trials are therefore needed to evaluate various systemic adjuvant therapies. Conventional trial designs require large numbers of patients, which are difficult to achieve in a rare disease. The aim of this study was to use existing data to estimate how sample size and study duration could be reduced by selecting high-risk patients and adopting multistage trial designs.

Methods. We identified 217 patients with a monosomy 3 melanoma exceeding 15 mm in basal diameter; these patients had a median survival of 3.27 years. Several trial designs comparing overall survival were explored for such a population. A power of 0.90 to detect a hazard ratio of 0.737 was set, and recruitment of 16 patients per month was assumed.

Results. A suitable single-stage study would require 960 patients and a duration of 76 months. A two-stage design with an interim analysis based on 852 patients after 53.3 months would have a 50% probability of stopping because no statistically significant treatment effect is seen. Encouraging but inconclusive results would require a further 108 patients and prolongation of the study to 77.2 months. A multistage design would have a 43% probability of stopping before 47 months having recruited 759 patients.

Conclusions. Prospects for clinical studies of systemic adjuvant therapy for uveal melanoma are enhanced by multistage trial designs enrolling only high-risk patients. (*Invest Ophthalmol Vis Sci.* 2012;53:4986–4989) DOI:10.1167/iovs.12-9858

Approximately 50% of patients with choroidal melanoma develop metastatic disease.[1] Despite systemic treatment, such disease is usually fatal within a year of becoming symptomatic. With some other cancers, survival is improved by systemic adjuvant therapy directed at undetectable micrometastases in high-risk patients. Previous clinical trials in uveal melanoma have not shown statistically significant benefit of adjuvant therapy, but they had inadequate sample sizes and included patients with low risk of metastatic disease.[2,3]

From the [1]Medical and Pharmaceutical Statistics Research Unit, Department of Mathematics and Statistics, Lancaster University, Lancaster, United Kingdom; [2]School of Health, University of Central Lancashire, Preston, United Kingdom; [3]Department of Infection and Population Health, UCL Royal Free Hospital, London, United Kingdom; and [4]Ocular Oncology Service, Royal Liverpool University Hospital, Liverpool, United Kingdom.

Submitted for publication March 14, 2012; revised June 21, 2012; accepted June 24, 2012.

Disclosure: **J. Whitehead**, None; **S. Tishkovskaya**, None; **J. O'Connor**, None; **B. Damato**, None

Corresponding author: John Whitehead, Medical and Pharmaceutical Statistics Research Unit, Department of Mathematics and Statistics, Lancaster University, Lancaster LA1 4YJ, UK; j.whitehead@lancaster.ac.uk.

It is difficult to enroll sufficient patients with high-risk uveal melanoma within a reasonable time frame into adjuvant therapy trials because the incidence of uveal melanoma is only six per million per year and because only about 50% of such tumors have metastatic potential.[4] Many patients will not enroll because they live far from the center or are too elderly. Many are lost to follow-up because involvement in the study is too onerous or because they die of unrelated disease. The feasibility of such trials could be enhanced by excluding patients with only a small risk of metastatic death, which is now possible through genetic typing of uveal melanomas.[5] Metastatic disease from choroidal melanoma occurs almost exclusively in patients whose primary tumor shows chromosome 3 loss or class 2 gene expression profile (i.e., metastasizing uveal melanoma). Survival time in these patients correlates inversely with clinical stage of disease and with histological grade of malignancy.[5]

Randomized studies conventionally follow a single-stage design: the sample size is fixed before recruitment begins and a single analysis is performed once all the data have been collected. Such studies require large numbers of patients in order to have sufficient statistical power to detect therapeutic benefit. In lethal diseases such as cancer, many studies have one or more interim analyses so that the trial can be stopped early, either because efficacy has already been established or because it would be futile to continue. Our approach is a sequential study design in which the trial is conducted in stages, with a succession of interim analyses determining whether the study should continue. At each interim analysis, a test statistic is calculated reflecting the survival advantage of treatment relative to controls. At the ith interim analysis, there will be an upper limit for continuation, u_i, and a lower limit, ℓ_i. If the test statistic exceeds u_i, the trial is stopped because the treatment is beneficial; if it is less than ℓ_i, the trial is stopped for futility. Otherwise the trial continues to the next interim analysis. A variety of sequential methodologies have been developed, and here we adopt the "boundaries approach."[6] The designs presented here have been constructed so that their expected sample sizes are appreciably smaller than both the corresponding fixed sample size and the expected sample sizes of other two-stage and multistage approaches.

The aim of this study was to determine how sequential methods could reduce the sample size required for randomized trials of systemic adjuvant therapy of high-risk uveal melanoma.

Methods

Analysis of an Existing Dataset on Survival in Uveal Cancer

High-risk patients with uveal melanoma from the computerized database of the Liverpool Ocular Oncology Service were selected if: (1) the patient was diagnosed clinically or histologically with uveal melanoma; (2) the patient resided in mainland Britain and was thus

flagged at the National Health Service Cancer Registry, which automatically notified us of the date and cause of any deaths; (3) the tumor involved the choroid; (4) genetic studies showed the tumor to have chromosome 3 loss; and (5) the largest basal tumor diameter exceeded 15 mm. Patients were excluded if they had bilateral uveal melanoma, the genetic tumor type was not identified, or the basal tumor diameter was not recorded.

The subset consisted of 217 patients (122 male, 95 female) with a median age of 65.6 years (range 28.4–89.4). The right eye was affected in 116 (53.5%) patients and the left eye in 101 (46.5%). The tumor was considered to have arisen in the choroid in 206 (94.9%) of patients and in the ciliary body in 11 (5.1%), with ciliary body involvement in 140 (64.5%) patients and extraocular extension in 49 (22.6%). The largest basal tumor diameter had a median of 18.0 mm (range 15.1–23.6 mm) and the thickness had a median of 8.9 mm (range 1.5–17.7 mm). The TNM size category (7th edition) was T1 in 0 tumors, T2 in 7 (3.2%), T3 in 90 (41.5%), and T4 in 120 (55.3%). The ocular treatment consisted of enucleation (190 patients), proton beam radiotherapy (12 patients), trans-scleral local resection (8 patients), and brachytherapy (7 patients). A total of 128 (59.0%) patients died, the cause of death being diagnosed as metastatic melanoma in 118 (54.4%). This subset comprised 5.3% of 4076 British patients with a unilateral uveal melanoma and 26.6% of 817 such patients whose chromosome 3 status was known.

Figure 1 shows Kaplan-Meier estimates of the corresponding survival curve, with a median survival time of 3.27 years.

Assumptions and Targets

We assumed that eligible patients will be recruited and randomized in equal numbers between a novel treatment and a placebo control, with each treatment group receiving the same background care. Up to k interim analyses will be conducted. We set $\ell_k = u_k$, so that the trial is certain to be conclusive and stop at the last analysis (if not before). The stopping limits were calculated to achieve fixed risks of type I and type II errors. We took the Kaplan-Meier curves shown in Figure 1 as representative of survival in the controls. The designs were constructed with a power of 90% to detect a probability of surviving beyond 3.27 years being 0.60 in the treated group (compared with 0.50 for placebo) as superior to control at the 5% (two-sided) level if the hazards of death are proportional over time. The advantage defined here corresponds to a hazard ratio of 0.737.

All analyses (interim and final) will be based on the standardized form of the log-rank statistic, which is calculated and compared with prespecified critical values. In practice, all analyses might allow for baseline prognostic factors using proportional hazards regression analysis.[7]

The number of deaths required to provide sufficient information from a single-stage design can be found using a conventional approach[8] that is easily extended to two-stage and multistage designs. The sample size and study duration required to produce these deaths were determined by the method described in Appendix 2 of Whitehead.[9] We assumed that 16 patients per month are recruited for 5 years and followed up until sufficient deaths are observed.

Results

A Single-Stage Study Design

To satisfy the 90% power requirement, 452 deaths are required. An additional follow-up of 16 months is needed beyond the recruitment period of 5 years. A total of $16 \times 60 = 960$ patients would be recruited. At the time of analysis, the standardized log-rank statistic would be computed and the novel treatment considered significantly superior to control if it exceeded 1.96.

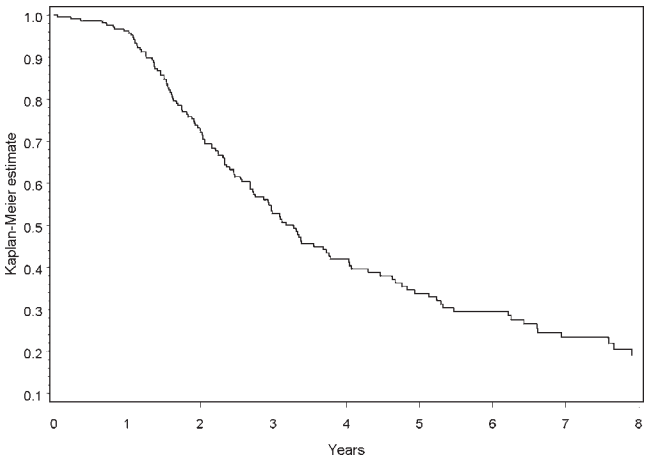

Figure 1. Kaplan-Meier estimates of survival function.

A Two-Stage Study Design

The two-stage design involves an interim analysis based on the standardized log-rank statistic calculated from all data available to date. If it is negative, the trial is stopped for futility. If it exceeds some predefined critical value u_1, then the trial is stopped to declare the novel treatment significantly superior to control. Otherwise, the trial is continued until a second and final analysis. At the final analysis, the log-rank statistic is computed from all available data, and if it exceeds a second critical value u_2, it is concluded that the novel treatment is superior. The design is constructed so that the number of deaths at the interim analysis would be one half the number at the final analysis (should the trial continue to that stage).

The design is determined by the values of u_1, u_2, and the number of deaths at the interim analysis. The required type I error rate (0.05, two-sided) and power (0.90) provide two equations that must be satisfied.

Several two-stage designs satisfy these requirements. We chose the one most likely to stop early. This design minimizes the expected required number of deaths for a hazard ratio of 0.858, which is the square root of the reference hazard ratio of 0.737. If the hazard ratio is 0.737 or 1 then early stopping is likely to declare a beneficial effect or futility, respectively. If the hazard ratio is 0.858 then early stopping is unlikely.

The optimal two-stage design involves an interim analysis after 53.3 months, by which time it is predicted that 852 patients would have been recruited and 232 deaths observed. The final analysis would take place after 77.2 months, by which time all 960 patients would have been recruited and a total of 463 deaths have occurred. The second analysis requires just 1.2 additional months of follow-up compared with the single-stage design and is based on 11 more deaths. The critical values are $u_1 = 2.54$ and $u_2 = 2.01$. Both u_1 and u_2 are more stringent than the value 1.96 used in the single-stage design. Table 1 presents further properties of the design.

A Multistage Study Design

Consider a design with 10 stages, involving an interim analysis after every 60 deaths, at the ith of which the standardized log-rank statistic Z_i would be computed and compared with the upper limit, u_i, and the lower limit, ℓ_i, as described above. This primary interim analysis requires: patient code number, recruitment date, date of death or date of last known status, and treatment arm.

The trial would recruit 16 patients per month for 5 years, and follow them up until a stopping boundary is reached. The

Table 1. Probability of Stopping at the Interim Analysis and Expected Terminal Sample Size and Trial Duration for the Two-Stage Design

Hazard Ratio	Probability of Stopping at Interim	Expected Final Sample Size	Expected Final Duration (Months)
1	0.506	905	65.1
0.858	0.207	938	72.2
0.737	0.425	914	67.1

study follows the triangular design,[6,10] an asymmetric design which stops for futility if no evidence that the novel treatment is superior is apparent. Following,[10] it can be established that the boundary points are given by the following equations:

$$\ell_i = -(3.591/\sqrt{i})\{1 - 3(i/10)\} \text{ and}$$
$$u_i = (3.591/\sqrt{i})\{1 + (i/10)\}, i = 1, \ldots, 10.$$

These boundary points are listed in Table 2, together with the number of deaths (d_i) that determine the timing of the ith interim analysis. Also shown are the probabilities of stopping on the upper boundary at or before each interim analysis, for the null hazard ratio of 1, the alternative hazard ratio of 0.737, and the intermediate value of 0.858. Corresponding probabilities for the lower boundary are also shown. A plot of $\ell_i\sqrt{(15i)}$ and $u_i\sqrt{(15i)}$ against the number of deaths (not shown) reveals the triangular shape that gives the test its name.

Table 3 shows the number of patients recruited at each interim analysis and the number of months that will have elapsed by then. Additional patients will be recruited while the interim analysis is being conducted and any decision to stop is being considered and confirmed. Stopping would have to occur by the third or fourth interim analysis to reduce sample size. Table 4 presents the expected sample size at study termination and the expected duration of the study in months (neglecting patients recruited and time elapsed during the conduct of the interim analysis and subsequent discussions about stopping). Reductions in sample size and study duration would be substantial if either the null or alternative values of hazard ratio were true, but less marked for the intermediate value. For all possible hazard ratios, these expected values improve on both the single-stage and two-stage designs. To achieve these savings, there must be a commitment to continue to the maximum trial duration of 97 months, should interim analyses prove inconclusive.

Figure 2 plots the expected sample sizes, for all three designs considered here against the value of $\theta = -\ln(\text{hazard ratio})$ (a scale that produces a symmetric plot). Note that $-\ln(1) = 0$; $-\ln(0.737) = 0.30$; and $-\ln(0.858) = 0.15$. The maximum sample sizes occur when the hazard ratio is slightly smaller than 0.858. Longer recruitment might allow the sample size to be reduced because a greater proportion of patients would die during the trial.

Table 3. Sample Size and Number of Months Elapsed at Each Interim Analysis

Interim	Sample Size at Interim	No. of Months Until Interim
1	491	31
2	640	40
3	759	47
4	868	54
5	960	61
6	960	67
7	960	73
8	960	79
9	960	87
10	960	97

Discussion

We have developed trial designs with 0.90 power to produce significant evidence that a novel systemic adjuvant therapy for high-risk uveal melanoma is superior to control at the 5% (two-sided) level, if the hazards of death in the two treatment groups are proportional over time with a ratio of 0.737. Such a hazard ratio implies that patients in the treated arm will have a 60% chance of surviving to the time that 50% of patients in the control arm are known to survive to.

At a realistic recruitment rate of 16 patients per month, a single-stage study design would require a trial enrolling 960 patients and lasting at least 6.5 years. A two-stage design would require 852 patients and would probably be 2 years shorter, because the treatment is either shown to be futile or beneficial. A multistage design, such as the triangular test, may require only 759 patients, shortening the trial even more, to 47 months, if clear results are available early on.

A randomized study of systemic adjuvant therapy in uveal melanoma would be a major undertaking. It would therefore be essential for it to be persuasive and definitive. Diminishing the power of a trial from 90% to 80%, would reduce the number of required deaths from 452 to 338, hence reducing

Table 2. Stopping Boundaries and Crossing Probabilities for a Triangular Test

				Probability of Stopping at ith Interim or Before on the Upper Boundary for Hazard Ratio			Probability of Stopping at ith Interim or Before on the Lower Boundary for Hazard Ratio		
Interim	d_i	ℓ_i	u_i	1	0.858	0.737	1	0.858	0.737
1	60	−2.514	3.950	0.000	0.000	0.003	0.006	0.001	0.0001
2	120	−1.016	3.047	0.001	0.014	0.085	0.155	0.032	0.004
3	180	−0.207	2.695	0.004	0.051	0.267	0.427	0.115	0.013
4	240	0.359	2.514	0.008	0.104	0.461	0.657	0.223	0.028
5	300	0.803	2.409	0.012	0.163	0.622	0.808	0.334	0.044
6	360	1.173	2.346	0.016	0.222	0.740	0.895	0.435	0.060
7	420	1.493	2.307	0.020	0.277	0.821	0.942	0.521	0.075
8	480	1.777	2.285	0.023	0.322	0.870	0.964	0.586	0.088
9	540	2.035	2.274	0.024	0.351	0.893	0.973	0.625	0.097
10	600	2.271	2.271	0.025	0.362	0.900	0.975	0.638	0.100

Table 4. Expected Terminal Sample Size and Trial Duration for the Triangular Design

Hazard Ratio	Expected Final Sample Size	Expected Final Duration (Months)
1	832	54.2
0.858	906	64.4
0.737	874	59.2

the sample size and study duration. However, most investigators would consider a 1 in 5 chance of an inconclusive result to be excessive.

The main strength of our study is the large number of high-risk patients we used to predict the survival pattern in the control arm of a future trial. Another strength is that we determined sample sizes using robust methods that were not based on parametric assumptions about the distribution of survival times. The inclusion of interim analyses can avoid continuing a study beyond the stage at which it becomes clear, either that the treatment is beneficial, or that it is futile. Omitting such analysis could lead to a study that is wasteful and unethical. With sequential trial designs, more trials can be performed more quickly, thereby increasing the likelihood of identifying an effective treatment.

Our analyses are based on all-cause mortality, an objective clinical measure usually acceptable to drug regulators and persuasive to clinicians. It does not rely on certified cause of death and accommodates any fatal iatrogenic complications. Our estimated sample sizes do not take account of loss to follow-up, which is likely to be common if many patients are elderly or live far from the research center. However, an intention-to-treat analysis is appropriate for such a trial, so for a patient to be included it would be necessary to know only the date of death or whether the patient was alive at a specified date.

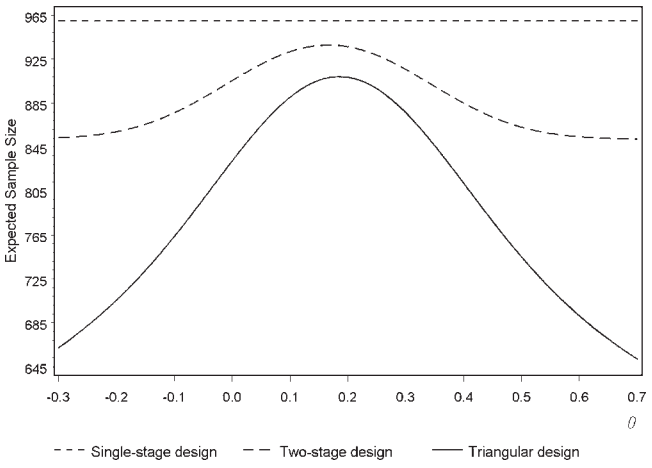

Figure 2. Expected terminal sample size.

Any studies on systemic adjuvant therapy would need to take account of patient compliance, which will depend on factors such as the drug being evaluated, its side effects and complications, the duration of treatment, the method of delivery (e.g., oral versus intravenous), and the frequency with which hospital visits are required. Another obstacle may be the reluctance or inability of some centers to perform prognostic biopsies. However, our impression is that a growing number of centers are offering prognostic biopsy and that this is rapidly becoming the standard of care.

There is no escaping the fact that a randomized trial evaluating systemic adjuvant therapy would require the enrollment of many patients and hence the participation of many centers. We hope that this study will facilitate the planning of such multicenter investigations, also preventing wasteful, inconclusive studies from being initiated.

Acknowledgments

During the preparation of this work, J O'C was supported by a studentship from GlaxoSmithKline and ST by funding from the Higher Education Innovation Fund.

References

1. Kujala E, Makitie T, Kivela T. Very long-term prognosis of patients with malignant uveal melanoma. *Invest Ophthalmol Vis Sci*. 2003;44:4651–4659.
2. Desjardins L, Dorval T, Lévy C, et al. Randomised study on adjuvant therapy by DTIC in choroidal melanoma. *Ophtalmologie*. 1988;12:168–73.
3. Voelter V, Schalenbourg A, Pampallona S, et al. Adjuvant intra-arterial hepatic fotemustine for high-risk uveal melanoma patients. *Melanoma Res*. 2008;18:220–224.
4. Damato B, Dopierala JA, Coupland SE. Genotypic profiling of 452 choroidal melanomas with multiplex ligation-dependent probe amplification. *Clin Cancer Res*. 2010;16:6083–6092.
5. Damato B, Eleuteri A, Taktak AF, Coupland SE. Estimating prognosis for survival after treatment of choroidal melanoma. *Prog Retin Eye Res*. 2011;30:285–295.
6. Whitehead J. *The Design and Analysis of Sequential Clinical Trials*. Revised 2nd ed. Chichester, UK: Wiley; 1997.
7. Cox DR. Regression models and life-tables. *J R Stat Soc B*. 1972;34:187–202.
8. Collett D. *Modelling Survival Data in Medical Research*. 2nd ed. Boca Raton, FL: Chapman and Hall/CRC; 2003.
9. Whitehead J. Predicting the duration of sequential survival studies. *Drug Info J*. 2001;35:1387–1400.
10. Whitehead J. Group sequential trials revisited: simple implementation using SAS. *Stat Methods Med Res*. 2011;20:636–656.