On-Line Appendix

Additional methodology regarding the patient cohorts

Three different databases were used to identify the patients used in this study:

(a) A database of consecutive patients presenting to North Bristol NHS Trust, UK was used to identify patients with an ultimate diagnosis of MPE, as confirmed independently by 2 respiratory consultants. From this database, two separate cohorts were extracted:
    a. The 'UK Cohort 1' comprised of 221 patients with MPE entered into the database between March 2008 and April 2012. All patients were followed up for a minimum of 12 months or until death.
    b. The 'UK Cohort 2' comprised of 83 patients with MPE entered into the database between May 2012 and April 2013. Patients were followed up for a minimum of 6 months or until death.

(b) The 'Dutch cohort' was collected from the Netherlands Cancer Institute between June 2006 and May 2012. It comprised of 315 consecutive patients with proven cancer and MPE.

(c) The 'Australian cohort' was obtained from the 'Longitudinal follow up of management of MPE in Western Australia' Study conducted between August 2009 and October 2011, and comprised of 253 patients. This study prospectively recorded patient data from the time of diagnosis of MPE for a minimum of 6 months or until death.

For the purposes of survival analysis, the date of diagnosis was defined using the date pleural malignancy was confirmed by pleural biopsy or cytology. In the absence of histo-cytological confirmation from the pleura, the date of diagnosis was defined by the date that the consistent radiology was performed.

197 of 221 patients in 'UK Cohort 1' died during the study follow up. The remaining 24 patients were censored after a median follow up of 774 days (IQR 473-1058).

52 of 83 patients in 'UK Cohort 2' died during the study follow up. The remaining 31 patients were censored after a median follow up of 220 days (IQR 284-388).

286 of 315 patients in 'Dutch cohort' died during the study follow up. The remaining 29 patients were censored after a median follow up of 941 days (IQR 314-1043).

195 of 253 patients in the 'Australian Cohort' died during the study follow up. The remaining 58 patients were censored after a median follow up of 708 days (IQR 476-922).

Additional methodology regarding sample processing for VEGF and NT-proBNP

For the 'Bristol Cohort' plasma (collected in EDTA), serum and pleural fluid (collected in EDTA and serum gel tubes), were collected at the time of study entry. These samples were centrifuged at 1000G for 20mins and the supernatant frozen at -70^{o}C for the purpose of subsequent analysis.

Pleural fluid VEGF quantification

Commercially available enzyme linked immunoassay (ELISA) kits were used to quantify VEGF levels in the pleural fluid samples (Quantikine © for human VEGF, R&D Systems, Abingdon, UK). Samples were processed in duplicate according to manufacturer's guidelines. The accurate measurement range as determined by the standard solutions is 31.2-2000pg/mL and samples were diluted if necessary.

Serum NT-proBNP quantification

Serum samples were utilized to measure NT-proBNP levels. This was performed using an automated, quantitative Enzyme Linked Fluorescent Assay (ELFA) technique, according to the manufacturers' guidelines (VIDAS ® NT-proBNP, bioMerieux SA, France). The accurate measurement range of the assay is 20-25,000 pg/mL.

A priori statistical analysis plan

Background

The 'pleural investigation trial' is a prospective observational series consisting of consecutive patients who presented to North Bristol NHS Trust with an undiagnosed pleural effusion. Written informed consent was taken from all patients included in the study, following which blood and pleural fluid samples (at the time of trial entry) were stored as per the study protocol. All patients were followed up for a minimum of 6 months or until death. A final diagnosis as to the cause of their effusion was documented.

This series includes the 221 patients who were enrolled into the study between March 2008 and April 2012, who had a final diagnosis of malignant pleural effusion (MPE). The diagnosis of MPE was confirmed independently by 2 respiratory consultants, with at least one of the following criteria having to be present:

- Histocytological proof of pleural malignancy on a pleural biopsy or cytology;
- Histological proof of malignancy elsewhere with a presumed malignant effusion;
- Radiological diagnosis of pleural malignancy without histocytological proof

Analysis Plan

Descriptive statistics will be used to summarize patient characteristics and clinical data. The appropriate statistical tests will be selected to compare the groups and look for correlations depending on whether the data is parametric or non-parametric. For the purposes of analysis of continuous variables, a linear relationship will be assumed in the first instance.

The overall survival of patients will be calculated in days from the date of trial entry to the date of death. Initially, each of the following potential prognostic factors will be evaluated individually using a univariable Cox model and Kaplan Meier Survival Curves to determine if they are independently associated with survival.

Patient/Tumor variables:

Variable	**Variable type**	**% missing data**
Age	Continuous	0
Baseline ECOG performance Score	Categorical (0-4)	5
Malignant cell type	Categorical (lung, mesothelioma, breast/ gynecological, other)	0

Comorbidities that may exacerbate MPE:

Variable	**Variable type**	**% missing data**
Baseline serum albumin level	Continuous	2
Baseline eGFR	Analyze as categories: <15, 15-29, 30-59, 60-89, >90	0.5
Serum BNP level at baseline	Continuous	10

Inflammation based variables:

Variable	**Variable type**	**% missing data**

Neutrophil to lymphocyte ratio in the blood	Continuous	1
Modified Glasgow Prognostic Score (mGPS)	Categorical (0-2)	5
Platelet to lymphocyte ratio (PLR)	Continuous	1
Baseline serum CRP	Continuous	3.1
Baseline pleural fluid VEGF	Continuous	8.5

Pleural Fluid characteristics:

Variable	**Variable type**	**% missing data**
Size of the effusion at baseline on CXR	Categorical (1-5)	4
Pleural fluid LDH	Continuous	3.1
Pleural fluid pH and glucose	Categorical (High or Low) (Patients will be divided into 2 categories according to their pleural fluid pH (<7.3; ≥7.3). If no pleural fluid pH value is available, but the pleural fluid glucose has been measured, those with a pleural fluid glucose <3.3mmol/l will be included in the 'low' group and those with a value ≥3.3mmol/l will be included in the 'high' group)	10

In order to determine which variables are still prognostic, even when used in conjunction with other variables, multivariable analysis will be used to assess their adjusted contribution on survival. Only variables with ≤10% missing data will be included in the multivariable analysis. Multiple imputation will be used to account for missing baseline data.

The multivariable analysis will be performed using a stepwise backward Cox proportional hazard model. Variables with a p value >0.1 will be removed from the model.

Depending on these results, we will then attempt to make a prognostic model if it makes sense to do so. For continuous variables that are being considered for inclusion in the prognostic model, the linearity assumption will be checked to see if a more complex relationship is more appropriate.

Statistical software will be used to perform the analyses. Statistical significance will be taken at the 5% level.

This plan was agreed by B Kahan, A Clive and N Maskell on February 4th 2013

NB. In order to combine the data with the Dutch and Australian cohorts, it was felt that defining survival from date of diagnosis to death would be more appropriate than trial entry to death. Analysing the data with this definition did not substantially alter the findings.

Table: Characteristics of the 221 patients in the 'UK Cohort 1'

Variable			**Result**	**Missing Data, n**
Age, Median (IQR)			74 (67-80)	0
Effusion Size, n (%)	1	Blunting of costophrenic angle	9 (4)	8
	2	More than blunting but less than 25% of hemithorax occupied by pleural fluid	22 (10)	
	3	Pleural fluid occupying 25% to 50% of hemithorax	90 (42)	
	4	Pleural fluid occupying 50% to 75% of hemithorax	61 (29)	
	5	Pleural fluid occupying more than 75% of hemithorax	31 (15)	
ECOG Performance Score, n (%)	0		24 (11)	9
	1		91 (43)	
	2		67 (32)	
	3		27 (13)	
	4		3 (1)	
Malignant cell type, n (%)	Mesothelioma		58 (26)	0
	Lung cancer		66 (30)	
	Breast/Gynaecological cancer		46 (21)	
	Other		51 (23)	
Serum Albumin (g/l), median (IQR)			33 (28-36)	4
Serum eGFR, n (%)	<15		1 (1)	0
	15-29		5 (2)	
	30-59		51 (23)	
	60-89		103 (47)	
	>90		61 (28)	
Serum NT-proBNP (pg/ml) Median (IQR)			282 (112-1001)	22
Serum NLR, median (IQR)	Absolute neutrophil count divided by the absolute lymphocyte count		5.15 (3.56-7.35)	2
Modified Glasgow Prognostic Score, n (%)	0	CRP ≤10mg/l and albumin ≥35g/l	47 (22)	11
	1	CRP >10mg/l and albumin <35g/l, or CRP ≤10mg/l and albumin < 35g/l	48 (23)	
	2	CRP >10mg/l and Albumin <35g/l	115 (55)	
Serum PLR, median (IQR)	Absolute platelet count divided by the absolute lymphocyte count		255.7 (178.3-360.0)	2
Serum CRP (mg/L), Median (IQR)			30 (11-83)	7
Pleural fluid VEGF (pg/mL), Median (IQR)			2251.5 (831.3-5722)	19
Pleural fluid LDH (IU/L), Median (IQR)			627.5 (352-1164)	7
Pleural fluid pH/glucose score, n (%)	High	Pleural fluid pH≥ 7.3 or if pH unavailable, pleural fluid glucose ≥3.3mmol/l	145 (74)	24
	Low	Pleural fluid pH <7.3 or if pH unavailable, pleural fluid glucose <3.3mmol/l	52 (26)	

Definition of abbreviations: HR = Hazard Ration; ECOG Performance Score = Eastern Cooperative Oncology Group Performance Score; NLR = Neutrophil-to-Lymphocyte Ratio; LDH = Lactate Dehydrogenase; IU/L = International units per litre; 95% CI = 95% Confidence Interval; IQR = interquartile range; VEGF = Vascular Endothelial Growth Factor; eGFR= estimated glomerular filtration rate; NT-proBNP = N-terminal prohormone of brain natriuretic peptide; CRP = C Reactive Protein; PLR = Platelet-to-Lymphocyte Ratio. * Variables with '-' were dropped from analysis due to non-significance (ie $p>0.1$)

Results of univariable analysis of the multiply imputed 'UK Cohort 1' dataset

Variable		Results of univariable analysis			Results of multivariable analysis		
		Hazard Ratio	**95% CI**	***p* value**	**Hazard Ratio**	**95% CI**	***p* value**
Age		1.03	1.01-1.04	0.0013	-	-	-
Effusion Size	1	1.01	0.49-2.09	0.974	1.01	0.47-2.19	0.977
	2	1.59	0.98-2.58	0.061	1.53	0.91-2.56	0.106
	3	-	-	-	-	-	-
	4	0.71	0.50-1.00	0.048	0.69	0.48-1.00	0.051
	5	0.92	0.59-1.43	0.714	0.78	0.49-1.24	0.302
ECOG Performance score	0	-	-	-	-	-	-
	1	1.31	0.78-2.18	0.304	1.42	0.84-2.39	0.192
	2	2.63	1.56-4.43	0.000	2.35	1.34-4.12	0.003
	3	3.41	1.88-6.20	0.000	2.67	1.39-5.10	0.003
	4	4.60	1.35-13.72	0.015	4.52	1.26-16.22	0.021
Malignant cell type	Mesothelioma	-	-	-	-	-	-
	Lung cancer	2.27	1.56-3.30	<0.001	1.46	0.97-2.20	0.072
	Breast/Gynecological cancer	0.87	0.57-1.34	0.529	0.79	0.50-1.25	0.314
	Other	1.32	0.88-1.96	0.178	0.96	0.62-1.48	0.858
Serum Albumin (g/l)		0.94	0.91-0.96	<0.001	-	-	-
Serum eGFR	<15	3.24	0.44-23.70	0.247	-	-	-
	15-29	1.39	0.56-3.50	0.479	-	-	-
	30-59	1.04	0.70-1.54	0.844	-	-	-
	60-89	0.82	0.58-1.15	0.248	-	-	-
	>90	-	-	-	-	-	-
Serum NT-proBNP/100 (pg/ml)***		1.01	1.01-1.01	<0.001	1.01	1.00-1.01	0.033
Serum NLR		1.08	1.05-1.12	<0.001	1.07	1.03-1.11	0.000

mGPS	0	-	-	-	-	-	-
	1	0.91	0.59-1.41	0.685	-	-	-
	2	1.66	1.15-2.38	0.006	-	-	-
Serum PLR		1.00	1.00-1.00	0.0201	-	-	-
Serum CRP (mg/L)		1.00	1.00-1.01	0.0024	-	-	-
Pleural fluid VEGF (pg/mL)		1.00	1.00-1.00	0.4130	-	-	-
Pleural fluid LDH/100 (IU/L)**		1.01	1.00-1.03	0.0368	1.02	1.00-1.03	0.010
Pleural fluid pH/glucose score		0.96	0.70-1.32	0.8000	-	-	-

Definition of abbreviations: ECOG Performance Score = Eastern Cooperative Oncology Group Performance Score; NLR = Neutrophil-to-Lymphocyte Ratio; LDH = Lactate Dehydrogenase; IU/L = International units per liter; 95% CI = 95% Confidence Interval; IQR = interquartile range; VEGF = Vascular Endothelial Growth Factor; eGFR= estimated glomerular filtration rate; NT-proBNP = N-terminal prohormone of brain natriuretic peptide; CRP = C Reactive Protein; PLR = Platelet-to-Lymphocyte Ratio;

* Variables with '-' were dropped from analysis due to non-significance

** The Hazard Ratios and 95% Confidence Intervals for pleural fluid LDH are for a 100IU/L increase in LDH.

*** The Hazard Ratios and 95% Confidence Intervals for NT-proBNP are for a 100pg/ml increase in NT-proBNP.

Table: Characteristics of the 83 patients in the 'UK Cohort 2' and the results of multivariable analysis

Variable		**Result**	**Missing Data**	**Results of multivariable analysis**		
				Hazard Ratio	**95% CI**	***p* value**
ECOG Performance Score, n (%)	0	13 (16)	2	-	-	-
	1	32 (40)		3.42	0.92-12.73	0.066
	2	21 (26)		2.18	0.57-8.34	0.256
	3	15 (19)		4.65	1.20-18.11	0.027
	4	0 (0)		-	-	-
Malignant cell type, n (%)	Mesothelioma	17 (21)	1	-	-	-
	Lung cancer	33 (40)		3.09	1.10-8.69	0.033
	Breast/Gynaecological cancer	15 (18)		2.83	0.88-9.14	0.081
	Other	17 (21)		3.62	1.09-12.02	0.036
Serum NLR, median (IQR)		4.97 (3.33-7.95)	1	1.09	1.03-1.15	0.004
Pleural fluid LDH (IU/L), median (IQR)		620.5 (352.5-1153)	3	1.02*	0.99-1.04*	0.163

Definition of abbreviations: ECOG Performance Score = Eastern Cooperative Oncology Group Performance Score; NLR = Neutrophil-to-Lymphocyte Ratio; LDH = Lactate Dehydrogenase; IU/L = International units per litre; 95% CI = 95% Confidence Interval.

* The Hazard Ratios and 95% Confidence Intervals for pleural fluid LDH are for a 100IU/L increase in LDH.

Sensitivity analysis of LENT Score using the 'UK Cohort 1'

SCORE	ASSUMING BEST CASE						ASSUMING WORST CASE					
	n	HR (95% CI)	Median (IQR) survival (days)	Survived to 6 months – n (%)	Survived to 3 months- n(%)	Survived to 1 month- n(%)	n	HR (95% CI)	Median (IQR) survival (days)	Survived to 6 months, n (%)	Survived to 3 months- n(%)	Survived to 1 month- n(%)
LOW RISK (score 0-2)	49	-	319 (236-549)	42 (86)	48 (98)	49 (100)	44	-	319 (228-549)	38 (86)	43 (98)	44 (100)
MODERATE RISK (score 3-4)	140	1.61 (1.13-2.28)	125 (41-427)	62 (44)	81 (58)	112 (80)	138	1.50 (1.04-2.17)	160 (48-443)	65 (47)	85 (62)	114 (83)
HIGH RISK (score 5-7)	32	5.33 (3.27-8.67)	44 (22-77)	2 (6)	5 (16)	21 (66)	39	5.46 (3.41-8.77)	43 (22-77)	3 (8)	6 (15)	24 (62)
Harrell's C	0.6921						0.706					
AUC for 1 month mortality	0.76 (0.69-0.83)						0.77 (0.71-0.84)					
AUC for 3 months mortality	0.84 (0.79-0.89)						0.83 (0.78-0.88)					
AUC for 6 month's mortality	0.83 (0.78-0.89)						0.83 (0.78-0.89)					

Definition of abbreviations: HR= Hazard Ratio; 95% CI = 95% Confidence Interval; IQR = interquartile range; AUC = Area Under Curve

NA * Unable to provide data as insufficient patients have died.

Sensitivity analysis of LENT Score using the 'UK Cohort 2'

SCORE	ASSUMING BEST CASE						ASSUMING WORST CASE					
	n	HR (95% CI)	Median (IQR) survival (days)	Survived to 6 months – n (%)	Survived to 3 months- n(%)	Survived to 1 month- n(%)	n	HR (95% CI)	Median (IQR) survival (days)	Survived to 6 months, n (%)	Survived to 3 months- n(%)	Survived to 1 month- n(%)
LOW RISK (score 0-2)	15	-	347 (336-NA*)	13 (87)	14 (93)	15 (100)	14	-	NA* (347-NA*)	12 (86)	13 (93)	14 (100)
MODERATE RISK (score 3-4)	50	2.50 (0.97-6.45)	198 (80-NA*)	29 (58)	33 (66)	47 (94)	49	2.74 (0.96-7.81)	249 (81-NA*)	29 (59)	33 (67)	46 (94)
HIGH RISK (score 5-7)	18	8.67 (3.10-24.21)	51 (24-119)	3 (17)	6 (33)	13 (72)	20	9.97 (3.30-30.17)	51 (24-119)	4 (20)	7 (35)	15 (75)
Harrell's C	0.7385						0.7446					
AUC for 1 month mortality	0.80 (0.67-0.93)						0.77 (0.64-0.91)					
AUC for 3 months mortality	0.75 (0.64-0.85)						0.73 (0.63-0.84)					
AUC for 6 month's mortality	0.80 (0.70-0.89)						0.78 (0.68-0.88)					

Definition of abbreviations: HR= Hazard Ratio; 95% CI = 95% Confidence Interval; IQR = interquartile range; AUC = Area Under Curve

NA * Unable to provide data as insufficient patients have died.

Sensitivity analysis of LENT Score including only patients with a histocytological diagnosis of malignancy (using 'UK Cohort 1')

SCORE						
	n	HR (95% CI)	Median (IQR) survival (days)	Survived to 6 months – n (%)	Survived to 3 months- n(%)	Survived to 1 month- n(%)
LOW RISK (score 0-2)	44	-	319 (228-549)	38 (86)	43 (98)	44 (100)
MODERATE RISK (score 3-4)	125	1.41 (0.97-2.04)	181 (63-467)	63 (50)	82 (66)	104 (83)
HIGH RISK (score 5-7)	33	6.20 (3.76-10.23)	43 (22-60)	2 (6)	4 (12)	20 (61)
Harrell's C	0.7127					
AUC for 1 month mortality	0.79 (0.72-0.85)					
AUC for 3 months mortality	0.85 (0.80-0.90)					
AUC for 6 month's mortality	0.84 (0.79-0.89)					
Definition of abbreviations: HR= Hazard Ratio; 95% CI = 95% Confidence Interval; IQR = interquartile range; AUC = Area Under Curve						

Sensitivity analysis of LENT Score excluding patients with mesothelioma (using 'UK Cohort 1')

SCORE						
	n	HR (95% CI)	Median (IQR) survival (days)	Survived to 6 months – n (%)	Survived to 3 months- n(%)	Survived to 1 month- n(%)
LOW RISK (score 0-2)	6	-	484 (228-NA*)	6 (100)	6 (100)	6 (100)
MODERATE RISK (score 3-4)	118	2.46 (0.90-6.70)	130 (47-467)	53 (45)	70 (59)	97 (82)
HIGH RISK (score 5-7)	39	7.83 (2.73-22.44)	43 (22-77)	3 (8)	6 (15)	24 (62)
Harrell's C	0.69					
AUC for 1 month mortality	0.69 (0.60-0.78)					
AUC for 3 months mortality	0.76 (0.70-0.83)					
AUC for 6 months mortality	0.81 (0.74-0.87)					
Definition of abbreviations: HR= Hazard Ratio; 95% CI = 95% Confidence Interval; IQR = interquartile range; AUC = Area Under Curve						

Sensitivity analysis evaluating only patients with a new diagnosis of malignancy (using 'UK Cohort 1')

SCORE						
	n	HR (95% CI)	Median (IQR) survival (days)	Survived to 6 months – n (%)	Survived to 3 months- n(%)	Survived to 1 month- n(%)
LOW RISK (score 0-2)	42	-	319 (239-549)	36 (86)	41 (98)	42 (100)
MODERATE RISK (score 3-4)	89	1.56 (1.05-2.31)	168 (47-443)	42 (47)	54 (61)	74 (83)
HIGH RISK (score 5-7)	24	5.97 (3.45-10.33)	33 (13-57)	2 (8)	3 (13)	13 (54)
Harrell's C	0.7264					
AUC for 1 month mortality	0.83 (0.76-0.89)					
AUC for 3 months mortality	0.88 (0.82-0.93)					
AUC for 6 month's mortality	0.84 (0.78-0.90)					
Definition of abbreviations: HR= Hazard Ratio; 95% CI = 95% Confidence Interval; IQR = interquartile range; AUC = Area Under Curve						