A Quality-of-Life Study in High-Risk (Thickness $\geqslant$ 2 mm) Cutaneous Melanoma Patients in a Randomized Trial of 1-cm versus 3-cm Surgical Excision Margins

Julia A. Newton-Bishop, Clarissa Nolan, Faye Turner, Margaret McCabe, Candida Boxer,* J. Meirion Thomas,† Gill Coombes,‡ Roger P. A'Hern,† and Jennifer H. Barrett

Division of Genetic Epidemiology, Cancer Research UK Clinical Center in Leeds, St. James's University Hospital, Beckett Street, Leeds; *Frimley Park Hospital, Camberley; †Royal Marsden NHS Trust and ‡Institute of Cancer Research, Sutton, UK

A quality-of-life study was carried out to test the hypothesis that melanoma patients treated with a 3-cm margin of excision suffer greater impairment of their quality of life than those treated with a 1-cm margin. The secondary aim was to determine the predictors of a poor patient perception of their excision scar. A postal questionnaire study was carried out using Hospital Anxiety and Depression (HAD), Psychosocial Adjustment of Illness Scale–Self-Report (PAIS-SR), Medical Outcomes Survey–Short Form 36 (MOS-SF36), and the Cassileth Scar questionnaires. Data were collected from 426 of the 537 patients who were mailed the questionnaires (response rate 79%). Fourteen percent had clinically significant anxiety and 5% had significant depression. A poor attitude toward quality of health care was associated with youth. Patients treated with a 3-cm margin excision had significantly poorer mental and physical function 1 mo after surgery, which disappeared within 6 mo. The greater difficulties experienced by the 3-cm margin group were particularly in their domestic, sexual, and social roles. Women, younger patients, those with poor physical and mental function after surgery, and those treated by a 3-cm margin were more likely to report a poorer perception of their scar. The poorer scar perception of patients in the 3-cm group persisted throughout the study period. Use of a 3-cm margin of excision for melanoma is associated with significantly more morbidity than use of a 1-cm margin, but this effect disappears in 6 mo. Patients treated by 3-cm excision were more likely, however, to have a persistent poor view of their scar. Youth and being female were also predictors of poor perception of the scar.

Key words: scar/HAD/PAIS-SR/MOS-SF36
J Investig Dermatol Symp Proc 9:152–159, 2004

Cutaneous melanoma presenting with a Breslow thickness of 2 mm or thicker is associated with a significant risk of recurrence, locoregionally or in distant viscera (Balch *et al*, 2001a). The treatment is surgical, but there are as yet no proven adjuvant therapies. The margin of excision chosen has changed over the years. An early pathologic study produced evidence of micrometastases in the skin around a tumor (Sampson, 1907) and led to the use of wide excision margins in an attempt to control a form of cancer, which at times appeared capricious. The use of wide margins with excision of fat down to fascia became the norm, and margins of many inches necessitated the use of grafts.

In more recent years a view has evolved that these large excision margins might be unnecessary. Nevertheless, although there have been several randomized clinical trials for thinner melanomas, providing reasonable evidence that, in these patients, margin has no effect on survival (Veronesi and Cascinelli, 1991), there are few data on outcome for thicker tumors. A study published by Balch *et al* compared 2-cm margins with 4-cm margins (Balch *et al*, 1995, 2001a, b), but only 213 patients in this trial had melanomas of 2 mm or thicker. A randomized clinical trial comparing 1-cm margins with 3-cm margins in melanomas 2 mm or thicker was therefore designed. These same margins were previously compared by the WHO Melanoma Group for melanomas thinner than 2-mm Breslow thickness (Veronesi *et al*, 1988; Veronesi and Cascinelli, 1991). In this WHO trial, margin of excision had no effect on disease-free survival or overall survival. Five patients in the 1-cm-margin group, however, had a local or in transit recurrence compared to one in the 3-cm-margin group. This difference was not statistically significant. The hypothesis to be tested in our MSG/BAPS UK Melanoma Excision Trial was that margin of excision has no effect on disease-free survival or overall survival, but that there may be an increased risk of local recurrence in the 1-cm-margin group. This MSG/BAPS UK Melanoma Excision Trial has shown that use of a 1-cm margin of excision for thick melanomas is associated with a higher risk of locoregional recurrence (particularly of regional lymph node recurrence) (M. Thomas *et al*, NEJH

Abbreviations: HAD, Hospital Anxiety and Depression; MCS, mental component summary; MOS-SF36, Medical Outcomes Survey–Short Form 36; PAIS-SR, Psychosocial Adjustment of Illness Scale–Self-Report; PCS, physical component summary; QOL, quality of life

in press, Feb. 2004). Although there is currently no significant evidence that there is any effect on survival, until the data have been allowed to mature, these data suggest that use of a 1-cm excision margin is untenable. These data are in press elsewhere (M. Thomas *et al*, NEJM Feb. 2004), and the quality-of-life (QOL) component of the study is reported here.

The treatment of melanoma is inevitably associated with a scar, but the nature of the scar depends very much on the margin of excision. A 1-cm margin can usually be carried out under local anesthetic as an outpatient and can be closed by primary intention. A 3-cm margin is much more likely to require a graft as an inpatient, with a cosmetic result surgeons would view as less satisfactory. Our hypothesis was therefore that use of a 3-cm margin of excision would be associated with a worse effect on QOL. It was predicted that this effect on QOL would be measurable in terms of the effect on physical measures and emotional measures. It was also predicted that the perception of the cosmetic appearance of the scar would be poorer in the 3-cm-margin group.

A postal QOL study was performed to test this hypothesis in a sample of the patients who were randomly assigned to the MSG/BAPS UK Melanoma Excision Trial.

Results

A total of 672 questionnaires were sent to 291 prospective subjects over a period of 2 y after surgery, and data were collected from 240 of these subjects. Data were not collected from 50 subjects as a result of patient refusal and from 1 subject as a result of consultant refusal.

A total of 246 questionnaires (i.e., 1 per patient) were sent to the retrospective patients, and data were collected from 186 of these patients. Fifty-nine patients declined to participate, and 1 patient was excluded because of consultant refusal.

There was significantly more nonresponse from retrospective than prospective participants over all time points, with 24% (60/246) and 15% (101/672) of questionnaires not returned, respectively (chi-square test, $p = 0.001$). Altogether 288 patients were measured only once, 186 of whom were in the retrospective group (Table I). Overall, the age of responders was similar to that of those who refused to take part and there were slightly more women in this group (Table II).

Of the 757 questionnaires returned, there were 82 missing PCS and MCS scores, 27 missing scar scores, and between 28 and 252 missing observations over all PAIS domain scores. The missing PAIS domain scores related mainly to vocational and sexual function. That is that when patients were retired they indicated as such on the form and left the section blank. Patients who were not sexually active generally speaking indicated as such on the form and left the section blank. Three patients had no scar site data and 98 patients did not have a scar score measured at 6 mo or after.

The mean age of the study population was 62.2 y and 57% of the study population were men. Significantly more women than men had an exposed scar, 48% (89/185) and 18% (43/241), respectively (chi-square test, $p < 0.0001$). This was expected, because melanoma is more common on the legs in women.

Fourteen percent (61/426) had scores consistent with a significant degree of anxiety and an additional 18.8% (80/426) of patients had a HAD score suggestive of borderline anxiety. Five percent (20/426) of patients had scores consistent with a significant degree of depression at one time point or more and 7.7% (33/426) were possibly depressed at one time point or more. This result concurs with evidence that, in cancer patients, clinical anxiety is more often present than depression (Skarstein *et al*, 2000).

Correlation between the different QOL measures was assessed as an internal validation of these measures. The weakest pairwise correlation was between PCS and MCS (correlation coefficient, 0.05; $p = 0.17$). This was expected because PCS and MCS were constructed to make the distinction between physical and mental health outcomes (Ware *et al*, 1994).

The following seven scores appear to be particularly intercorrelated: HAD Depression, HAD Anxiety, Psychological Distress (PAIS), MCS, Vocational Role (PAIS), Domestic Role (PAIS), and Social Role (PAIS). The strongest correlation of all the variables was between HAD Anxiety and PAIS Psychological Distress (coefficient, 0.75; $p < 0.0001$). HAD Anxiety was also highly correlated with MCS (coefficient, -0.62; $p < 0.0001$), a result that has been demonstrated previously (Fossa and Dahl 2002). Psychological Distress was most strongly correlated with MCS (coefficient, -0.66; $p < 0.0001$) and HAD Depression was most strongly correlated with Psychological Distress (coefficient, 0.65; $p < 0.0001$). HAD Depression was also strongly correlated with Social Role (coefficient, 0.64; $p < 0.0001$). Within the

Table I. The number of questionnaires returned at each time point by treatment group and the number of nonresponders at each time point

	<1 mo	1 mo	3 mo	6 mo	1 y	2 y	Retrospective	Total
1-cm margin	6	73	55	64	54	35	94	381
3-cm margin	11	64	56	67	52	34	92	376
Responders	17	137	111	131	106	69	186	757
Nonresponders (% of total)	0 (0)	30 (18)	15 (12)	19 (13)	23 (18)	14 (17)	60 (24)	161
Total number sent questionnaires at each time point	17	167	126	150	129	83	246	918

Table II. Number of responders and nonresponders by age, sex, and margin

	Margin (cm)			
	Responders (n = 426)		Nonresponders (n = 111)	
	1	3	1	3
N	216	210	53	58
Mean age, years (SD)	61.2 (14.0)	62.1 (14.4)	60.4 (15.2)	58.5 (18.3)
% men	59.3	53.8	45.3	39.7

PAIS domains there were strong correlations between Domestic Role and Vocational Role (coefficient, 0.69; $p < 0.0001$), and between Domestic Role and Social Role (coefficient, 0.69; $p < 0.0001$).

PCS and MCS The PCS and MCS scales computed using the MOS-SF36 as overall measures of physical and mental well-being correlated well with other measure of QOL as described above. Both the PCS and the MCS were significantly lower at 1 mo among the wide margin group than the narrow margin group (Mann–Whitney U tests, $p = 0.03$ and $p = 0.008$, respectively). The mean PCS and MCS were somewhat lower among the wide margin group overall, but these results were not statistically significant at the 5% level ($p = 0.06$ and $p = 0.15$, respectively). A low PCS score was associated with patients of older ages ($p < 0.0001$) and a low MCS score was associated with patients of younger ages ($p = 0.001$). The mean PCS and MCS did not differ with sex or with respect to whether the scar was exposed or not.

To determine whether the effect of time on the PCS and MCS differed between margin groups, random effects models were used incorporating time as a continuum in years and an interaction term between margin and time (Table III). These models showed that the PCS scores improved at a significantly faster rate for the wide margin group compared to the narrow margin group ($p = 0.005$), the wide margin group having worse scores initially. The pattern is similar but less marked for the MCS scores (the interaction is not significant, $p = 0.11$).

To illustrate this, Figs 2 and 3 describe the change in the mean PCS and MCS score for each margin group over time. There was a large improvement in PCS score (Fig 1) among the wide margin group between 3 and 6 mo, and from 6 mo onward there appears to be little difference between the treatment groups (Mann–Whitney U test for difference at 2 y, $p = 0.85$). The narrow and wide margin groups never quite reached the UK general sample PCS mean throughout the 2 y. Figure 3 shows the lower mean MCS score among the wide margin group at 1 mo and an obvious improvement in MCS score for the wide margin group between 1 and 6 mo. Like the PCS score there was little difference in MCS score between the treatment groups from 6 mo onward (Mann–Whitney U test for difference at 2 y, $p = 0.63$). The narrow margin group had a higher mean MCS than the UK general sample throughout the 2 y, but the wide margin group reached the UK general sample mean after only 3 mo. At one time point or more: 8% of subjects had an MCS less than 2 SD below the UK mean and 11% had a PCS less than 2 SD below the UK mean (in each case significantly more than 2.5%).

Paired t tests were carried out between the first and last scores for patients measured at least twice. Subjects were included if they had a first score in the first 3 mo and a last score at 6 mo or later. There were no significant differences, among the narrow margin group, between the first and last mean PCS scores ($p = 0.15$; mean difference, −1.88; 95% CI, −4.47 to 0.71) or between the first and last mean MCS scores ($p = 0.98$; mean difference, −0.03; 95% CI, −2.19 to 2.14). Nevertheless, among the wide margin group, the first mean PCS score was significantly lower than the final mean PCS score ($p < 0.0001$; mean difference, −7.44; 95% CI, −10.64 to −4.24). Likewise, the first mean MCS score

Table III. Random effects models for PCS and MCS including time[a]

	PCS (model 1)		MCS (model 2)	
Covariates	**Coefficient (SE)**	**p value**	**Coefficient (SE)**	**p value**
Margin[b]	−157.0 (83.5)	0.06	−133.1 (91.6)	0.15
12 mo[c]	−29.3 (82.7)	0.72	−67.7 (87.3)	0.44
6 mo[c]	−21.5 (77.6)	0.78	−71.7 (82.0)	0.38
3 mo[c]	−297.2 (83.9)	<0.0001	−52.3 (89.0)	0.56
1 mo[c]	−559.2 (80.3)	<0.0001	−241.6 (85.3)	0.005
<1 mo[c]	−617.8 (187.4)	0.001	−235.0 (197.6)	0.23
Sex[d]	−58.2 (90.4)	0.52	61.5 (99.1)	0.54
Age	−11.1 (2.94)	<0.0001	11.2 (3.2)	0.001
Exposed[e]	−66.6 (97.1)	0.49	−55.7 (106.5)	0.60

[a]Number of observations, 673; number of subjects, 392. PCS and MCS are squared; a lower score indicates a worse outcome.
[b]3-cm margin compared with the baseline of 1-cm margin.
[c]The baseline of the time variable is 2 y.
[d]Women compared with the baseline of men.
[e]Exposed scar compared with the baseline of not exposed.

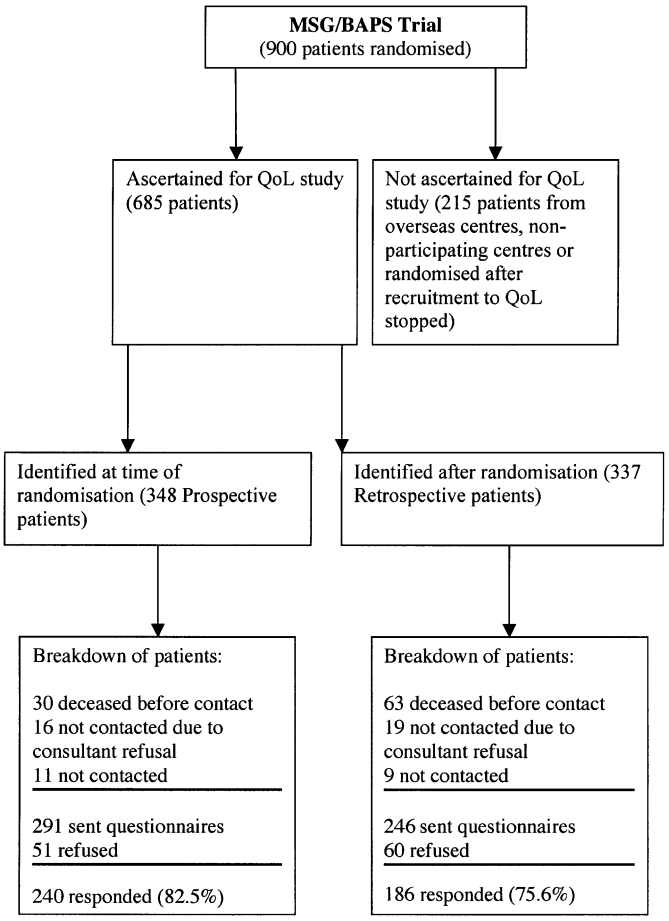

Figure 1
MSG/BAPS trial.

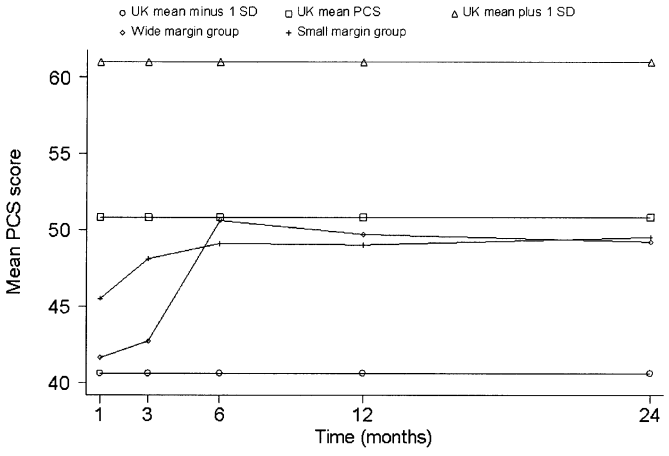

Figure 2
Change in mean PCS over time, by margin group.

was significantly lower than the final mean MCS score among the wide margin group ($p = 0.02$; mean difference, −2.63; 95% CI, −4.74 to −0.52).

PAIS domains

Health-Care Orientation (PAIS domain 1) and Psychological Distress (PAIS domain 7) Mann–Whitney U tests at 1 mo showed no differences between the narrow and the wide margin groups with respect to psychological distress or attitude toward the quality of medical care, treatment, and their illness (as measured using PAIS Health-Care Orientation). A poor attitude toward the quality of medical care was associated with younger age overall ($p < 0.0001$), but there was no effect of sex or whether the scar was exposed. Overall, the wide and narrow margin groups did not differ in their attitude toward the quality of medical care, to treatment and to their illness ($p = 0.4$). Overall, there was no evidence of a significant change in these attitudes over time and no difference in change of attitudes between treatment groups over time (interaction term, $p = 0.88$).

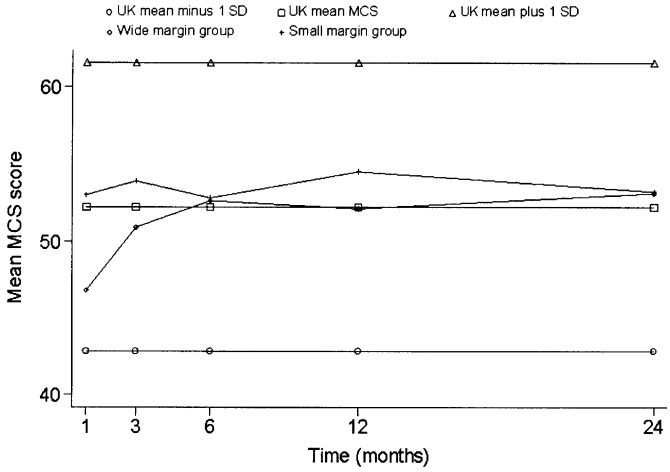

Figure 3
Change in mean MCS over time, by margin group. Scores measured at less than 1 mo were not incorporated into the graphical displays because of small numbers.

Results for the Psychological Distress domain (PAIS 7) were in all respects very similar to those for the MCS score, with which this domain is highly correlated.

PAIS domain 2 through PAIS domain 6 Chi-square tests between the binary PAIS domains and margin size at each time point revealed the following results: significantly more problems among the wide margin group in domestic role and sexual role at 1 mo (chi-square (1), 7.17; $p = 0.007$; and chi-squared (1), 6.35; $p = 0.01$, respectively); significantly more problems among the wide margin group in social role at 1 and 3 mo (chi-square (1), 7.21; $p = 0.007$; and chi-square (1), 5.52; $p = 0.02$, respectively). There were no significant differences in vocational role and extended family relations between treatment groups at any time point.

Results of the generalized estimating equation models are shown in Table IV. Overall, there were significantly reduced odds of having poor adjustment to domestic role over time in the narrow margin group (OR of 0.43 for each year that passes) but not of the other domains. The chance of improvement in domestic role over time was significantly higher among the wide margin group than the narrow margin group (OR, $0.43 \times 0.51 = 0.22$, test for interaction; $p = 0.05$). Thus patients with a wide margin were more likely to find their domestic role a problem, but this was more likely to improve. There were significantly reduced odds of poor adjustment to social role over time among the wide

Table IV. Generalized estimating equation models for PAIS domain 2 through PAIS domain 6

	Domain[a]				
Parameter	**Vocational**	**Domestic**	**Sexual**	**Extended family**	**Social**
No. of observations	504	652	568	713	706
Margin[b]	1.66 (0.68–4.08)	**3.11 (1.17–8.27)**	1.92 (0.70–5.31)	1.09 (0.43–2.75)	**4.22 (1.54–11.55)**
Time in years	0.69 (0.44–1.06)	**0.43 (0.27–0.68)**	1.35 (0.84–2.15)	0.81 (0.52–1.27)	0.78 (0.51–1.21)
Margin × Time	0.67 (0.36–1.24)	**0.51 (0.26–1.00)**	0.60 (0.30–1.19)	0.93 (0.49–1.77)	**0.48 (0.25–0.92)**
Sex[c]	0.66 (0.36–1.22)	1.52 (0.75–3.08)	0.57 (0.26–1.27)	0.68 (0.32–1.42)	0.67 (0.33–1.39)
Age	0.98 (0.96–1.00)[d]	0.98 (0.95–1.00)[d]	1.02 (0.99–1.04)	1.00 (0.98–1.03)	1.00 (0.97–1.02)
Exposed[d]	1.82 (0.94–3.53)	1.21 (0.57–2.54)	1.43 (0.61–3.32)	1.04 (0.48–2.26)	**2.55 (1.17–5.56)**

OR (95% CI). Odds ratios relate to risk of a poor outcome. Results significant at the 5% level are bolded.
[a]3-cm margin compared with the baseline of 1-cm margin.
[b]Women compared with the baseline of males.
[c]Exposed scar compared with the baseline of not exposed.
[d]p values are 0.09 for vocational role and 0.06 for domestic role.

margin group only (OR, 0.37, interaction term; $p = 0.03$), and social role was two to three times more likely to be adversely affected in those with a scar on an exposed site.

HAD Depression and Anxiety scores Results for both of these scores were similar to those for the MCS score, with which they are highly correlated, except that the depression score showed no relationship with age.

Cassileth Scar score Women were found to have a poorer perception of their scar overall than males ($p < 0.0001$) (a result that has been demonstrated previously (Cassileth *et al*, 1983)), as did patients who were younger ($p < 0.0001$) or who had an exposed scar ($p = 0.001$). The wide margin group had a significantly poorer perception of their scar than the narrow margin group, and unlike the other QOL measures, this difference persisted over the whole period of follow-up. The scar scores were 19% lower in the wide margin overall ($p < 0.0001$; 95% CI, 15%–23%), 22% lower at 1 mo ($p < 0.0001$; 95% CI, 13%–30%) and 16% at 2 y ($p < 0.0001$; 95% CI, 10%–21%). Overall, these problems reduced significantly between 1 and 6 mo and 2 y, and the wide margin group improved at a significantly faster rate than the narrow margin group (interaction term, $p = 0.02$). Figure 4 illustrates this pattern over time. There was a significant correlation between the scar score and the HAD Depression score at all time points from 1 mo, ranging from 0.45 at 1 mo to 0.33 at 2 y (Pearson correlations). A poor scar score was therefore held to be clinically significant.

The predictors of a poor scar score between 6 mo and 2 y after surgery Table V shows the effect of initial PCS and MCS scores on whether or not a patient had a poor scar score at their final assessment (6 mo to 2 y after surgery) in those measured at least twice. The PCS and MCS scores were classified into approximate quartiles based on all time points. Patients in the lowest quarter of PCS scores initially were 4.5 times more likely to have a poor scar score than patients in the highest quarter ($p = 0.04$). Patients with a first MCS score less than the median were between three and four times more likely to have a poor scar score than

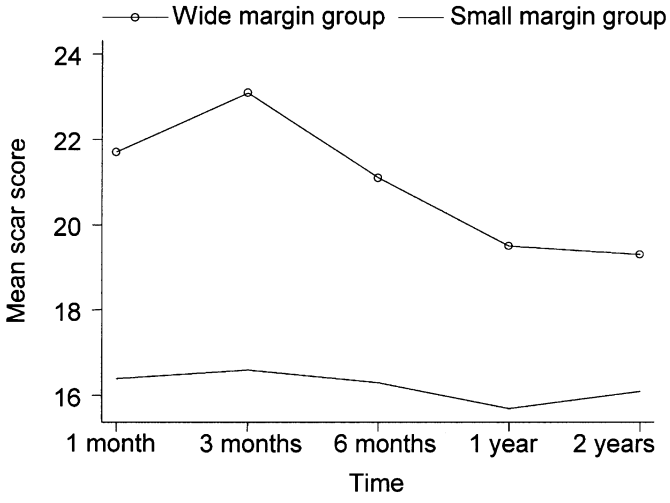

Figure 4
Change in mean scar score with time in the two treatment groups.

patients in the baseline category. This suggests that increased physical and mental limitations at the first time point were associated with an increased risk of a poor scar score at the last assessment (6 mo to 2 y after surgery). Similarly for psychological distress, extended family, and vocational role, subjects in the highest category (poorer adjustment to illness) were more likely to have a poor scar score than patients in the baseline category. There were no associations for the remaining PAIS domains or for the HAD scales.

Discussion

This was a postal questionnaire study of a sample of patients recruited to the first UK national surgical melanoma trial. Responses were received from 83% of patients sent questionnaires in the prospective study, and 76% of patients contacted from the retrospective study (79% overall). Of the 918 questionnaires sent out in total, 757

Table V. Logistic regression models for the last scar score, including PCS and MCS, in subjects measured at least twice: a multivariate model[a]

Last scar score	N	OR (95% CI)	p value
PCS			
1 (PCS > 55.4)	25	1	
2 (49.9 < PCS = 55.4)	31	1.37 (0.32–5.84)	0.67
3 (39 < PCS = 49.9)	33	2.30 (0.57–9.25)	0.24
4 (PCS = 39)	39	4.45 (1.11–17.93)	0.04
Sex[b] (male/female)	72/56	3.91 (1.40–10.94)	0.009
Age[c] (y)			
53	40	1	
53 < 69	47	0.30 (0.10–0.88)	0.03
> 69	41	0.05 (0.01–0.20)	< 0.0001
Exposed[c] (not exposed/exposed)	86/42	0.62 (0.20–1.92)	0.41
Margin[d] (1 cm/3 cm)	64/64	6.29 (2.28–17.38)	< 0.0001
MCS			
1 (MCS > 59.1)	32	1	
2 (55.3 < MCS = 59.1)	31	1.77 (0.41–7.58)	0.44
3 (46.9 < MCS = 55.3)	33	4.27 (1.01–18.07)	0.05
4 (MCS = 46.9)	32	3.28 (0.80–13.39)	0.11
Sex[b] (male/female)	72/56	4.00 (1.44–11.12)	0.008
Age (y)			
Baseline (age = 53)	40	1	
53 < 69	47	0.37 (0.13–1.09)	0.07
> 69	41	0.07 (0.02–0.27)	< 0.0001
Exposed[c] (not exposed/exposed)	86/42	0.97 (0.33–2.89)	0.96
Margin[d] (1 cm/3 cm)	64/64	5.55 (2.06–14.98)	0.001

[a]Number of subjects, 128.
[b]Women compared with the baseline of men.
[c]Exposed scar compared with the baseline of not exposed.
[d]3-cm margin compared with the baseline of 1-cm margin.

(82%) were returned. Only one-third of the participants (138/426) returned multiple questionnaires. Analyses performed on the data collected from patients who returned multiple questionnaires were consistent with the results from the whole data set.

The study used a series of well-described QOL measures, and good correlation was found between measures purporting to measure similar things. For example, there was strong correlation between HAD Anxiety, MCS, and Psychological Distress. We found the MOS-SF36 item questionnaire summary variables MCS and PCS useful in monitoring changes over time, and the PAIS domains were useful for looking at specific issues such as attitudes to the health-care services.

The HAD questionnaire showed significant levels of anxiety in the melanoma patients as was expected, but the MOS-SF36 item MCS score for mental health showed that the population of melanoma patients reached UK population means for mental function 3 mo after surgery. Overall then, the trauma of surgery for melanoma (and presumably the trauma of dealing with a diagnosis of cancer) was associated with measurable anxiety, but adjustment was made such that population means were reached in a comparatively short period of time. This adjustment took no account of relapse during the course of the study, which might have been expected to increase anxiety again. It is of note that lower MCS scores indicative of poorer mental health were associated with younger age at diagnosis.

It is of interest that a poorer attitude toward the quality of medical care was associated with younger age, and one could speculate whether this was a function of age or was a cohort effect reflecting changes in society generally. It did not differ according to margin of excision, sex, or whether or not the scar was exposed, which might suggest that the specific experiences of the patient were less important than their age.

The purpose of the study was to investigate the effect of margin of excision on QOL. The MSG/BAPS UK melanoma excision trial itself (M. Thomas *et al*, NEJM Feb. 2004,

in press) showed that use of a 3-cm margin of excision was associated with longer hospital stay, greater use of skin grafts, and higher surgical complication rates. Not surprisingly then, in this QOL study, use of a 3-cm margin was associated with significantly greater impairment of physical function, which took 3 to 6 mo to settle, and this impairment was greater in older individuals. The larger margin was associated, in addition, with a significant but less marked effect on mental function as measured using the MCS (SF-36 item). The PAIS domains allow the investigator to look at specific aspects of life. Use of the 3-cm margin was significantly more likely to have an adverse effect on the patient's function in the home, sexually and socially at 1 mo. But these functions (particularly at home and socially) improved with time.

Poorer perceptions of the surgical scar, as measured using the Cassileth Scar questionnaire were significantly more likely in the 3-cm margin group, even at 2 y after surgery. At 2 y , the degree of physical and mental health impairment associated with wider excision margins, as measured by mean questionnaire scores, settled. The implication is therefore that patients who have had wider excision margins view their scar as poorer at 2 y but the degree of physical and mental distress associated with that is not detectable with MOS-SF36 in the group overall. The correlation of poor scar scores with HAD Depression scores at all time points, however, suggests that poor perception of the scar was clinically relevant.

A poor perception of the scar was more likely in women and younger patients as was predicted, but whether or not the scar was exposed had no effect. Patients who had low PCS scores when first measured were more likely to perceive their scar to be poor. Patients whose graft does not take, or who have sepsis, are more likely to be physically impaired around the time of surgery. They are more likely also to have a poorer scar, which may explain this correlation. It is not possible for us to distinguish this from a general effect of depression, however, and indeed a low MCS score was also indicative of a poor perception of the scar.

This study showed significantly greater impairment of physical and mental components of QOL after 3-cm margins of excision rather than after 1-cm margins in melanoma patients. Nevertheless, with the exception of perception of their scar site, there were no persisting differences in these measures 2 y after surgery. Use of a wider excision margin was associated with a persistently poorer perception of the scar.

The excision trial has to date shown a greater locoregional relapse rate if a 1-cm margin is used albeit with not yet significantly increased mortality (M. Thomas *et al*, NEJM Feb. 2004, in press). The choice for the patient and his/her doctor therefore involves a balance between poorer scars with attendant, albeit transient greater morbidity, and an increased risk of a lymph node dissection if a 1-cm margin is used rather than a 3-cm margin.

Materials and Methods

Patients A total of 900 patients were recruited to the MSG/BAPS UK Melanoma Excision Trial (in press, NEJM). These patients were informed about the QOL study by surgeons randomizing to the margins trial, who sought permission for a postal consent process to be carried out by a study coordinator based in Leeds. The QOL component of the trial was funded for only 2 of the 8 y in which the excision trial was running. During this period, 685 English-speaking subjects were identified either at the time of randomization to the MSG/BAPS UK Melanoma Excision Trial or as having being randomized previously. Of the 685 subjects, 348 were identified at the time of randomization (the prospective group) and 337 were identified after randomization (the retrospective group). Of 348 prospective subjects 291 were asked by post to consent to completion of a series of questionnaires from diagnosis for a period of 2 y, and 246 of 337 retrospective subjects were asked to complete a single set of questionnaires 2 y or more after surgery. The main reasons for not posting consent forms and questionnaires to the remaining patients identified were patient death ($n = 93$) or consultant refusal ($n = 35$) (Fig 1).

The questionnaires chosen were the Hospital Anxiety and Depression (HAD), Psychosocial Adjustment of Illness Scale–Self-Report (PAIS-SR), Medical Outcomes Survey–Short Form 36 (MOS-SF36), and the Cassileth Scar score as described below. All four questionnaires were sent to patients at each time point. The original protocol stipulated that a set of questionnaires would be given to the patient by the randomizing surgeon at surgery and that the other five sets would be dispatched at 1 mo, 3 mo, 6 mo, 1 y, and 2 y after surgery. Nevertheless, few patients were given the questionnaires by the randomizing surgeon (17 in all) at time point 1. Losses of patients from the study occurring as a result of patient illness, death, or withdrawal from the study are indicated by falling numbers of questionnaires dispatched in Table I.

QOL measures

HAD scale The HAD scale was developed to screen for significant anxiety and depression in patients with physical illness, using two subscales (HAD Anxiety and HAD Depression). The HAD Anxiety and the HAD Depression scales each consist of seven items, scored from 0 (no problems) to 3 (maximum distress), resulting in total scores ranging from 0 to 21. It is suggested that scores between 8 and 10 indicate "possible cases" and scores of $11+$ indicate "cases" of anxiety and depression (Skarstein *et al*, 2000; Fossa and Dahl, 2002).

PAIS-SR The PAIS-SR questionnaire was designed to assess the quality of a patient's psychosocial adjustment to a current medical illness or its resultant effects. The PAIS-SR consists of seven domains, which have been identified as having high predictive relevance for adjustment to illness. These domains relate to health-care orientation (perception of services), vocational environment, domestic environment, sexual relationships, extended family relationships, social environment, and psychological distress. A total of 46 items encompass the Seven domains, each item being scored on a 4-point scale with alternated scale direction on the even-numbered items to reduce position response biases (Derogatis, 1983). A high PAIS domain score indicates poor adjustment.

MOS-SF36 The MOS-SF36 is a 36-item questionnaire assessing health function (Ware *et al*, 1994). It consists of eight scales: physical functioning, vitality, social functioning, general health, bodily pain, physical role, emotional role, and mental health (Ware *et al*, 1994). The eight MOS scales were aggregated, using the US scoring algorithm, to produce a physical component summary (PCS) scale and a mental component summary (MCS) scale. It has been shown that the correlations between summary scales using the UK-specific algorithms and those using the US algorithms are sufficiently high to justify using the US scoring algorithms, which are well documented and the most widely used (Ware *et al*, 1998). A low PCS score indicates substantial limitations in self-care, physical and social activities, severe bodily pain, and frequent tiredness. A low MCS score indicates frequent psychological

distress and substantial social disability owing to emotional problems (Ware *et al*, 1994).

Cassileth Scar score The Cassileth Scar score questionnaire was developed to investigate patients' opinions of the size and cosmetic implications of their excisions (Cassileth *et al*, 1983). The questionnaire contains 12 items, each scored from 1 (minimal negative cosmetic impact) to 4 (maximum negative cosmetic impact). The questionnaire also contains front and back outline drawings of the human body, on which patients were asked to indicate the size, shape, and location of their scar. From this, a scar site variable was calculated and grouped into the following four categories: trunk, upper arm/leg, head, or lower arm/leg. An additional variable was calculated based on whether or not a patient's scar was predicted to be usually exposed, and this was coded as follows: 0 = trunk or upper arm/leg; 1 = head or lower arm/leg.

Statistical methods Spearman's rank correlation coefficients were produced for all pairwise combinations of the untransformed MCS, PCS, PAIS, HAD, and scar score variables.

The frequency distributions of MCS and PCS were left skewed, so squaring transformations were used to normalize the data. Log transformations were used for the right-skewed frequency distributions of HAD, PAIS domain 1, PAIS domain 2, and the scar scores (adding one to the scores before log transformation). Mann–Whitney U tests were carried out to identify possible differences in these scores between the two margin groups at particular time points, and random effects models were used to examine the distribution of these scores over time for each margin group.

Graphs of the mean MCS, PCS, and scar scores at each time point were used to display changes over time. As a comparative measure, the graphs also display the corresponding means and standard deviations from data collected on a sample of 1760 UK subjects by means of a general population survey (also calculated using the US scoring algorithms) (Ware *et al*, 1998).

Two-thirds (288/426) of patients had only completed one questionnaire so graphs of the mean MCS and PCS scores at each time point were also plotted for the 138 patients who were measured at least twice. This was to ascertain that any improvements in the mean MCS or PCS scores over time were due to changes within patients rather than differences between patients. In addition, for patients measured at least twice, paired t tests were carried out between their first and last PCS scores and between their first and last MCS scores, for each margin group.

For PAIS domain 2 through 6, log transformations did little to normalize the data because at least one-third of all the observations were zero. On this basis, a binary variable was calculated for PAIS domain 2 to PAIS domain 6, and these variables were analyzed by means of generalized estimating equation logit models.

The random effects and generalized estimating equation models examined the effect of treatment margin and how this changed over time and were adjusted for age, sex, and whether a patient's scar was exposed or not. Models including interaction terms between margin and time were used to determine whether changes in scores over time differed between margin groups.

One aim was to identify whether a poor initial HAD, PAIS, MCS, or PCS score could predict a poor scar score at least 6 mo after surgery. For patients whose scar score was last measured at least 6 mo after surgery, their last scar score was recorded. A binary variable was then calculated and coded as zero for observations in the lower two-thirds of final scores and coded as 1 for observations in the upper one-third of final scores (poor scar). The first MCS, PCS, HAD, and PAIS scores were then identified for each patient, and logistic regression models were used to determine which of the initial measures predicted a poor scar score at least 6 mo after surgery. The analysis was carried out in those patients whose QOL was measured on at least two occasions. All logistic regression models adjusted for age, sex, whether exposed, and margin size. All analyses were carried out using Stata, version 7 (StataCorp, 2001).

This work was funded by the Northern and Yorkshire NHS Executive and Cancer Research UK. We thank Joanne Walker who assisted greatly with data management. We are grateful to participating surgeons and patients for their time and forbearance. We are also grateful to Professor Leroy Fallowfeld for advice in setting up the study.

DOI: 10.1046/j.1087-0024.2003.09118.x

Manuscript received May 14, 2003; revised July 29, 2003; accepted for publication August 20, 2003

Address correspondence to: Julia Newton-Bishop, Genetic Epidemiology Division, Cancer Research UK, Cancer Genetics Building, St. James's University Hospital, Beckett Street, Leeds, LS9 7TF, UK. Email: j.newton-bishop@cancer.org.uk

References

Balch CM, Soong SJ, Gershenwald JE, *et al*: Prognostic factors analysis of 17,600 melanoma patients: Validation of the American Joint Committee on Cancer melanoma staging system. J Clin Oncol 19:3622–3634, 2001a

Balch CM, Soong SJ, Smith T, *et al*: Long-term results of a prospective surgical trial comparing 2 cm vs. 4 cm excision margins for 740 patients with 1–4 mm melanomas. Ann Surg Oncol 8:101–108, 2001b

Balch C, Urist MM, Karakousis CP, *et al*: Efficacy of 2-cm surgical margins for intermediate thickness melanomas (1 to 4 mm): Results of a multi-institutional randomized surgical trial. Ann Surg 218:262–267, 1995

Cassileth BR, Lusk EJ, Tenaglia AN: Patients' perceptions of the cosmetic impact of melanoma resection. Plast Reconstr Surg 71:73–75, 1983

Derogatis L: The prevalence of psychiatric disorder among cancer patients. J Am Med Assoc 249:751–757, 1983

Fossa SD, Dahl AA: Short Form 36 and Hospital Anxiety and Depression Scale: A comparison based on patients with testicular cancer. J Psychosom Res 52:79–87, 2002

Sampson H: The pathology of melanotic growths in relation to their operative treatment. Lancet 32:362–367, 1907

Skarstein J, Aass N, Fossa SD, Skovlund E, Dahl AA: Anxiety and depression in cancer patients: Relation between the Hospital Anxiety and Depression Scale and the European Organization for Research and Treatment of Cancer Core Quality of Life Questionnaire. J Psychosom Res 49:27–34, 2000

StataCorp. Statistical Software: Release 7.0. College Station (TX): Stata Corporation, 2001

Veronesi U, Cascinelli N: Narrow excision (1-cm margin): A safe procedure for thin cutaneous melanoma. Arch Surg 126:438–441, 1991

Veronesi U, Cascinelli N, Adamus J, *et al*: Thin stage I primary cutaneous malignant melanoma: Comparison of excision with margins of 1 or 3 cm. N Engl J Med 318:1159–1162, 1988

Ware JE Jr, Gandek B, Kosinski M, *et al*: The equivalence of SF-36 summary health scores estimated using standard and country-specific algorithms in 10 countries: Results from the IQOLA Project. International Quality of Life Assessment. J Clin Epidemiol 51:1167–1170, 1998

Ware J, Kosinski M, Keller SD: SF-36 Physical and Mental Health Summary Scales: A User's Manual. Boston: Health Assessment Lab, 1994