Genetic Markers in Long-Term Survivors of Glioblastoma Multiforme

Jan Shaw

SID: 0438597

Project Title: Genetic Markers in long term survivors of Glioblastoma Multiforme

Location and dates: Brain Tumour Project, Cancer Genetics, the Kolling Institute, Royal North Shore Hospital, St Leonards, NSW, 2007

Context: The data used for this project was collected by my work supervisor Kerrie McDonald in association with surgeons working at Royal North Shore Hospital, North Shore Private Hospital and Prince of Wales Hospital in the Sydney Neuro-oncology Group. The immunohistochemistry on sections from archived paraffin blocks of tumour tissue obtained from glioblastoma patients at initial biopsy or tumour debulking was carried out by Kerrie and her Research Assistant, Cathy Payne and the data on patients was obtained by Kerrie or by Jonathan Parkinson, a trainee brain surgeon and PhD student in the laboratory, from primary patient records.

I was working on the brain tumour project as a molecular and cell biologist working on gene therapy and expression projects relating to Glioblastoma Multiforme (GBM) and also using skills gained through the Master of Biostatistics degree to analyse data generated from these projects. Judy Simpson supervised the statistical part of the project.

Contribution of the student: My contribution to the project was the data validation and analysis.

Statistical issues involved: One major statistical issue was the dataset to use. The original dataset included short term survivors who were diagnosed with GBMs less than 1000 days before the census date of 30th May 2007. Kerrie was of the opinion that as they were all already dead it was legitimate to include them as short term survivors, but Judy pointed out that long term survivors who were diagnosed during the same time period were not able to be included which could lead to bias since long term survivors who were

diagnosed after the 31st August 2004 were not given the equivalent opportunity to be included in the study and so were not available as a comparison group for short term survivors diagnosed this date. Because of this the short term survivors diagnosed after 31st August 2004 were removed. However this did reduce the number of patients available for the study from 50 (37 short term and 13 long term) to 41 (28 short term and 13 long term), which introduced the problem of interpreting results obtained with a smaller sample size. Another possible source of bias was the inclusion of the original three long term survivors, since the hypothesis about the association between long term survival and low expression in two genes was formed because it was present in 2 of the 3 original long term survivors, so including them could give a significant outcome when a sample of patients not including them might not. This has been addressed in the Appendix, which repeats the analyses without them. The choice of levels to use for the variables was another statistical issue that caused some difficulty. It was resolved by using the levels that gave a good linear trend of the coefficients in logistic regression and by using the levels that gave good separation of Kaplan Meier curves in survival analysis. Concerns about collinearity and confounding were also encountered.

Signed declaration by the student: I declare that the work presented here is my own work, with suggestions and changes from my two supervisors, Kerrie McDonald and Judy Simpson.

Signed:

Comments by the project supervisor:

Jan initially worked quite independently on this project with her content supervisor, as she has described above and in the preface. She then responded appropriately to my comments on each draft of the project report. She has worked diligently on this project, although there have been delays, mainly due to the fact that her funding ran out and she needed to find new employment. She now has a good new job with CSIRO.

Signed:

Judy M Simpson
Professor of Biostatistics
University of Sydney

Genetic markers in long term survivors of Glioblastoma Multiforme

1. Introduction

The most common type of primary brain tumour is glioblastoma multiforme or GBM, which accounts for around 38% of brain tumours. It is also the most aggressive and has a median survival time of 14.6 months even with intensive therapy and is invariably fatal. GBMs have several characteristics that make them difficult to treat. One is that their growth is invasive with diffuse infiltration into adjacent brain tissue. This is unlike many other tumour types that initially grow within a discrete capsule. This diffuse infiltration makes it impossible to remove the entire tumour surgically without removing a large area of healthy brain. Another characteristic is that tumour cells accumulate mutations and changes in gene expression over time that make chemotherapy and radiotherapy treatments less effective.

A subgroup of GBM patients, however, display significantly longer overall survival. This subgroup has been documented previously as "Long Term Survivors" (LTS) where patients show survival of 3 years or more after the initial diagnosis of a GBM. Scott et al (1999) in a population based case control study of LTS GBM patients matched for age, sex and year of diagnosis with short term survivors (STS), found that LTS patients had a higher Karnofsky Performance Score and were more likely to have had gross total resection and adjuvant chemotherapy and their tumours were more likely to have fewer mitoses than control patients. The incidence of LTS in the GBM population is less than

3%. Molecular markers that can predict which GBM patients will fall into the LTS category are currently unknown and need to be elucidated.

A recent microarray and qPCR study in the Cancer Genetics Laboratory at the Kolling Institute identified two marker genes, IQGAP and IGFBP2 as being significant in predicting the aggressiveness of glioma tumours (McDonald et al, 2007). In this study three GBM patients had survival times of over three years. Two of these three LTSs were found to have low expression of IQGAP and IGFBP2 as measured by immunohistochemistry. The third patient, with a survival time over 1000 days had high expression in these genes but had received unusually aggressive treatment at their own request and expense.

The expression of IGFBP2 (Insulin-like Growth Factor Binding Protein 2) is increased in a range of tumour types, with a positive correlation between tumour grade and the level of expression. However the mechanisms by which IGFBP2 is involved in tumorigenesis are not clear. Wang et al (2007) found that cell lines over-expressing IGFBP2 RNA showed significantly enhanced invasiveness and also had increased expression of matrix metaloproteinase-2 (MMP-2), which plays a role in tumour progression by degrading the extracellular matrix.

IQGAP (IQ-motif-containing GTPase activation protein 1) reduces the ability of cells to adhere to each other by binding to E-cadherin and preventing it from binding the catenins

and attaching to the actin cytoskeleton. Hence its increased expression is associated with reduced cell-cell adhesion which allows cells to migrate and metastasize.

Additional genes that have been found to be important in tumour progression are PTEN and MGMT. PTEN (phosphatase and tensin homolog) gene, when correctly expressed, acts as a tumour suppressor gene by regulating cell proliferation, promoting apoptosis and regulating cell migration and invasion. Mutations in the PTEN gene are responsible for Cowden Syndrome characterised by overgrowth of tissue. PTEN regulates cell proliferation by promoting cell-cycle arrest in the G_1 phase through inhibition of the PI3K/Akt pathway. PTEN sensitises glioma cells to apoptosis when irradiated or treated with chemotherapy drugs also through the PI3K/Akt pathway. PTEN expression reduces glioma cell invasion *in vitro* but the molecular mechanism for this is not yet fully understood. The PTEN gene is often mutated to an inactive form in glioma tumours. PTEN is also important in controlling angiogenesis during tumour formation again through the PI3K/Akt pathway by regulating HIF1α and VEGF expression (Park et al, 2002).

Hegi et al (2005) found that MGMT promoter methylation as shown by methylation specific PCR, was a favourable prognostic factor for longer survival, irrespective of treatment. The methylation of the MGMT promoter was also found to be associated with a more favourable response of patients to temozolomide and radiotherapy treatment. The MGMT gene encodes the enzyme O-6-methylguanine-DNA methyltransferase, a DNA repair enzyme which removes methyl groups from the O-6 position of guanine.

Alkylating chemotherapy drugs such as temozolamide, act by adding methyl groups to DNA which then targets the cell for death by apoptosis. So the MGMT enzyme can repair damage caused by chemotherapy drugs. Methylation of promoter regions of genes inhibits their expression and so the methylation of the MGMT promoter region would be expected to result in lower expression and better tumour response to alkylating drugs. However MGMT plays conflicting roles in cancer since the ability of MGMT to repair DNA damage is vital for cell survival. The silencing of MGMT results in an increased frequency of mutations which can lead to more aggressive tumour behaviour in mutated cells. As a result of the paper by Hegi showing that patients with high levels of methylation of the MGMT gene have a better response to the temozolomide, GBM patients with low MGMT promoter methylation are denied temozolomide treatment in some parts of the world.

Now a total of 13 LTSs have been identified where long term survival was defined as survival for more than 1000 days. This study compared immunohistochemistry expression data for IQGAP and IGFBP2 in these LTSs to patients with survival times of less than 1000 days (STSs) to see if the observation of lower IQGAP and IGFBP2 expression in LTSs was confirmed. We also looked at the expression of additional genes: MGMT and PTEN as assessed by immunohistochemistry. The study population were all consenting patients aged less than 65 years, diagnosed with GBM between January 1998 and 31st August 2004, treated at Royal North Shore Hospital or North Shore Private Hospital, or by participating clinicians at the Prince of Wales Private Hospital, who

received the standard treatment of tumour resection, if achievable, radiotherapy and temozolomide chemotherapy.

1.1. Aim

The aim of this study was to determine if the expression of a range of proteins that have been shown to be differentially expressed in glioblastoma tissue, can be used as an aid to predicting survival time in patients with GBMs. In particular, the aim was to determine whether long term survival, defined as survival for more than 1000 days, can be predicted by expression levels of any of the proteins: IGFBP2, IQGAP1 or PTEN in tumour tissue or by the % nuclei staining for MGMT expression.

1.2. Hypotheses

Low expression of IGFBP2 and IQGAP1 and high levels of expression of PTEN in samples of tumour tissue obtained at initial biopsy or surgery, as measured by immunohistochemistry, are predictive of survival for more than 1000 days in glioblastoma patients. The hypothesis relating to MGMT expression is more complicated as this gene has conflicting roles in tumour development. Because some GBM patients with unmethylated MGMT promoter regions are being denied temozolomide treatment, it is important to test the hypothesis behind this policy. This hypothesis is that low MGMT expression, which is the expected result of high levels of promoter methylation, is associated with a better prognosis in patients receiving the standard treatment of radiotherapy and chemotherapy with temozolamide.

2. Methods:

2.1. Patient Selection

Patients were selected from the Neuroendocrine Tumour Bank over a seven year period (1998-2004) and their survival outcomes determined from the Sydney Neuroendocrine database. Tumour samples and clinical data including patient age, gender, tumour location and treatment were obtained in Sydney under the auspices of Dr Ray Cook, a leading member of the Sydney Neuro-oncology Group (SNOG). All consenting patients diagnosed with a GBM during this interval and aged less than 65 years, were categorised into two groups on the basis of their survival time: short-term survivors (STS) with survival times of less than 1000 days after diagnosis and long-term survivors (LTS) with survival times greater than 1000 days. Age was restricted to less than 65 years as all LTSs were younger than this and as age over 65 is one of the most significant predictors of poor survival. The census date was 30th June 2007 as this was when the immunohistochemistry and scoring were completed and the analysis started. No patients diagnosed after 31st August 2004 were included in the study as they would not have had the opportunity to become a long-term survivor before the census date. 41 patients fulfilled these criteria of which 28 were STSs and 13 were LTSs. At the time of the census date, four LTSs were still alive.

An experienced neuropathologist, Dr Janice Brewer, from the Department of Anatomical Pathology at Royal North Shore Hospital, collected archived paraffin blocks of tumour tissue obtained from glioblastoma patients at the time of initial biopsy or surgery, i.e. very close to the time of initial diagnosis. The tumours were diagnosed as GBMs by Dr Brewer

by means of histological examination of sections cut from this embedded tissue according to the WHO 2000 criteria. Ethics approval for this study was covered by Protocol 0612-228M and was obtained from the Human Research Ethics Committees of participating institutions.

2.2. Immunohistochemistry Detection of Proteins in GBM samples

Paraffin sections (4μm) were cut from the blocks by the Department of Anatomical Pathology at RNSH and immunohistochemistry was used to detect the cytoplasmic protein expression of IGFBP2 and IQGAP1 and PTEN and the nuclear-specific protein MGMT. Trained technicians, blinded to the survival-time, assessed the degree of staining.

2.2.1. IGFBP2 Immunohistochemistry

Positive IGFBP2 staining is typically observed in the cytoplasm of the tumour samples since it is a secreted protein. Scores were allocated to represent all observed degrees of staining. Subjects were given scores between 0 and 3 for IGFBP2 staining, where 0 represented no cytoplasmic staining, 1 for weak cytoplasmic staining (<5% of examined tumour cells), 2 for moderate cytoplasmic staining (<25% of examined tumour cells) and 3 for strong membranous and cytoplasmic staining (>25% of examined tumour cells).

2.2.2. IQGAP1 Immunohistochemistry

Positive IQGAP1 staining was typically visualised as a strong cytoplasmic reaction. To score for IQGAP1 staining, subjects were given scores between 0 and 4 with where 0

represented no staining, 1 for weak cytoplasmic staining (<5% of examined tumour cells), 2 for moderate cytoplasmic staining (<20% of examined tumour cells), 3 for moderate to strong cytoplasmic staining (<25% of examined tumour cells) and 4 for strong cytoplasmic staining (>25% of examined tumour cells).

2.2.3. PTEN Immunohistochemistry

PTEN was also localised in the cytoplasm. The scoring for PTEN was more complex and staining was graded according to a previously established scale of 0-2 (Mellinghoff et al, 2005). A score of 2 was given when the cytoplasmic staining intensity was equal to that of the vascular endothelium, 1 if it was less intensely stained than the endothelium, and 0 if it was undetectable. Tumours with a score of 0 or 1 were then considered PTEN negative and a score of 2 was positive, giving a binary final score.

2.2.4. MGMT Immunohistochemistry

MGMT expression was quantified by microscopically examining 500–1000 tumour cells under high-power and counting cells with stained nuclei. The percent of stained cells was calculated by dividing the number with stained nuclei by the number of cells counted.

These IHC scoring methods were dictated by the appearance of the staining in the tissue being examined and was guided by previous practice in similar published studies.

2.3. Statistical Analyses

Initial exploratory analysis looked at the frequencies of the categorical variables and at the distribution of the continuous variable, age. The variable for percentage of nuclei staining for MGMT was divided into categories at the overall quartiles and treated as an ordered categorical variable because its distribution was highly skewed and because the most frequent result was zero. As the sample size was small, Fisher's exact test was used to test for significance differences between LTSs and STSs in categorical variables and the Mantel-Haenszel test which is also valid for small sample sizes was used to test for significant trends in the ordered categorical variables (Kay, 2007). Univariate logistic regression was carried out to assess the linear trend of the regression coefficients for each level of the ordered categorical variables since the logistic regression model used for multivariate modelling, assumes a linear relationship between the outcome and the risk factor. If there was not a linear trend, categories with similar coefficients were merged until a linear trend or a binary variable was achieved. The Mann Whitney test was used to compare the mean age of LTSs and STSs as age was found not to be normally distributed. The p-values for the Mantel-Haenszel test for trend were adjusted using a modified Bonferroni correction as the cutpoints used had been chosen to maximise trend (Lausen and Schumaker, 1996).

All statistical methods were carried out using the data set that included the three original LTSs and also on a second data set without these three patients as the presence of the original patients may have biased the analysis since two of the three were known to have low expression of IQGAP1 and IGFBP2 and including them could give a more significant

outcome than a sample of patients not including them. The results for this alternative analysis are given in the appendix.

Univariate logistic regression was used to find if the level of protein expression could be used to predict whether a subject would be a LTS using those genes with $P < 0.25$ in the exploratory analysis. For those genes that were predictive of long-term survival in this univariate analysis, the joint effect of these genes in predicting long-term survival was assessed using multivariate logistic regression. The influence of the presence of genes on the size of the odds ratio and significance of the other gene variables in the model was assessed to find evidence of confounding. A full model was fitted containing all gene variables with $P < 0.25$ in the univariate models. A parsimonious model was arrived at by eliminating the least significant gene variable and comparing the resulting model containing one less variable using the likelihood ratio test and the unweighted residual sum of squares (RSS) test (Kuss, 2002). This process was continued until eliminating a variable resulted in a χ^2 statistic from the likelihood ratio with a p-value less than 0.05 which indicated the model was significantly worse fitting.

Clinicians prefer to have a cut-point above which one prognosis is more likely than the alternative, rather than a continuous value with the likelihood of a disease outcome increasing as the value increases. Receiver Operating Characteristic (ROC) analysis helps identify the most appropriate classification rules and was used to examine whether gene expression information would be useful at predicting prognosis in patients diagnosed with GBM. ROC curve analysis was used to find cut-points for IHC scores that maximised the

sensitivity and specificity of predicting prognosis in the genes that were in the final model. These cut-points were compared to the sensitivity and specificity of the scores in their original form. IHC scores from predictive genes were combined and analysed with ROC curves to examine whether the combination was better at prediction that the genes separately.

Survival analysis was used to examine the relationship between gene expression as measured by IHC, and survival time as this could differ from its relationship with the binary outcome of survival for more than 1000 days. Survival analysis uses information from all time points from diagnosis until death while logistic regression only considers whether the subject survived for more than 1000 days. If the expression of a gene has an effect on survival within the first few months after diagnosis, but not later, this would not be evident from logistic regression. While long term survival is the primary focus of this study, it was thought that survival analysis might reveal more detailed effects of the expression of these genes on prognosis.

The outcome variable was survival time in days, counted from the patient admission date for first surgery or biopsy of the tumour. A second exploratory analysis used Kaplan Meier curves, incidence rates and logrank tests was carried out and categories with similar incident rates were merged for Cox regression analysis of the ordered categorical variables to ensure the ordered categorical variables were linear in the log hazard. Univariate Cox regression models were fitted for each gene variable with $P < 0.25$ in exploratory analysis. A multivariate Cox regression model was fitted using the variables in the univariate

models and the influence of the presence of each gene on the size of the hazard ratio and significance of the other gene variables in the model was assessed to find evidence of confounding. A parsimonious model was arrived at using a similar procedure to that for the logistic regression model, testing each model for the proportional hazards assumption.

All analyses were carried out using Stata 8.2.

3. Results

3.1. Exploratory analysis

The results of exploratory analysis are given in Table 1. 28 patients in the study survived less than 1000 days from their first surgery or biopsy and 13 survived more than 1000 days. Fifteen were female and 26 male with ages ranging from 26 to 62. Histograms of age for the full dataset with the original three LTSs are shown in Figure 1. In all cases, corresponding figures for the dataset without the three originals are given in the Appendix. Age was found not to be normally distributed by the Shapiro-Wilks test ($P = 0.002$), so a Mann-Whitney test was used to test for significant differences in age between the long term and STSs. No significant difference in age was found ($P = 0.11$).

Table 1: Comparison of long term and STSs by age, gender, IGFBP2, IQGAP1, PTEN and MGMT score.

		Frequency (%)			
		Short term	Long term	coeff by level *	Tests for significant effects and trend
Number	Levels	28	13		
Gender	Female	11 (73)	4 (27)	0	Fisher's exact P = 0.73
	Male	17 (65)	9 (35)	0.38	
IGFBP2 (original IHC scores)	0	6 (60)	4 (40)	0	Fisher's exact P < 0.0005 Mantel Haenszel test for trend P = 0.005
	1	1 (12)	7 (88)	2.35	
	2	6 (86)	1 (14)	-1.39	
	3	15 (94)	1 (6)	-2.30	
IGFBP2 (3 levels)	0&1	7 (39)	11 (61)	0	Fisher's exact P < 0.001 Mantel Haenszel test for trend P = 0.0006 (0.002)#
	2	6 (86)	1 (14)	-2.24	
	3	15 (94)	1 (6)	-3.16	
IQGAP1 (original IHC scores)	0	1 (50)	1 (50)	0	Fisher's exact = 0.18 Mantel Haenszel test for trend P = 0.10
	1	0 (0)	1 (100)	-	
	2	5 (50)	5 (50)	0	
	3	7 (87)	1 (13)	-1.95	
	4	15 (75)	5 (25)	-1.10	
IQGAP1 (3 levels)	0&1	1 (33)	2 (67)	0	Fisher's exact = 0.09 Mantel Haenszel test for trend P = 0.04 (0.09)#
	2	5 (50)	5 (50)	-0.69	
	3&4	22 (79)	6 (21)	-2.00	
MGMT % stained nuclei in quartiles	<7%	10 (100)	0 (0)	-	Fisher's exact = 0.013 Mantel Haenszel test for trend P = 0.003
	7-39%	7 (70)	3 (30)	-1.41	
	40-67%	7 (70)	3 (30)	-1.41	
	>67%	4 (36)	7 (64)	0	
MGMT (3 levels)	<7%	10 (100)	0 (0)	-	Fisher's exact = 0.005 Mantel Haenszel test for trend P = 0.002 (0.008)#
	7-67%	14 (70)	6 (30)	17.5	
	>67%	4 (36)	7 (64)	19.0	
PTEN IHC score	0	8 (53)	7 (47)	0	Fisher's exact P = 0.17
	1	20 (77)	6 (23)	-1.07	

		Short term	Long term	
Survival time	Lower quartile	255	1126	
	Median	389	1192	
	Upper quartile	752.5	1477	
Age at 1st Surgery	Mean	49.1	45.8	Mann Whitney test P = 0.11
	Stand dev	9.4	8.5	

* Coeff by level: the logistic regression coefficient at each level of IHC score calculated with the scores either in their original or final form.

adjusted p-value using modified Bonferroni correction in groups merged to maximise trend.

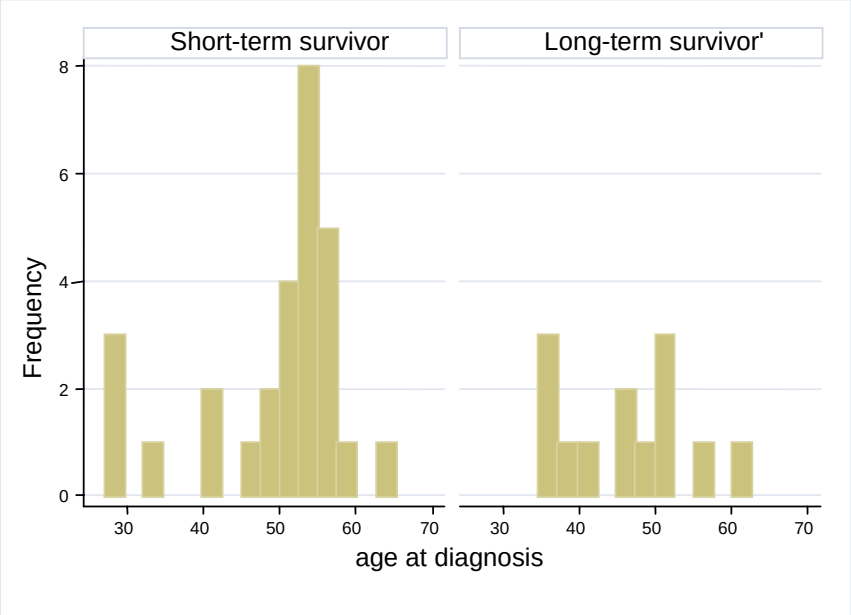


Figure 1: Frequency distribution of age at first operation in STSs and LTSs.

There appeared to be no difference in the proportion of males between LTSs and STSs. It was noted however that males in this study outnumbered females by around 2:1. This was consistent with earlier findings of the proportion of males to females with GBMs.

Fisher's exact test showed that there was a highly significant difference in IGFBP2 score between the LTSs and STSs with lower scores associated with long-term survival. The Mantel-Haenszel test for trend showed that this followed a trend of increasing long term survival with decreasing IHC scores. Logistic regression coefficients were calculated for each level of the IHC score. The coefficients for levels 0 and 1 did not display a linear trend with levels 2 and 3 and so they were combined to give 3 levels of IHC score for IGFBP2 to improve the linear trend. The linearity between IHC scores and the logit of the proportion of patients who were LTSs was improved with 3 levels of IHC scores. The p-

value for this trend was adjusted using a modified Bonferroni correction to adjust for this data determined cutpoint.

There were only 2 patients with IQGAP1 IHC scores of 0 and only one patient with a score of 1 and the coefficients for levels 0, 1 and 3, 4 did not display a linear trend and so these scores were combined to give 3 levels of IHC score for IQGAP1. The logistic regression coefficients showed a linear trend after combining the groups which was not significant after adjusting the p-value using a modified Bonferroni correction.

There was no significant difference in PTEN IHC score between LTSs and STSs.

With the percent nuclei staining for MGMT divided into quartiles there was a significant linear trend as measured by the Mantel Haenszel Test for Trend, but the logistic regression coefficients for 7-39% and 40-67% were exactly the same. The 7-39% and 40-67% quartiles were combined to give a 3 level variable: <7%, 7-67% and >67% to give a linear relationship between categories of nuclear staining and the logit of the proportion of patients who were LTSs. The p-value for this trend was adjusted using a modified Bonferroni correction to adjust for the data determined cutpoint.

Figure 2 below and a comparison in Table 1 of the quartile and the 3 level outcomes for percent nuclei staining for MGMT showed that there were differences between LTSs and STSs. No LTSs had less than 10% staining and most had more than 50%, while 10 of the

STSs had 0-5% staining. With MGMT coded as a 3 level variable the difference between STSs and LTSs was highly significant using Fisher's exact test.

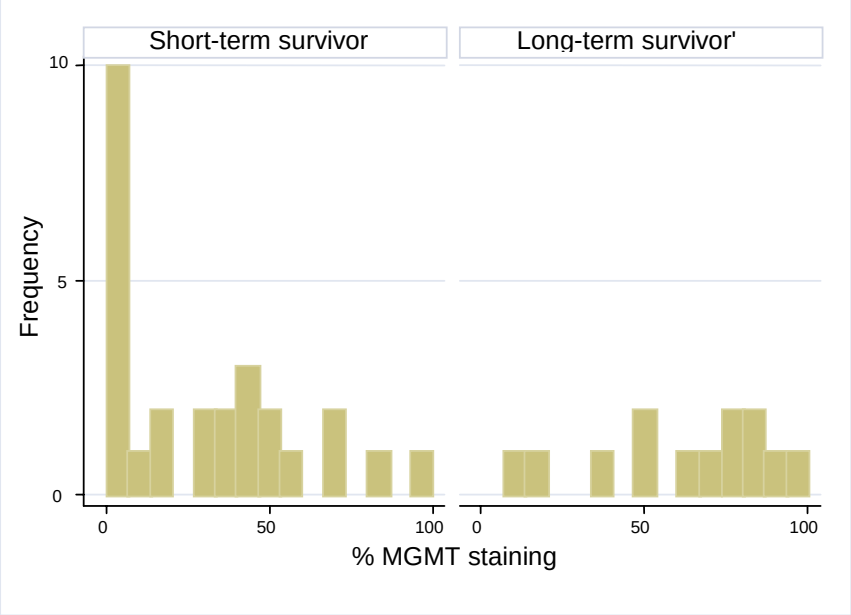


Figure 2: Frequency distribution of % nuclei staining for MGMT in STSs and LTSs.

3.2. Logistic Regression for predictors of long-term survival

Univariate Analysis with Logistic Regression

Table 2 gives odds ratios with 95% confidence intervals and p-values that resulted from univariate logistic regression with LT or ST survival as the outcome. Gender was not included as it had a $P > 0.25$ in exploratory analysis and was not a variable of interest. Age and PTEN expression did not significantly predict the outcome. IGFBP2 and MGMT IHC score were good predictors of LTS while IQGAP1 IHC score was just significant in predicting LTS. An increase of one level in the IGFBP2 IHC score; that is an increase from 0 or 1 to a score of 2 or from 2 to 3, was associated with an approximately 80%

reduction in the odds of becoming a LTS, while a one level increase in MGMT staining was associated with a 6 fold increase in the odds of becoming a LTS.

3.3 Fitting the baseline model

All variables with P < 0.25 in univariate analysis were included in the baseline model. These were IGFBP2 (3 levels), IQGAP1 (3 levels), PTEN and MGMT (3 levels). Age was also included although its univariate P = 0.28 was greater than 0.25, because it is usually a very strong predictor of survival and to control for differences between the ages of LT and STSs.

Table 2: Odds Ratios and p-values for predictors of LTS or STS in univariate models compared to the multivariate baseline model.

	Univariate		Baseline multivar model		
	Odds Ratio	P	Odds Ratio	P	VIF
IGFBP2–3 levels: 0&1, 2, 3	0.18	0.003	0.17	0.031	1.55
IQGAP1–3 levels: 0&1, 2, 3&4	0.33	0.047	0.44	0.31	1.33
PTEN	0.34	0.12	0.27	0.26	1.22
MGMT-3 levels <7, 7-67, >67%	6.4	0.005	9.33	0.012	1.04
Age (per year)	0.96	0.28	1.06	0.35	1.39
Log likelihood			-10.79		

The Odds Ratios and P-values for each predictor were compared between the univariate and multivariate models to check for large changes in significance or odds ratios as an indicator of confounding or collinearity. The variance inflation factor (VIF) was calculated as a measure of collinearity. The results are shown in Table 2. MGMT showed a large change in significance between the univariate and multivariate model, changing from P = 0.005 to P = 0.012 with its odds ratio changing from 6.4 to 9.3, a change of 46%.

When IGFBP2 was dropped from the baseline model the odds ratio for MGMT returned to its univariate value. This indicated that there could be confounding or collinearity between MGMT and IGFBP2 however the value for VIF indicated that collinearity was not the cause. There was also some confounding between IGFBP2 and both IQGAP1 and PTEN as the odds ratio for IQGAP1 changed by 30% and for PTEN by 23% and returned to their original values when IGFBP2 was removed, however there was again no indication of collinearity as measured by the VIF.

To investigate this further, the relationships between IHC results were examined using crosstabs and Fisher's exact test. This is shown in Table 3. The relationship between age in quartiles and IHC result was also examined, but only results for the two IHC variables with a Fisher's exact test significant at the 5% level have been included in Table 3. There was a significant relationship between IGFBP2 and IQGAP1. Higher values of IQGAP1 were associated with high values of IGFBP2. The p-value for Fisher's exact test was 0.02, so the relationship was not strong. Because Fisher's exact test indicated that the relationship between IGFBP2 and IQGAP1 was only just significant and because the VIFs in Table 2 were all below 2, collinearity did not seem to be a problem.

There was a very strong relationship however, between age and IGFBP2 score with a p-value of 0.008 and a just significant relationship between age and PTEN score with older age associated with higher scores of both.

Table 3: Crosstabs of frequency. The frequency, expected frequency and probability under the null hypothesis of no association between IHC scores using Fisher's exact test.

IGFBP2	IQGAP1 -3 level			MGMT -3 level			PTEN	
3 level	0	1	2	<7%	7-67	>67%	0	1
0	3 (1.3)	0 (0.5)	0 (1.2)	3 (4.4)	8 (8.8)	7 (4.8)	9 (6.6)	9 (11.4)
1	7 (4.4)	2 (1.7)	1 (3.9)	3 (1.7)	3 (3.4)	1 (1.9)	1 (2.6)	6 (4.4)
2	8 (12.3)	5 (4.8)	15 (10.9)	4 (3.9)	9 (7.8)	3 (4.3)	5 (5.9)	11 (10.1)
Fisher's exact p-value = 0.02				= 0.5			= 0.2	

IQGAP1	MGMT –3 level			PTEN	
3 level	<7%	7-67	>67%	0	1
0	1 (0.7)	1 (1.5)	1 (0.8)	1 (1.1)	2 (1.9)
1	3 (2.4)	3 (4.9)	4 (2.7)	4 (3.7)	6 (6.3)
2	6 (6.8)	16 (13.7)	6 (7.5)	10 (10.2)	18 (17.8)
Fisher's exact p-value = 0.5				= 1.0	

	MGMT- 3 level		
PTEN	<7%	7-67	>67%
0	3 (3.7)	8 (7.3)	4 (4.0)
1	7 (6.3)	12 (12.7)	7 (7.0)
Fisher's exact p-value = 0.9			

Age in	PTEN		IGFBP2 -3 level		
quartiles	0	1	0	1	2
<43	7 (4.0)	4 (7.0)	9 (4.8)	0 (1.9)	2 (4.3)
43-51	5 (3.7)	5 (6.3)	6 (4.4)	2 (1.7)	2 (3.9)
52-54	2 (3.7)	8 (6.3)	1 (4.4)	2 (1.7)	7 (3.9)
>54	1 (3.7)	9 (6.3)	2 (4.4)	3 (1.7)	5 (3.9)
Fisher's exact p-value = 0.04			= 0.008		

frequency
(expected frequency)

3.4. Finding the best model

To find the most parsimonious model, the least significant variable was removed one at a time and the remaining model retested for fit using the likelihood ratio test. The final model was also tested for fit using the residual sum of squares test. The results are shown in Table 4. Although the Hosmer Lemeshow test is the usual test of choice for model fit with logistic regression models, it has very poor power to reject a poorly fitting model with this sample size (Kuss; Hosmer and Hjort) and was unable to reject any of the models. The residual sum of squares test (RSS) displays better power with sparse data (Kuss, 2002).

Table 4: Assessing the fit of the models using the likelihood ratio test compared to the previous model and the unweighted residual sum of squares (RSS) test.

	Likelihood ratio test			RSS test	
Variables modelled	-2 log likelihood	Ch^2 (df)	P	z	P
IGFBP2,IQGAP1,MGMT,PTEN, age	25.68	-	-		
IGFBP2,IQGAP1,MGMT,PTEN	26.61	0.93 (1)	0.33		
IGFBP2,IQGAP1,MGMT	27.76	1.15 (1)	0.28		
IGFBP2,MGMT	28.58	0.82 (1)	0.37	0.65	0.5
IGFBP2	37.60	9.02 (1)	0.003		

From Table 4 we can see that the final model with only IGFBP2 and MGMT was the most parsimonious model to fit the data. Table 5 gives the Odds Ratios, 95% confidence interval, coefficients and P-values for this final model. For each increase in IGFBP2 IHC score (with 3 levels) there was around an 80% decrease in the odds of becoming a LTS.

The effect of having 7-67% compared to <7% nuclei staining for MGMT or >67% compared to 7-67% was around an 8 fold increase in the odds of becoming a LTS.

Table 5: Parsimonious logistic regression model for LTS or STS, with IGFBP2 and MGMT expression as predictors.

	Odds ratio	95% CI	coeff	P
IGFBP2 -3 levels:0&1, 2, 3	0.17	0.05-0.60	-1.76	0.006 (0.02)#
MGMT -3 levels: <7%,7-67%,>67%)	7.6	1.5-38	2.03	0.014 (0.05)#
c-statistic = 0.91 (Hanley and McNeil, 1982)				

adjusted p-value using modified Bonferroni correction to adjust for groups merged to maximise trend.

3.5. ROC analysis

Only IGFBP2 and MGMT were considered for ROC analysis as only these two predictors were present in the final logistic regression model. Figure 3A shows separate ROC plots for IGFBP2 and for MGMT expression with the MGMT in its original form and IGBP2 with 3 levels. IGFBP2 and MGMT had the same area under the curve of 0.80. Figure 3B shows IGFBP2 and MGMT categorised into binary variables at the best point for predicting long-term survival which was between 1 and 2 of the original IHC scores for IGFBP2 and between 51 and 52 percent nuclei staining for MGMT. The area under the curve for these binary variables was similar to the variables in their original form, indicating that if clinicians would like to advise their patients whether they were likely to become a LTS based on IHC results for these genes, choosing these as cut-offs would give a similar balance of sensitivity and specificity as the original scores.

Clinicians may also like to know if combining information from IHC scores from both genes is more useful in predicting which patients are likely to be LTSs than each gene taken separately. To test this, a combined variable using both the IGFBP2 and MGMT result was created. The combined variable IGFBP2 & MGMT had a value of 1 if a patient had a score of 0 or 1 for IGFBP2 and 52% or more nuclei staining for MGMT and zero otherwise. The area under a ROC plot of this variable was greater than the area under the individual curves for binary IGFBP2 or MGMT but not significantly greater using the roccomp test in Stata, so this combined variable does not provide more information for clinicians. Table 6 shows that IGFBP 2 has better sensitivity; MGMT has better specificity while the combined variable has worse sensitivity and better specificity than using the variables separately. In this case sensitivity means correctly predicting which patients will become LTSs and specificity means correctly predicting who will become a STS. From the patients' perspective it is preferable to be able to predict LTSs, making IGFBP2 expression status the better clinical predictor.

Table 6: Sensitivity, specificity, positive and negative predictive value and % correctly predicted, using IGFBP2, MGMT or a combined variable obtained by multiplying a binary variable from IGFBP2 (scores of 0 &1 versus scores of 2 & 3) by a binary MGMT variable (less than or more than 52% staining).

	Sensitivity	Specificity	% correct
IGFBP2 -3 levels: 0&1, 2, 3	84.6%	75.0%	78.1%
IGFBP2 –binary: 0&1, 2&3	84.6%	75.0%	78.1%
MGMT -3 levels: <7%, 7-67%, >67%	53.9%	85.7%	75.6%
MGMT -binary: <52%, ≥52%	76.9%	81.5%	80.0%
IGFBP2 (bin) & MGMT (bin)	69.2%	92.6%	85.0%

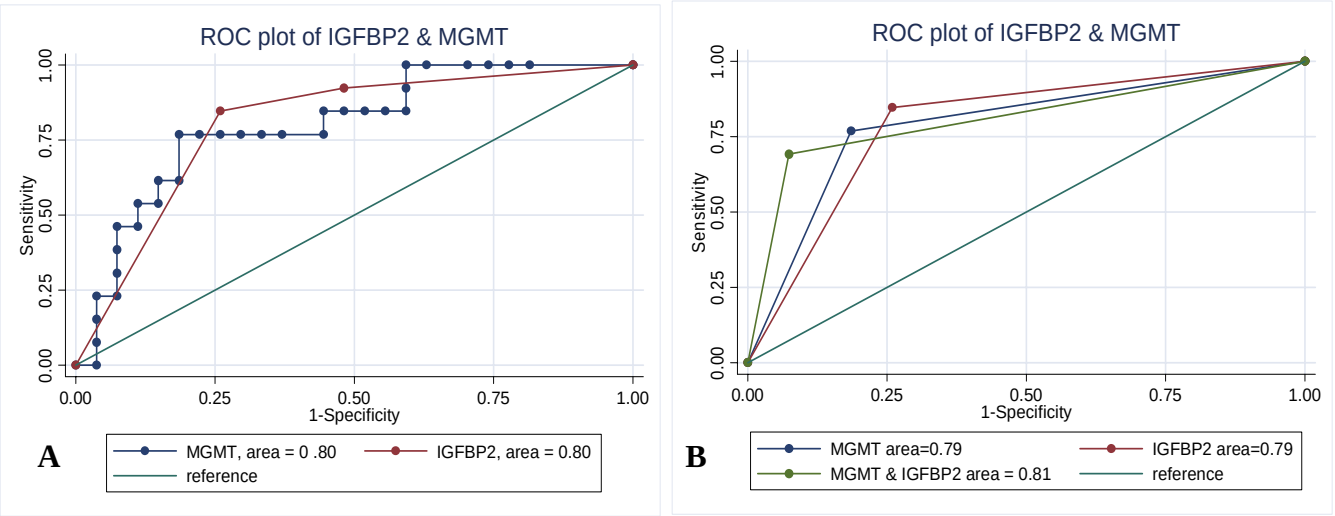


Figure 3: ROC plots showing sensitivity and 1-specificity of the classification of patients as LTSs or STSs based on their IGFBP2 and MGMT immunohistochemistry. Plot A shows all levels of MGMT and 3 levels of IGFBP2. The values for IGFBP2 are in reverse order to allow them to be plotted on the same graph. Plot B has MGMT categorised into <52% or ≥52% stained nuclei and IGFBP2 into 0 or 1 compared to 2 or 3, again in reverse order.

3.6. Survival Analysis

Survival analysis uses survival time as the outcome with information from patients with the shortest to those with the longest interval from diagnosis until death. This compares to logistic regression which only considers whether the subject was a LTS or a STS. Because of this, the relationship between predictors and outcome may not necessarily be the same since the expression levels of a gene may have one effect on the risk of death within the first months after diagnosis and a different effect later if the patient survives these first months. While the primary interest of this project is on predictors of long term survival it was thought that survival analysis might reveal more detailed effects of the expression of these genes on prognosis. An additional exploratory analysis was carried

out using Kaplan Meier curves and incidence rates to find which IHC scores had similar survival curves. This is shown in Table 7.

3.7: Exploratory Analysis

Table 7: Results of exploratory survival analysis.

Variable	Level	No.	Incidence rate/day	Logrank test Chi^2 (df)	P
IGFBP2 (3 levels)	0, 1	18	0.0009	5.80 (2)	0.055
	2	7	0.0018		
	3	16	0.0016		
IGFBP2 (binary)	0, 1	18	0.0009	5.80 (1)	0.016
	2, 3	23	0.0016		(0.06)#
IQGAP1 (3 levels)	0, 1	3	0.0005	1.87 (2)	0.39
	2	10	0.0012		
	3, 4	28	0.0014		
IQGAP1 (binary)	0, 1	3	0.0005	1.71 (1)	0.19
	2, 3, 4	38	0.0013		(0.8)#
MGMT (quartiles)	< 7%	10	0.0031	21.35 (3)	0.0001
	7-39%	10	0.0012		
	40-67%	10	0.0013		
	>67%	11	0.0007		
MGMT (3 levels)	<7%	10	0.0031	20.68 (1)	<0.0001
	7-67%	20	0.0012		(0.0001)#
	>67%	11	0.0007		
PTEN	0	15	0.0010	1.25	0.26
	1	26	0.0013		
Gender	Female	15	0.0014	0.70 (1)	0.40
	Male	26	0.0011		
Age (in quartiles)	<43	11	0.0009	3.13 (3)	0.37
	43-51	10	0.0012		
	52-54	10	0.0016		
	>54	10	0.0014		

adjusted p-value using modified Bonferroni correction to adjust for groups merged to maximise trend.

From the incidence rates and Kaplan Meier curves it could be seen that there was a difference in survival between IGFBP2 IHC scores of levels 0 and 1 compared to levels 2 and 3 but none between level 2 and 3, so a binary form of IGFBP2 was used for survival analysis. IQGAP1 had a difference in exploratory analysis between levels 0 and 1 compared to 2, 3 and 4, so again a binary form would have been better to use for survival analysis. However as there were only three subjects with level 0 or 1 scores, three levels of score were used as in the previous section.

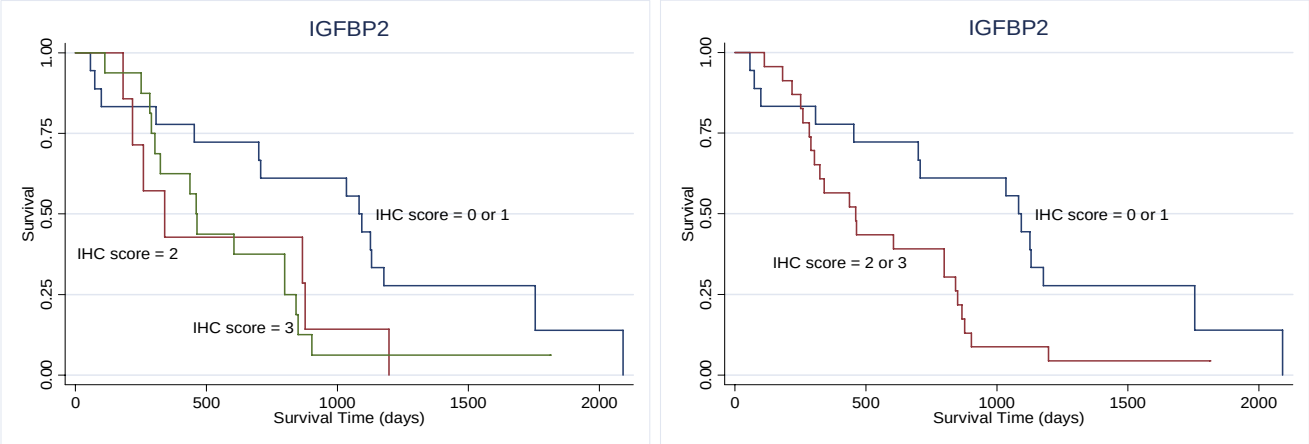


Figure 4: Kaplan Meier Curves for IGFBP2 with three or two levels of IHC score.

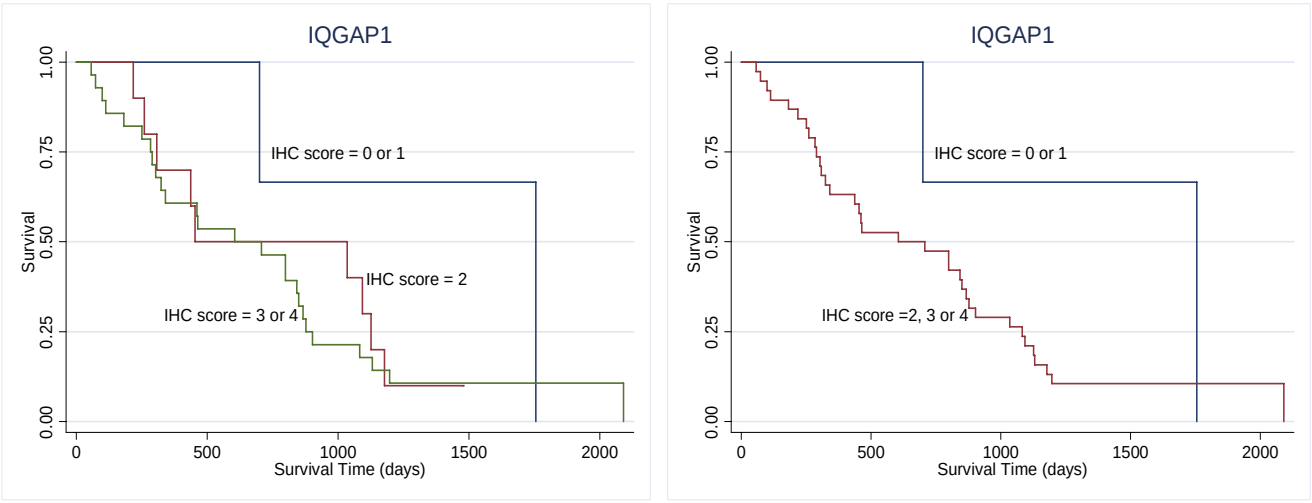


Figure 5: Kaplan Meier Curves for IQGAP1 with three or two levels of IHC score.

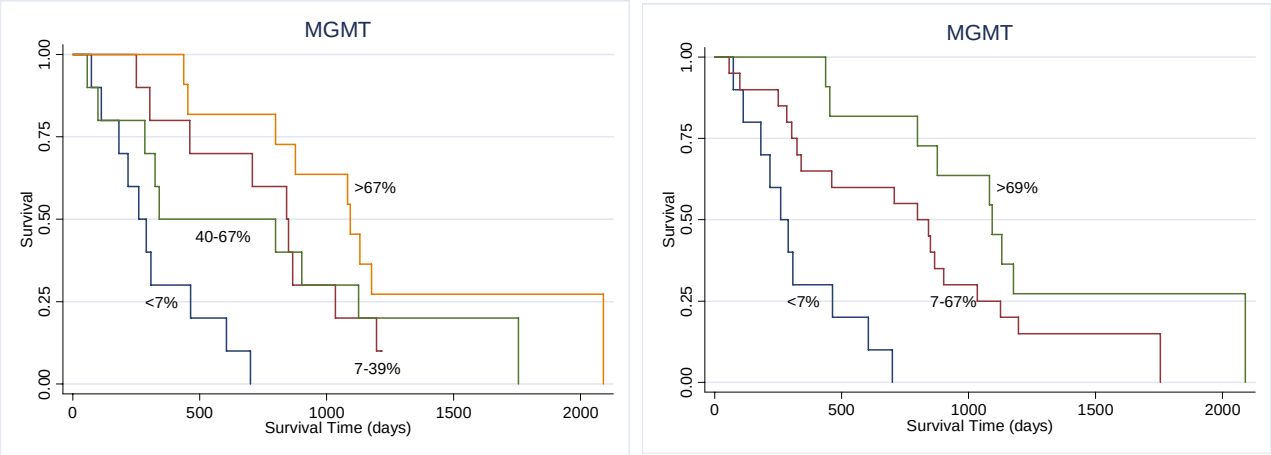


Figure 6: Kaplan Meier curves for MGMT percentage stained nuclei, divided at quartiles or divided into3 levels at <7%, 7-67% and >67%.

Exploratory analysis of MGMT showed a difference in survival between <7% stained nuclei compared to 7-67% and between 7-67% compared to >67%, with no difference between 7-39% and 40-67%. With these two quartiles merged to give 3 levels: <7%, 7-67% and >67% stained nuclei the difference in survival between the groups was clear so as before the 3 level form of MGMT was used.

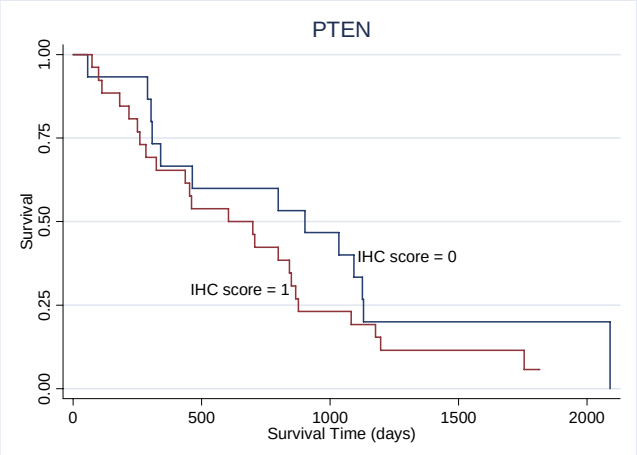


Figure 7: Kaplan Meier curves for PTEN IHC score

As with the earlier analysis for predictors of long term survival, there was no influence apparent from PTEN IHC score, gender or age on survival time in exploratory analysis.

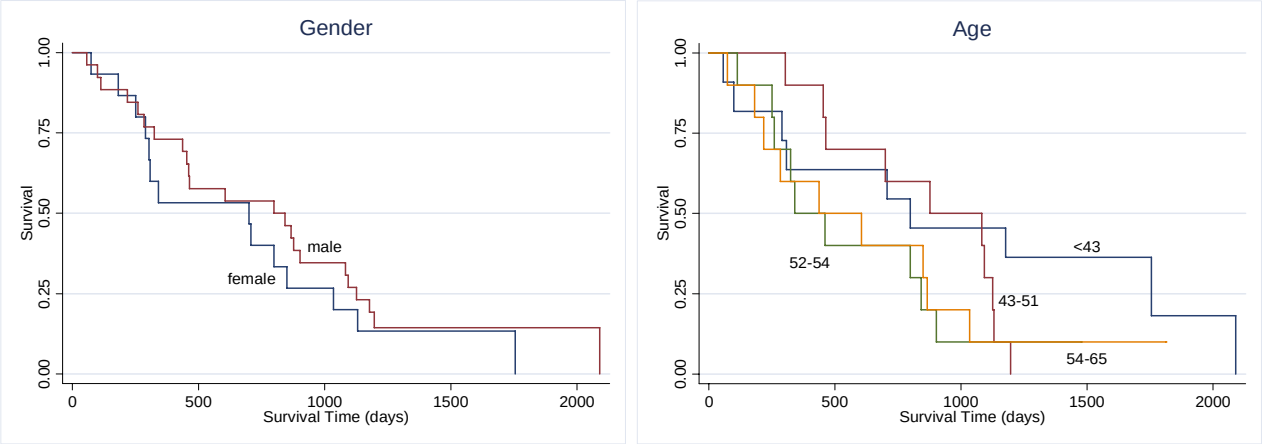


Figure 8: Kaplan Meier curves for gender and age.

3.8. Fitting the baseline multivariate model

All variables with P < 0.25 in univariate analysis using the logrank test were included in the baseline model. These were IGFBP2 (binary), IQGAP1 (3 levels), and MGMT (3 levels) based on the results of the previous section. Age was included although its univariate p-value was 0.30, because it is usually a very strong predictor of survival. PTEN with P = 0.27 was not included.

The Hazard Ratios and P-values for each predictor were compared between the univariate and multivariate models to check for confounding. The results are shown in Table 8. The predictor with the greatest changes in the size of the Hazard Ratio were IGFBP2 which changed from 2.32 to 1.56 a change of 33% indicating confounding between IQGAP1 and IGFBP2 due to correlation between the two variables which was increased when IGFBP2

was expressed as a binary variable. The p-value for IGFBP2 also went from slightly significant at P = 0.017 to not significant at P = 0.9 in the presence of IQGAP1. However from the size of the VIFs there did not appear to be any serious collinearity.

Table 8: Hazard Ratios and P-values using predictors of survival time in univariate models compared to the multivariate baseline model.

	Univariate		Baseline multiv model		
	Hazard Ratio	P	Hazard Ratio	P	VIF
IGFBP2 -binary: 0&1, 2&3	2.32	0.019	1.56	0.88	1.75
IQGAP1 -3 levels: 0&1, 2, 3&4	1.42	0.20	1.29	0.73	1.30
MGMT -3 levels: <7%, 7-67%, >67%	0.34	<0.0001	0.34	0.001	1.08
Age	1.02	0.30	1.00	0.98	1.32
Log likelihood			-99.02		

3.9. Finding the best model

To find the most parsimonious model, the least significant variable was removed one at a time and the remaining model retested for fit using the likelihood ratio test. The global proportional hazards assumption was tested for each model. The results are shown in Table 9.

The final model with only MGMT was the most parsimonious model that fitted the data. The Hazard Ratio was 0.34 (95% confidence interval: 0.19-0.60) which meant patients with 7-67% MGMT staining had a 65% reduction in the instantaneous risk of death from GBM compared to patients with <7% staining. Likewise patients with >67% staining had

a similar reduction of risk when compared to patients with 7-67% staining. The p-value was 0.0001 and the adjusted p-value 0.0004 using a modified Bonferroni to adjust for optimizing the cutpoint to maximize trend.

Table 9: Assessing the fit of the models using the likelihood ratio test compared to the previous model, and global test of the proportional hazards assumption using Schoenfeld residuals.

	Likelihood ratio test			Test of global proportional hazards assumption	
Variables modelled	-2 log likelihood	Ch^2 (df)	P	Ch^2 (df)	P
IGFBP2,IQGAP1,MGMT, age	198.03	-	-	2.33 (4)	0.67
IGFBP2,IQGAP1,MGMT	198.03	0.00 (1)	1.0	2.04 (3)	0.56
IGFBP2,MGMT	198.57	0.54(1)	0.46	0.34 (2)	0.84
MGMT	202.21	3.64 (1)	0.056	0.44 (1)	0.50

For the results of analysis without the original three LTSs, see the Appendix.

4. Discussion

The data have been analysed using both the original dataset including the original three LTSs who prompted the study, and in the dataset without these the original patients. Logistic regression and ROC analysis with survival for more than 1000 days as the outcome were used to find proteins whose over- or under-expression as measured by IHC were able to predict long term survival in GBM patients. Cox regression analysis was used to find which proteins were able to predict survival time.

These methods gave slightly different results. Patients with low IGFBP2 immunohistochemistry scores and/or a high percentage of nuclei staining for MGMT in tumour sections taken at initial biopsy or surgery were more likely to survive for more than 1000 days. Having an IHC score of 2 for IGFBP2 compared to a score of 0 or 1, or a score of 3 compared to 2, reduced the odds a patient will survive for more than 1000 days by 85%. Having more than 7-67% MGMT stained nuclei compared to less than 7% or more than 67% compared to 7-67%, increased the odds of surviving more than 1000 days by around 8 fold. However, the best predictor of survival time as opposed to survival for more than 1000 days was having a high percentage of nuclei staining for MGMT in multivariate analysis. Having >7% staining or more than 67% staining reduced the instantaneous risk of dying by 65% when compared to <7% or 7-67% respectively. Patients with less than 7% staining had a particularly poor prognosis with a median survival time of only 250 days.

Considering the IGFBP2 IHC score in these patients as well did not significantly improve the ability to predict survival time. MGMT may be more significant in survival analysis because the survival curves for IGFBP2 only show a better outcome for subjects with low IHC scores after 250 days. Before 250 days after diagnosis, subjects with IGFBP2 scores of 0 or 1 have worse survival than those with higher scores. This contrasts with the Kaplan Meier curves for MGMT which show better survival throughout in patients with higher percentages of MGMT staining. This was not apparent when looking at whether a patient is likely to become a LTS as by 1000 days the difference in survival between patients with low compared to high IHC scores for IGFBP2 was very large. The primary aim of this project was to find predictors of LTS, and both IGFBP2 and MGMT protein expression are the best at this, but it was interesting that patients with low levels of IGFBP2 expression have a poorer prognosis for the first 250 days after diagnosis and so it was not as good a predictor of the instantaneous risk of death as measured by Cox regression as MGMT protein expression.

An advantage of survival analysis over logistic regression modelling in this situation is that the Kaplan Meier curve reveals such changes in risk over time and so provides a more detailed picture of the relationship between the explanatory variables and the risk of death. The Kaplan-Meier curve requires few assumptions and is easy to calculate but is inefficient compared to parametric survival estimators such as Cox regression. Cox regression and logistic regression typically produce the same p-values and very similar regression coefficients when used to analyse the same data with large N (Moriguchi et al, 1993), however both Cox regression and logistic regression are prone to type I and type II

errors with small numbers of events per variable (EPV) and Cox regression is more sensitive than logistic regression to this (Vittinghoff and McCulloch, 2006). This limits the usefulness of Cox regression with this data. In our study, EPV is near the limit at which serious problems can be expected with logistic regression and well within the area where serious problems can be expected with Cox regression. See the limitations section for more on this.

Clinicians need to be able to advise their patients about the likely course of their disease. The prognosis for patients diagnosed with GBM is never good, but since some patients do live for three years or more after diagnosis, clinicians are likely to be interested in whether studies such as this can help predict which patients will live longer than the median GBM patient. ROC analysis gives the sensitivity and specificity of predicting an event such as LT survival. This study has shown that IGFBP2 has better sensitivity for predicting LT survival while MGMT has better specificity. Since it is better from the patients' point of view to correctly predict who will become a LTS than to correctly predict who will become a STS, this makes IGFBP2 the better predictor for the clinician. However this should be tempered by the observation that some patients with low levels of IGFBP2 expression have worse survival in the first year after diagnosis. This effect should be taken into account by clinicians before using low IHC expression for IGFBP2 to predict the likelihood of a patient becoming a LTS.

The over-expression of IGFBP2 in cancer and its association with poor prognosis is well established, both in GBMs and many other tumour types; however the molecular

mechanisms by which IGFBP2 enhances tumour cell growth and increases tumorigenicity remain undefined (Wang et al, 2003). A number of studies have shown causal links between the level of expression of IGFBP2 and levels of expression of other genes upregulated in GBMs. Wang et al (2003) found MMP2 over-expression when IGFBP2 was over-expressed in cell lines and Fukushima et al (2007) found CD4 expression could be could suppressed by under expressing IGFBP2. Zhou et al (2005) also found significant correlation between the expression of IGFBP2 and MMP2 in gliomas. In addition they found significant correlation between IGFBP2 expression and VEGF (vascular endothelial growth factor) expression in the same study. In our study we found that IGFBP2 expression correlates strongly with IQGAP1 expression, another gene that appears to be involved in invasiveness of tumours. IGFBP2 appears to be a central player in controlling a number of genes promoting the invasiveness of GBMs.

There is a strong relationship between age and levels of both IGFBP2 and PTEN expression with high levels of expression of both these genes found in GBMs from older patients. Age over 60 years is strongly associated with poor prognosis and high IGFBP2 expression levels is also associated with poor prognosis (Scott et al, 1999), so the finding that IGFBP2 expression correlates with age is interesting and could be part of the reason for poorer survival in older patients. However higher expression of PTEN, according to earlier findings should be associated with a better prognosis, while older age is associated with a worse prognosis. Park et al (2002) found that the introduction of the wild-type PTEN genes reduced invasion invitro, and Zhou et al (2005), found that patients with high levels of PTEN mRNA expression as measured by quantitative PCR, were more likely to

have favourable outcomes. However in this study no association between PTEN protein expression and survival time was found, either in univariate or multivariate analysis. It could be that the relationship between high levels of PTEN expression and longer survival is only seen in older patients who were not included in this study or alternatively that the increased mRNA expression is not reflected in protein levels as measured by IHC.

Higher levels of protein expression of MGMT were significantly associated with better survival for more than 1000 days. However Hegi et al (2005) found that MGMT promoter methylation, which would be expected to result in lower MGMT expression, was a favourable prognostic factor for longer survival irrespective of treatment and also was associated with a more favourable response of patients to temozolomide and radiotherapy treatment. However low levels of expression of MGMT would also be expected to result in an increased frequency of mutations, particularly in these patients who are undergoing radiotherapy as well as chemotherapy. This may lead to mutations that give rise to more aggressive tumour behaviour. This may be having the predominant effect in this study.

The expression of MGMT does not correlate with IGFBP2 expression and so MGMT represents an independent gene or gene pathway associated with prognosis. In biological terms it means that the control of IGFBP2 expression and MGMT expression are independent of each other.

4.1 Analysis without the original Log Term Survivors.

IQGAP1 expression as measured by IHC was only a significant predictor of long term survival in the dataset that included the original three LTSs and then only with a p-value of 0.047. Since IQGAP1 expression was one of the variables found to be down-regulated in two of the original three long-term survivors, the fact that this is not confirmed in the larger group without these originals may mean the result of the analysis in the data set containing the originals was biased by including them. This affected the analysis in several ways. Although there appeared to be a linear trend between IQGAP1 score and LTS with the original three patients, this trend disappeared without them. Kaplan Meier curves appeared to show a weak effect of IQGAP1 IHC score on survival but this disappeared without the three original LTSs. Since there was a relationship between the protein expression of IQGAP1 and IGFBP2, any apparent effect of IQGAP1 may be through this relationship rather than a direct effect of IQGAP1 on tumour growth and survival.

Two of the original three LTSs were over 55 years of age while the average age of the other LTSs was 44 and so another way in which the inclusion of the original LTSs affected the analysis was through age. There was a significant difference in age between the LTSs and STSs without the original three LTSs using the Mann Whitney test, but not using logistic regression.

Both IGFBP2 and MGMT became less significant predictors of LTS in logistic regression without the three originals, presumably because of the loss of power associated with there being three fewer events to model. However IGFBP2 became more significant in survival

analysis without the original three, possibly because one of these three was the only LTS to have a score of 3 for IGFBP2.

4.2 Limitations

Small studies such as the one conducted for this project have disadvantages from the point of view of statistical analysis and the subsequent interpretation of results. Both logistic regression and Cox regression have problems when there are small numbers of events per variable (EPV). There were only 13 events (LTSs) in this study or 10 if we exclude the original LTSs. With 2 variables in the final model, we are at the limit below which one can expect problems with type I error in logistic regression. Power falls steadily as the number of EPV decreases from 20 to 10 in logistic regression and then falls sharply below 10 EPV leading to problems increasing type II error rates (Vittinghoff and McCulloch, 2006) and (Peduzzi et al, 1996). These problems are more extreme with Cox regression and occur at higher EPV. In our study, EPV is near the limit at which serious problems can be expected with logistic regression and well within the area where serious problems can be expected with Cox regression. However the most likely effect of small EPV on logistic regression analysis is a conservative conclusion due to lack of power and so significant results are likely to be found to even more strongly significant when repeated with more EPV. With Cox regression the problems are likely to be more extreme and so these results should be treated with caution.

The study described in this project used all the subjects available through the Kolling Institute's Brain Tumour Bank and the Sydney Neuro-Oncology Group at the time the

immuno-histochemistry was carried out. As time passes, more patients will become available and still more patients can be recruited by collaborating with other institutions, making a larger study feasible. My former supervisor, Kerrie McDonald, is presently seeking funding for a larger study, with more subjects and using more genetic markers.

References

Fukushima T, Kataoka H, 2007, 'Roles of insulin-like growth factor binding protein-2 (IGFBP-2) in glioblastoma.' *Anticancer Res*, vol 27, no. 6A, pp3685-92.

Hanley JA McNeil BJ, 1982, 'The meaning and use of the area under a receiver operating characteristic (ROC) curve.' *Radiology.* Vol 143, no. 1, pp29-36.

Hegi ME, Diserens AC, Gorlia T, Hamou MF, de Tribolet N, Weller M, Kros JM, Hainfellner JA, Mason W, Mariani L, Bromberg JE, Hau P, Mirimanoff RO, Cairncross JG, Janzer RC, Stupp R, 2005, 'MGMT gene silencing and benefit from temozolomide in glioblastoma.' *N Engl J Med*. vol 352, no. 10, pp997-1003.

Hosmer DW, Hjort NL, 2002, 'Goodness-of-fit processes for logistic regression: simulation results.' *Stat Med.* Vol 21, no. 18, pp2723-38.

Kay R, 2007, *Statistical thinking for non-statisticians in drug evaluation,* Wiley-Interscience, p76.

Kuss O, 2002,'Global goodness-of-fit tests in logistic regression with sparse data,' *Stat Med,* vol 21, no24, pp3789-801.

Mellinghoff IK, Wang MY, Vivanco I, Haas-Kogan DA, Zhu S, Dia EQ, Lu KV, Yoshimoto K, Huang JH, Chute DJ, Riggs BL, Horvath S, Liau LM, Cavenee WK, Rao PN, Beroukhim R, Peck TC, Lee JC, Sellers WR, Stokoe D, Prados M, Cloughesy TF, Sawyers CL, Mischel PS, 2005, 'Molecular determinants of the response of glioblastomas to EGFR kinase inhibitors,' *N Engl J Med*, vol 353, no. 19, pp2012-24.

McDonald KL, O'Sullivan MG, Parkinson JF, Shaw JM, Payne CA, Brewer JM, Young L, Reader DJ, Wheeler HT, Cook RJ, Biggs MT, Little NS, Teo C, Stone G, Robinson BG; 2007, 'IQGAP1 and IGFBP2: valuable biomarkers for determining prognosis in glioma patients,' *J Neuropathol Exp Neurol,* vol 66, no. 5, pp405-17.

Moriguchi S, Hayashi Y, Nose Y, Maehara Y, Korenaga D, Sugimachi K, 1993, 'A comparison of the logistic regression and the Cox proportional hazard models in retrospective studies on the prognosis of patients with gastric cancer,' *J Surg Oncol*, vol 52, no.1, pp9-13.

Ohgaki H, Kleihues P, 2007, 'Genetic Pathways to Primary and Secondary Glioblastoma,' *The American Journal of Pathology*, vol 179, no. 5, pp1445-1453.

Park M-J, Kim M-S, Park I-C, Kang H-S, Yoo H, Park SH, Rhee CH, Hong SI, Lee SH, 2002, 'PTEN Suppresses Hyaluronic Acid-induced Matrix Metalloproteinase-9

Expression in U87MG Glioblastoma Cells through Focal Adhesion Kinase Dephosphorylation,' *Cancer Research, vol* 62, pp6318-6322.

Peduzzi P, Concato J, Kemper E, Holford TR & Feinstein AR, 1996, 'A Simulation Study of the Number of Events per Variable in Logistic Regression Analysis,' *J Clin Epidemiol* , vol. 49, no. 12, pp. 1373-1379.

Scott JN, Rewcastle NB, Brasher PM, Brahser MA, Fulton D, MaKinnon JA, Hamilton M, Cairncross JG, Forsyth P, 1999, 'Which Glioblastoma Patient Will Become a Long-Term Survivor? A Population-Based Study,' *Ann Neurol*, vol 46, no. 2, pp183-8.

Tsuyoshi Fukushima, Tomoaki Tezuka, Takeshi Shimomura, Shinichi Nakano, and Hiroaki Kataoka, 2007, 'Silencing of Insulin-like Growth Factor-binding Protein-2 in Human Glioblastoma Cells Reduces Both Invasiveness and Expression of Progression-associated Gene CD24,' *The Journal of Biological Chemistry,* vol. 282, no. 25, pp. 18634–18644.

Vittinghoff E and McCulloch CE, 2006, 'Relaxing the Rule of Ten Events per Variable in Logistic and Cox Regression,' *American Journal of Epidemiology*, vol. 165, no. 6, pp710-718.

Wang H, Wang H, Shen W, Huang H, Hu L, Rambas L, Zhou YH, Liao WS, Fuller GN, Zhang W, 2003,'Insulin-like growth factor binding protein 2 enhances glioblastoma invasion-enhancing genes,' *Cancer Research,* vol 63, pp4315-4321.

Zhou Y, Hess KR, Liu L, Linskey ME, Yung WK, 2005, 'Modelling prognosis for patients with malignant astrocytic gliomas: Quantifying the expression of multiple genetic markers and clinical variables,' *Clinical Neuro-Oncology*, vol 7, pp485-494.

Appendix: Analysis without the original three LTSs.

3.1. Exploratory analysis

The results of exploratory analysis without the original three LTSs are given in Appendix Table 1. The results were similar to those with the three originals except a significant difference in age was found (P = 0.03) with the LTSs about five years younger on average than the STSs.

Another difference was that the linear trend between long term survival and IQGAP1 score was worse in the dataset without the original LTSs and was still not present with 3 levels of IQGAP1 score.

A comparison of Figure 1 and Appendix Figure 1 shows the reason for the significant difference in age between LTSs and STSs: two of the original three LTSs were over 55 years of age.

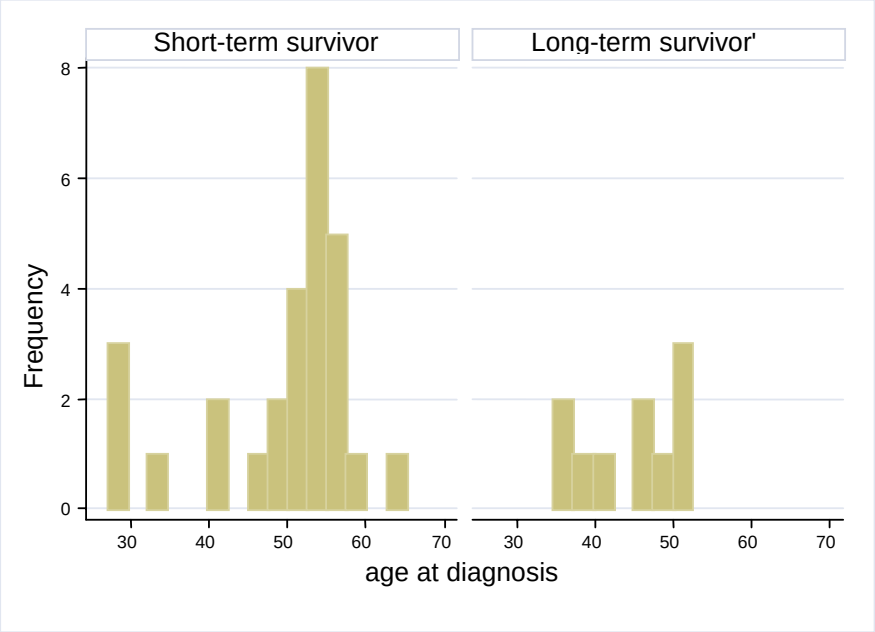
Appendix Table 1: Comparison of LTS and STSs by age, gender, IGFBP2, IQGAP1, PTEN and MGMT score without the original 3 LTSs.

		Frequency (%)			
		Short term	Long term	coeff by level*	Tests for significant effects and trend
Number		28	10		
Gender	Female	11 (84.6)	2 (15.4)		Fisher's exact = 0.44
	Male	17 (68)	8 (32)		
IGFBP2 (original IHC scores)	0	6 (60)	4 (40)	0	Fisher's exact P < 0.0001 Mantel Haenszel test for trend P = 0.003
	1	1 (16.7)	5 (83.3)	2.01	
	2	6 (85.7)	1 (14.3)	-1.39	
	3	15 (100)	0 (0)	-	
IGFBP2 (3 levels)	0&1	7 (43.7)	9 (56.3)	0	Fisher's exact P < 0.001 Mantel Haenszel test for trend P = 0.0004 (0.002)#
	2	6 (85.7)	1 (14.3)	-2.04	
	3	15 (100)	0 (0)	-	
IQGAP1 (original IHC scores)	0	1 (50)	1 (50)	0	Fisher's exact = 0.34 Mantel Haenszel test for trend P = 0.18
	1	0 (0)	0 (0)	-	
	2	5 (55.6)	4 (44.4)	-0.22	
	3	7 (87.5)	1 (12.5)	-1.95	
	4	15 (78.9)	4 (21.1)	-1.32	
IQGAP1 (3 levels)	0&1	1 (50)	1 (50)	0	Fisher's exact = 0.18 Mantel Haenszel test for trend P = 0.10 (0.4)#
	2	5 (55.6)	4 (44.4)	-0.22	
	3&4	22 (81.5)	5 (18.5)	-1.48	
MGMT % stained nuclei in quartiles	<7%	10 (100)	0 (0)	-	Fisher's exact = 0.019 Mantel Haenszel test for trend P = 0.004
	7-39%	7 (77.8)	2 (22)	-1.66	
	40-67%	7 (77.8)	2 (22)	-1.66	
	>67%	4 (40)	6 (60)	0	
MGMT (3 levels)	<7%	10 (100)	0 (0)	-	Fisher's exact = 0.001 Mantel Haenszel test for trend P = 0.009
	7-67%	14 (77.8)	4 (22)	17.3	
	>67%	4 (40)	6 (60)	18.9	
PTEN IHC score	0	8 (57.1)	6 (42.9)		Fisher's exact P = 0.08
	1	20 (83.3)	4 (16.7)		

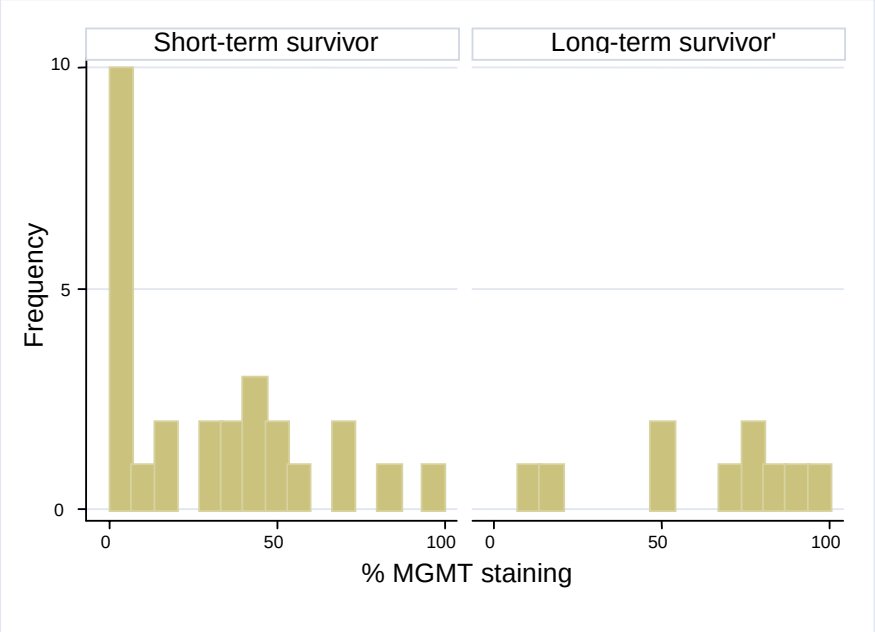
		Short term	Long term	
Survival time	Lower quartile	255	1126	
	Median	389	1184.5	
	Upper quartile	752.5	1214	
Age at 1st surgery	Mean	49.1	44.4	Mann Whitney test P = 0.03
	Stand dev	9.4	6.6	

* Coeff by level: the logistic regression coefficient at each level of IHC score with the scores in their original or final form.

p-value with modified Bonferroni correction in groups merged to maximise trend.



Appendix Figure 1: Frequency distribution of age at first operation in STSs and LTSs without the original LTSs.



Appendix Figure 2: Frequency distribution of % nuclei staining for MGMT in STSs and LTSs without the original LTSs.

3.2. Logistic Regression for predictors of long-term survival

Univariate Analysis with Logistic Regression

Appendix Table 2 gives the results of univariate logistic regression. They were very similar to those with the original LTSs, except IQGAP is no longer significant either in its original form with five levels of score or with three levels of score. Age was not significantly different between the two groups using this method of analysis.

3.3. Fitting the baseline model

Appendix Table 2: Odds Ratios and Ps for predictors of long term or short term survival in univariate models compared to the multivariate baseline model.

	Univariate		Baseline multivar model		
	Odds Ratio	P	Odds Ratio	P	VIF
IGFBP2 – 3 levels: 0&1, 2, 3	0.08	0.013	0.05	0.035	1.56
IQGAP1 – original scores	0.65	0.19	1.10	0.86	1.24
IQGAP1 – 3 levels: 0&1, 2, 3&4	0.38	0.12			
PTEN	0.27	0.086	0.16	0.17	1.40
MGMT – 3 levels: <7%, 7-67%, >67%	7.36	0.007	7.5	0.041	1.07
Age	0.94	0.16	1.10	0.22	1.66
Log likelihood			-9.42		

In the baseline model containing all variables with p-values <0.25, the results were very similar to those with the original LTSs. IQGAP1 has been used in its original form as merging scores made no difference to the linearity of trend.

Appendix Table 3: Crosstabs of frequency. Shows the frequency, expected frequency and probability under the null hypothesis of no association between IHC scores using Fisher's exact test.

	IQGAP1-3 level			MGMT-3 level			PTEN	
IGFBP2	0	1	2	<7	7-67%	>67%	0	1
0	2 (0.8)	6 (3.8)	8 (11.4)	3 (4.2)	6 (7.6)	7 (4.2)	8 (5.9)	8 (10.1)
1	0 (0.4)	2 (1.7)	5 (5.0)	3 (1.8)	3 (3.3)	1 (1.8)	1 (2.6)	6 (4.4)
2	0 (0.8)	1 (3.6)	14 (10.7)	4 (3.9)	9 (7.1)	2 (3.9)	5 (5.5)	10 (9.5)
Fisher's exact p-value = 0.07				= 0.3			= 0.3	

IQGAP1	MGMT-3 level			PTEN	
3 level	<7	7-67%	>67%	0	1
0	1 (0.5)	0 (0.9)	1 (0.5)	1 (0.7)	1 (1.3)
1	3 (2.4)	2 (4.3)	4 (2.4)	3 (3.3)	6 (5.7)
2	6 (7.1)	16 (12.8)	5 (7.1)	10 (9.9)	17 (17.1)
Fisher's exact p-value = 0.1				= 1.0	

	MGMT-3 level		
PTEN	<7	7-67%	>67%
0	3 (3.7)	7 (6.6)	4 (3.7)
1	7 (6.3)	11 (11.4)	6 (6.3)
Fisher's exact p-value = 0.9			

Age in	PTEN		IGFBP2-3 level		
quartiles	0	1	0	1	2
<43	7 (3.7)	3 (6.3)	8 (4.2)	0 (1.8)	2 (3.9)
43-51	5 (3.7)	5 (6.3)	6 (4.2)	2 (1.8)	2 (3.9)
52-54	2 (3.7)	8 (6.3)	1 (4.2)	2 (1.8)	7 (3.9)
>54	0 (2.9)	8 (5.1)	1 (3.4)	3 (1.5)	4 (3.2)
Fisher's exact p-value = 0.008			= 0.007		

frequency
(expected frequency)

The relationship between the IHC variables was similar to those with the original LTSs. However the relationship between IGFBP2 and IQGAP1 was no longer significant (p-value = 0.07). The association between the IHC scores and age in quartiles is also similar with only PTEN and IGFBP2 having a significant association, but now the relationship between PTEN and age is highly significant with a p-value for Fisher's exact test of 0.008.

3.4. Finding the best model

Appendix Table 4: Assessing the fit of the models using the Likelihood ratio test and the unweighted residual sum of squares test (RSS).

	Likelihood ratio test			RSS test	
Variables modelled	-2 log likelihood	Ch^2 (df)	P	z	P
IGFBP2,IQGAP1,MGMT, PTEN, age	18.83	-	-		
IGFBP2,MGMT,PTEN,age	18.87	0.04 (1)	0.84		
IGFBP2,MGMT,PTEN	20.45	1.58 (1)	0.21		
IGFBP2,MGMT	21.59	1.14 (1)	0.29	0.31	0.76
IGFBP2	28.08	6.49 (1)	0.01		

The final model with only IGFBP2 and MGMT was the same as with the original LTSs. Appendix Table 5 gives the Odds Ratios, 95% confidence interval, coefficient and P-values for this final model. The effect of each increase in IGFBP2 IHC score was slightly greater than with the originals, giving a 90% decrease in the odds of becoming a LTS while the effect of having 7-67% compared to <7% or >67% compared to 7-67% nuclei staining for MGMT was slightly smaller than with the original LTSs giving around a 6

fold increase in the odds of becoming a LTS. The p-values were less significant than with the original LTSs probably due to reduced power with smaller numbers.

Appendix Table 5: Parsimonious logistic regression model for LTS or STS, with IGFBP2 and MGMT expression as predictors.

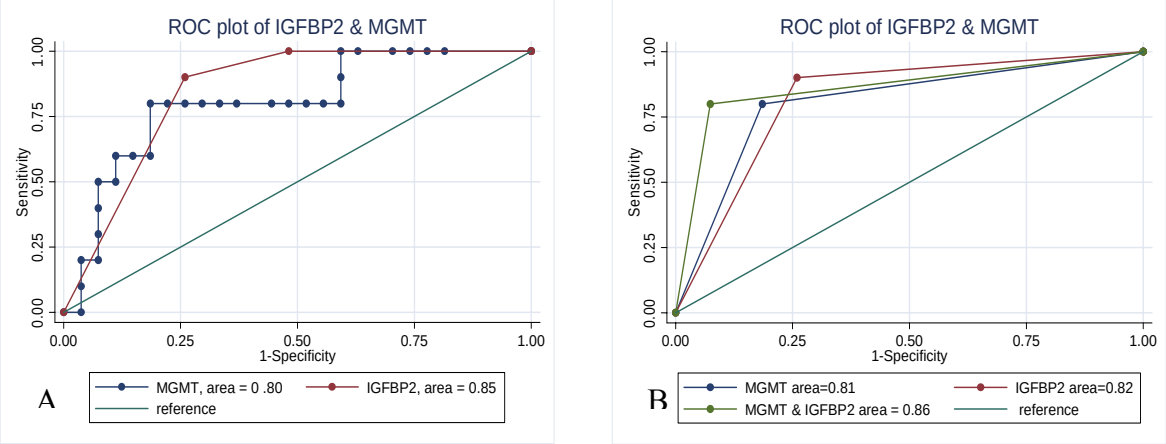
	Odds ratio	95% CI	coeff	P
IGFBP2 -3 levels:0&1, 2, 3	0.096	0.013-0.71	-2.34	0.02
MGMT -3 levels: <7%, 7-67%, >67%)	6.4	1.2-34	1.85	0.03
c-statistic = 0.93				

3.5. ROC analysis

Appendix Figure 3 shows ROC plots for IGFBP2 and for MGMT expression without the original LTSs with IGFBP2 with three levels and MGMT scores in their original form also and categorised as binary variables at the best value for predicting LTS. IGFBP2 had slightly better area under the ROC curve (AUC) than with the original LTSs in both forms. The variable formed by multiplying the binary IGFBP2 and MGMT variables also gave a slightly better AUC than with the original long term survivors: 0.86 compared to 0.81.

Sensitivity using both IGFBP2 and MGMT to predict who will become a LTS has improved quite a bit without the original LTSs. See Appendix Table 6. IGFBP2 has sensitivity of 90% both as a 3 level and as a binary variable, up from 85% with the original LTSs. The sensitivity of MGMT has also improved slightly and the combined IGFBP2 and MGMT binary variable has improved sensitivity from 70% to 80%.

Specificity is unchanged by the removal of the original LTSs. IGFBP2 is still the best at predicting who will become a LTS using sensitivity.



Appendix Figure 3: ROC plots showing sensitivity and 1-specificity of the classification of patients as LTSs or STSs based on their IGFBP2 and MGMT immunohistochemistry. Plot A shows all levels of MGMT and 3 levels of IGFBP2. The values for IGFBP2 are in reverse order to allow them to be plotted on the same graph. Plot B has MGMT categorised into <52% or ≥52% stained nuclei and IGFBP2 into 0 or 1 compared to 2 or 3, again in reverse order.

Appendix Table 6: Sensitivity and specificity and % correctly predicted, achieved using IGFBP2 and MGMT and a combined variable obtained by multiplying a binary variable from IGFBP2 (scores of 0 &1 versus scores of 2 & 3) by a binary MGMT variable (less than or more than 52% staining).

	Sensitivity	Specificity	% correct
IGFBP2 -3 levels: 0&1, 2, 3	90.0%	75.0%	79.0%
IGFBP2 -binary: 0&1, 2&3	90.0%	75.0%	79.0%
MGMT -3 levels: <7%, 7-67%, >67%	60.0%	85.7%	79.0%
MGMT -binary: <52%, ≥52%	80.0%	81.5%	81.1%
IGFBP2 (bin) & MGMT (bin)	80.0%	92.6%	89.2%

3.6. Survival Analysis .

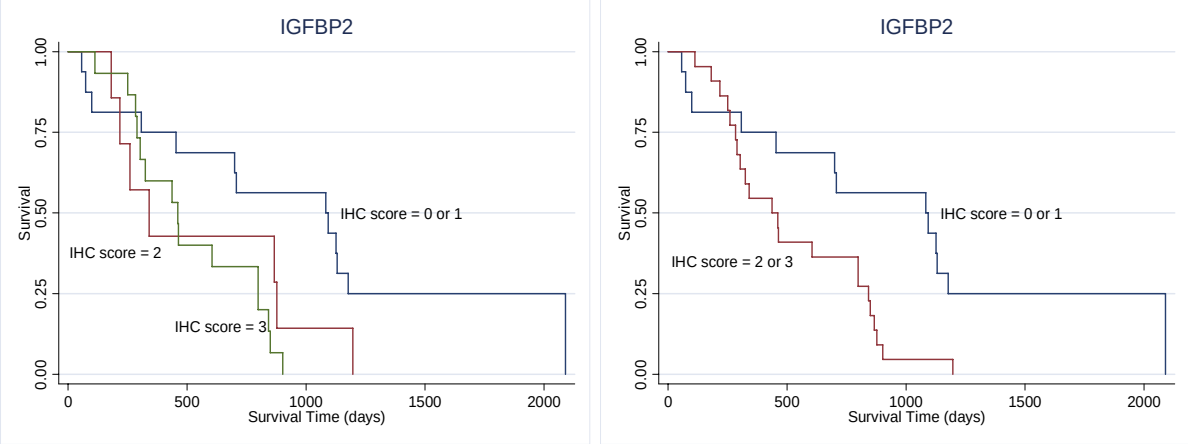
3.7: Exploratory Analysis

Appendix Table 7: Results of exploratory survival analysis.

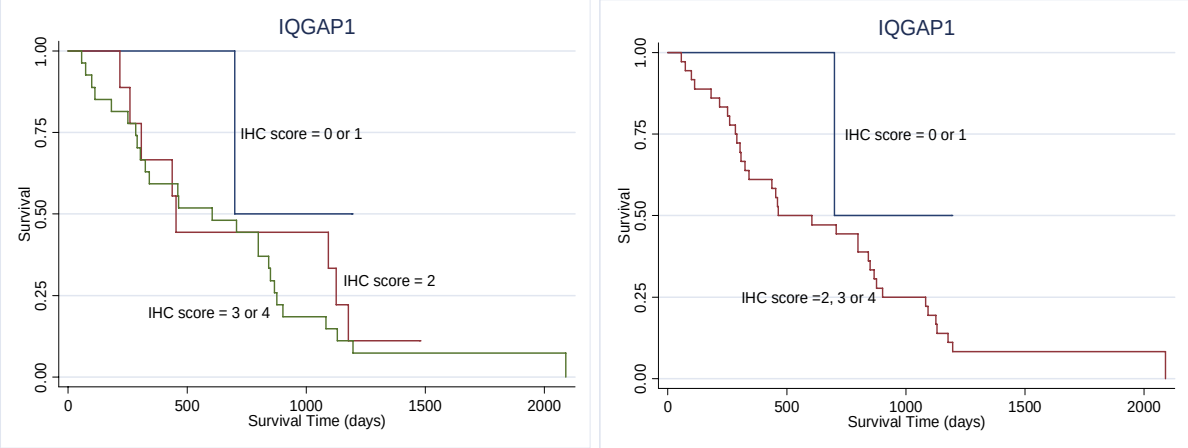
Variable	Level	No.	incidence rate/day	Logrank test Chi^2 (df)	P
IGFBP2 (3 levels)	0, 1	16	0.0009	8.16 (2)	0.017
	2	7	0.0018		
	3	15	0.0019		
IGFBP2 (binary)	0, 1	16	0.0093	7.37 (1)	0.007
	2, 3	22	0.0019		(0.03)#
IQGAP1 (3 levels)	0, 1	2	0.0053	1.56 (2)	0.46
	2	9	0.0012		
	3, 4	27	0.0015		
IQGAP1 (binary)	0, 1	2	0.0053	1.15 (1)	0.28
	2, 3, 4	36	0.0014		(1.0)#
MGMT (in quartiles)	< 7%	10	0.0031	17.23 (3)	0.0006
	7-39%	9	0.0012		
	40-67%	9	0.0015		
	>67%	10	0.0009		
MGMT (3 levels)	<7%	10	0.0031	20.68 (1)	0.0002
	7-67%	18	0.0013		(0.0008)#
	>67%	9	0.0009		
PTEN	0	14	0.0011	2.57 (1)	0.11
	1	24	0.0016		
Gender	Female	13	0.0016	1.09 (1)	0.30
	Male	25	0.0013		
Age (in quartiles)	<43	10	0.0010	6.60 (3)	0.086
	43-51	10	0.0012		
	52-54	10	0.0016		
	>54	8	0.0023		

p-value with modified Bonferroni correction in groups merged to maximise trend.

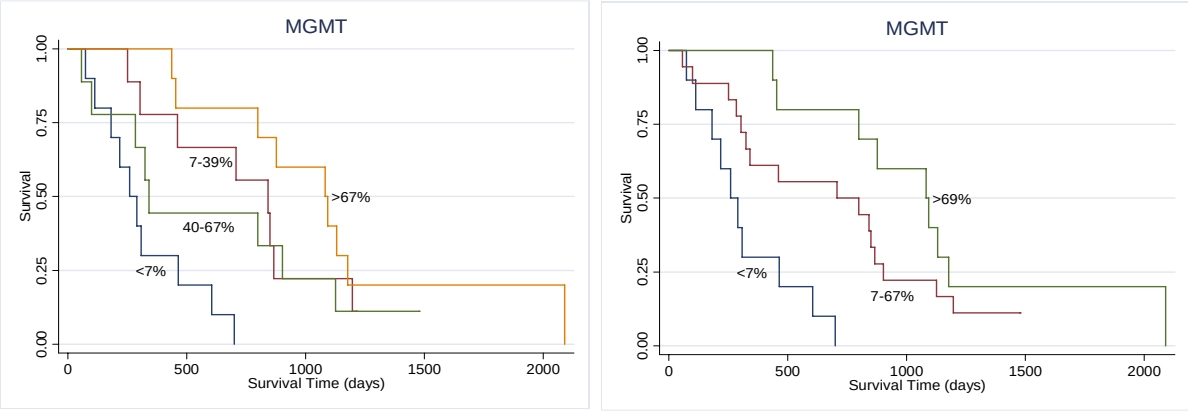
Without the original LTSs IGFBP2 became a more significant predictor of survival by the logrank test, but still showed worse survival in the first year after diagnosis with low IGFBP2 scores. There were only two subjects remaining with IQGAP1 scores of 0 or 1. See Appendix Table 7 and Figure 5.



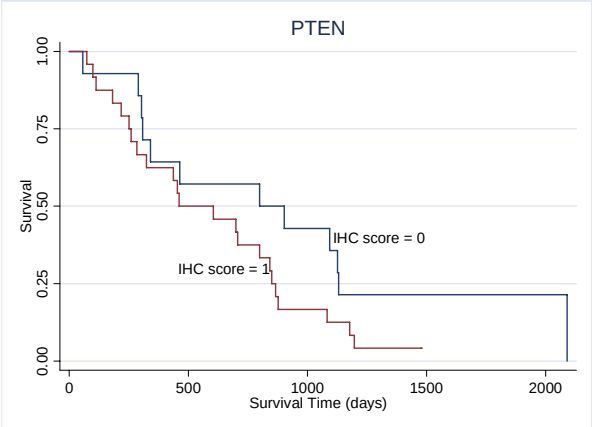
Appendix Figure 4: Kaplan Meier Curves for IGFBP2 with 3 or 2 levels of IHC score.

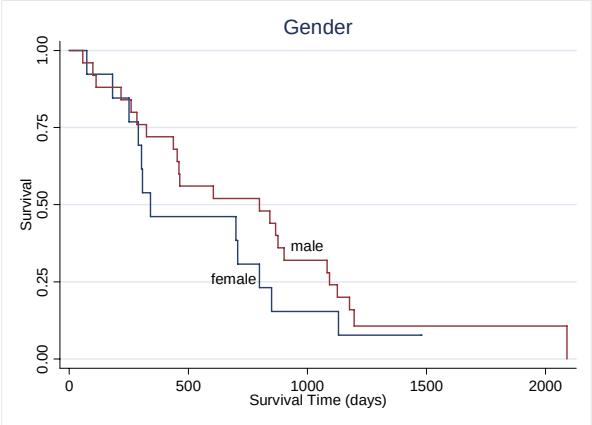


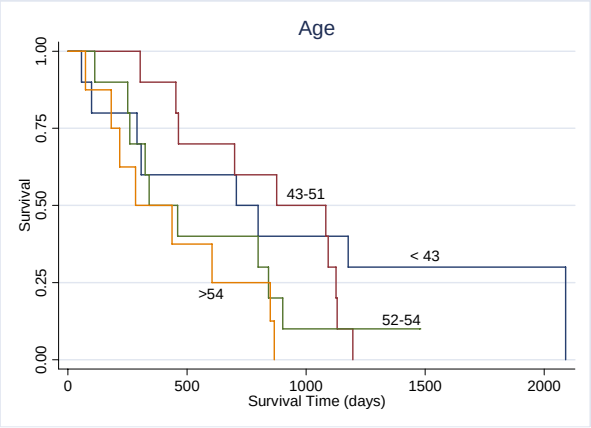
Appendix Figure 5: Kaplan Meier Curves for IQGAP1 with 3 or 2 levels of IHC score.



Appendix Figure 6: Kaplan Meier curves for MGMT percentage stained nuclei, divided at quartiles or divided into3 levels at <7%, 7-67% or >67%.







Appendix Figure 7: Kaplan Meier curves for PTEN IHC score

Appendix Figure 8: Kaplan Meier curves for gender and age.

3.8. Univariate Analysis using Cox Regression

Survival analysis was carried out using Cox regression. The results are given in Appendix Table 8. IGFBP2 has been analysed as a binary variable and IQGAP1 and MGMT with three levels as in the analysis with the three originals.

Compared to the results with the original LTSs, IGFBP2 was more significant and MGMT was slightly less significant. Each increase in IGFBP2 expression as measured by a binary IHC score resulted in an almost 3 fold increase in the risk of death while having 7-67% nuclei staining for MGMT compared to <7% or >67% compared to 7-67% resulted in a 60% fall in the risk of death.

3.9. Fitting the baseline multivariate model

All variables with $P < 0.25$ in univariate analysis were included in the baseline model. These were IGFBP2 (binary), IQGAP1 (3 levels), and MGMT (3 levels), PTEN and age.

Appendix Table 8: Hazard Ratios and P-values using predictors of survival time in univariate models compared to the multivariate baseline model.

	Univariate		Baseline multivar model		
	Hazard Ratio	P	Hazard Ratio	P	VIF
IGFBP2-binary: 0&1, 2&3	2.78	0.008	1.77	0.30	1.81
IQGAP1-3 levels: 0&1, 2, 3&4	1.47	0.23	1.17	0.71	1.23
MGMT– 3 levels: <7%, 7-67, >67%	0.37	0.001	0.40	0.005	1.11
PTEN	1.80	0.11	1.72	0.22	1.40
Age	1.03	0.20	0.98	0.47	1.85
Log likelihood			-91.14		

The Hazard Ratios and Ps for each predictor were compared between the univariate and multivariate models to check for confounding. The results are shown in Table 8. The predictor with the greatest change in the size of the Hazard Ratio was in IGFBP2 which changed by 36%, followed by IQGAP1 which changed by 20%. This indicated confounding of IQGAP1 by IGFBP2 and of IGFBP2 by MGMT due to a significant

relationship between the IQGAP1 and IGBP2 in this form (P = 0.45). The VIF are larger than before but still not large enough to indicate serious collinearity.

3.10. Finding the best model

To find the most parsimonious model, the least significant variable was removed one at a time as before. The results are shown in Table 9. As with the dataset with the original three long term survivors, the final model had only MGMT as a predictor of survival time. Compared to patients with <7% nuclei stained for MGMT, those with 7-67% staining had a 60% reduction in the instantaneous risk of death from GBM and a similar reduction was seen in patients with >67% staining compared to 7-67%.

Appendix Table 9: Assessing the fit of the models using the Likelihood ratio test compared to the previous model, and global test of the Proportional Hazards Assumption using Schoenfeld Residuals.

	Likelihood ratio test			Test of global proportional hazards assumption	
Variables modelled	-2 log likelihood	Ch^2 (df)	P	Ch^2 (df)	P
IGFBP2,IQGAP1,MGMT,PTEN,age	182.29	-	-	5.07 (5)	0.41
IGFBP2,MGMT,PTEN, age	182.44	0.15 (1)	0.70	3.63 (4)	0.46
IGFBP2,MGMT,PTEN	182.95	0.51 (1)	0.48	2.64 (3)	0.45
IGFBP2,MGMT	183.88	0.93 (1)	0.33	2.71 (2)	0.26
MGMT	187.11	3.23 (1)	0.07	1.11 (1)	0.29

Appendix Table 10: Final Cox regression model for LTS or STS with IGFBP2 and MGMT expression as predictors.

	Hazard ratio	95% CI	P

MGMT– 3 levels: <7%, 7-67%, >67%	0.37	0.21-0.67	0.001 (0.004)#

p-value with modified Bonferroni correction in groups merged to maximise trend.