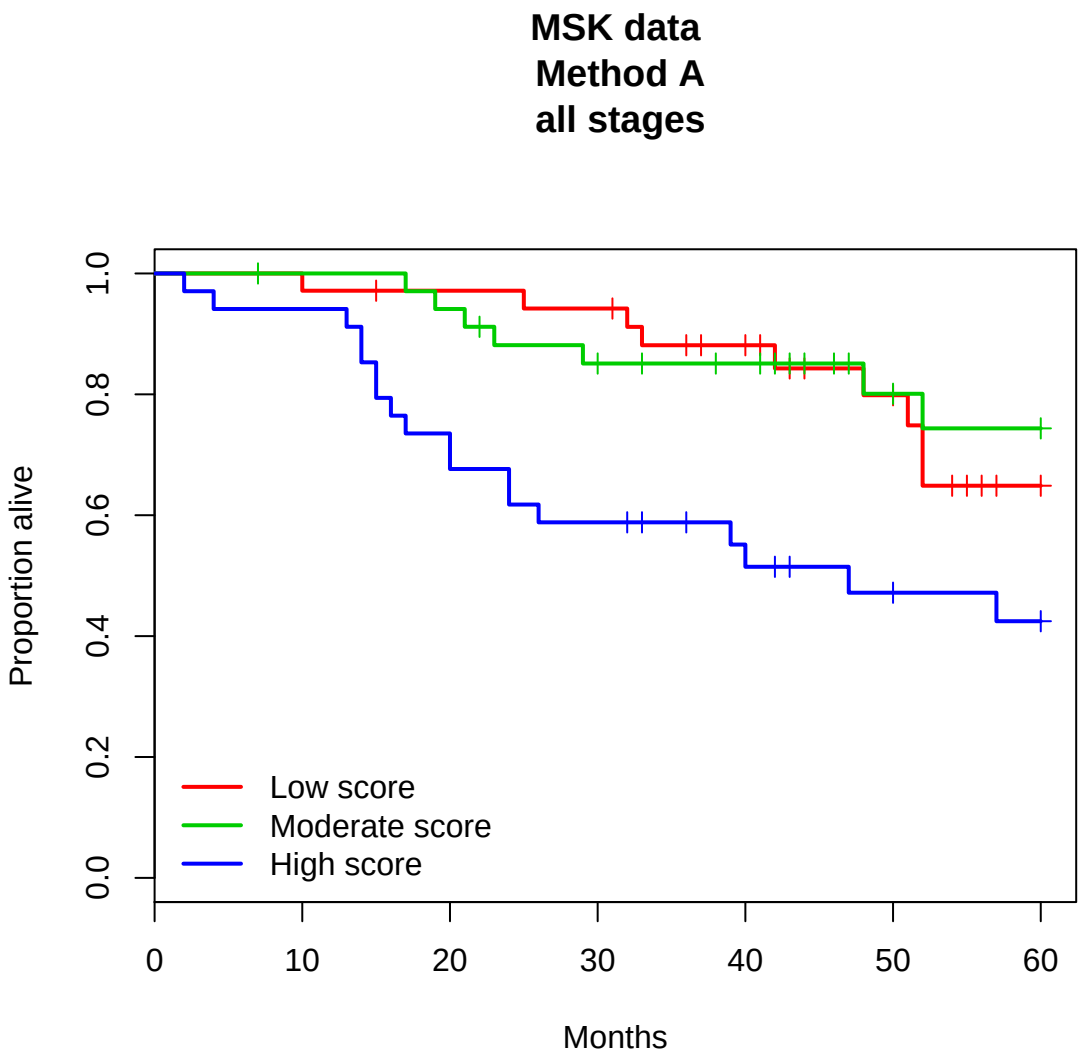
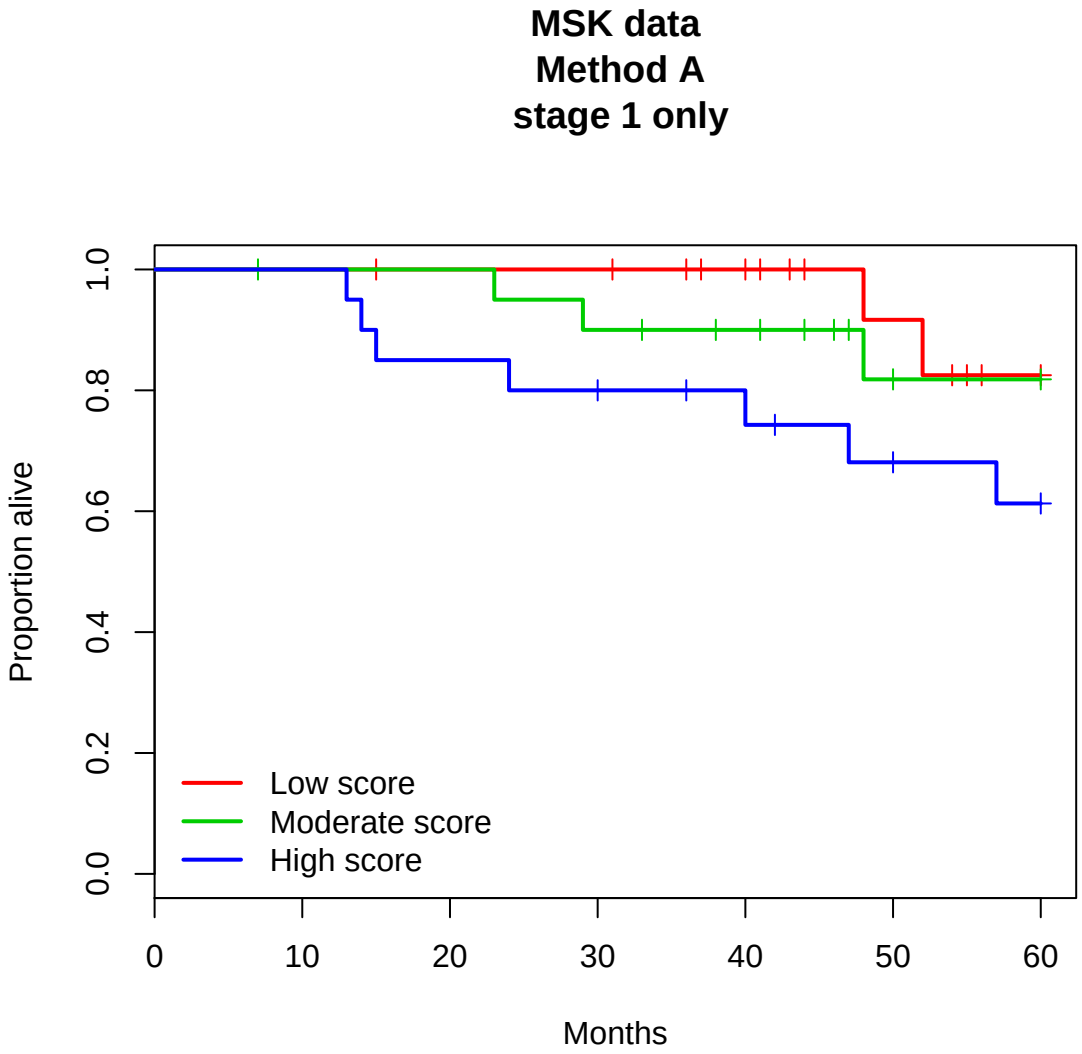
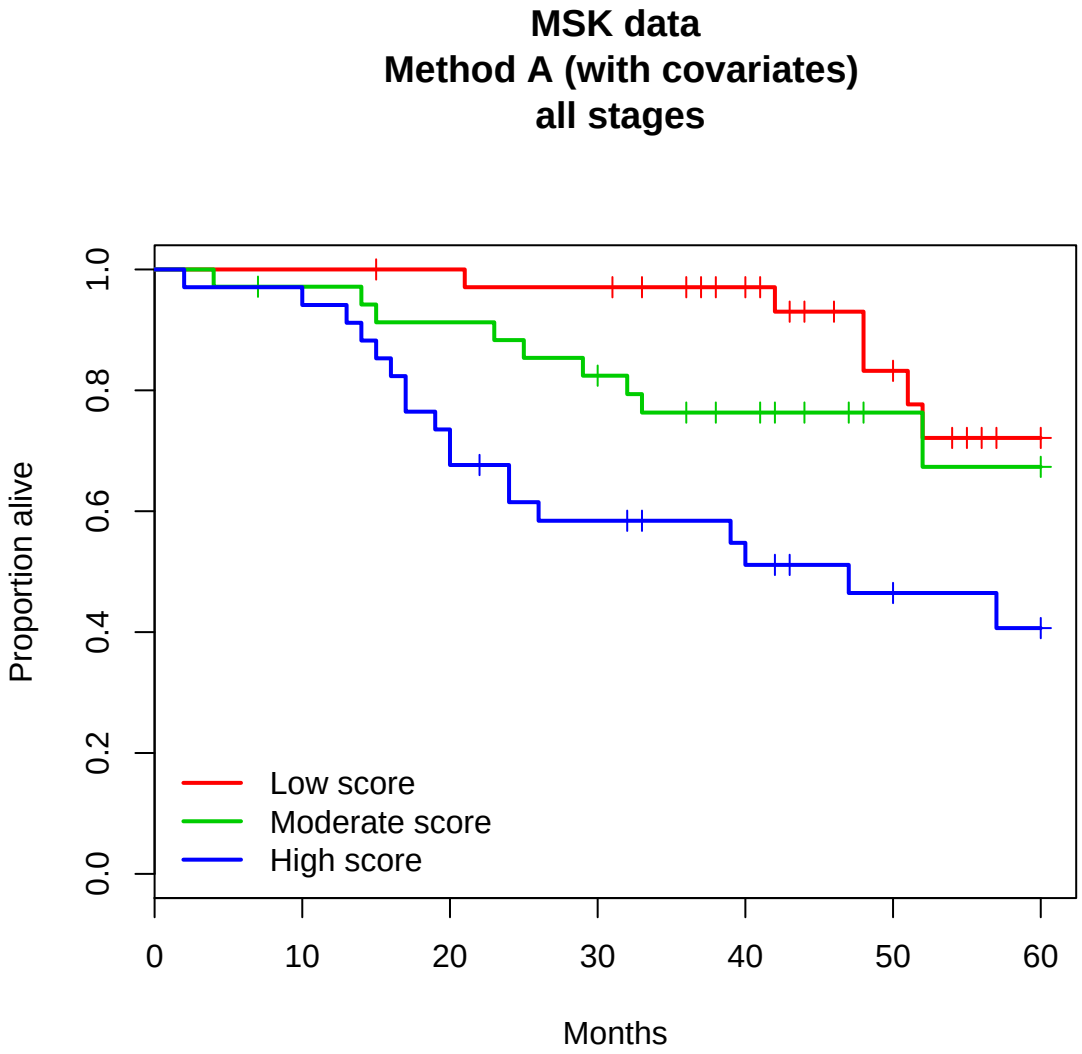
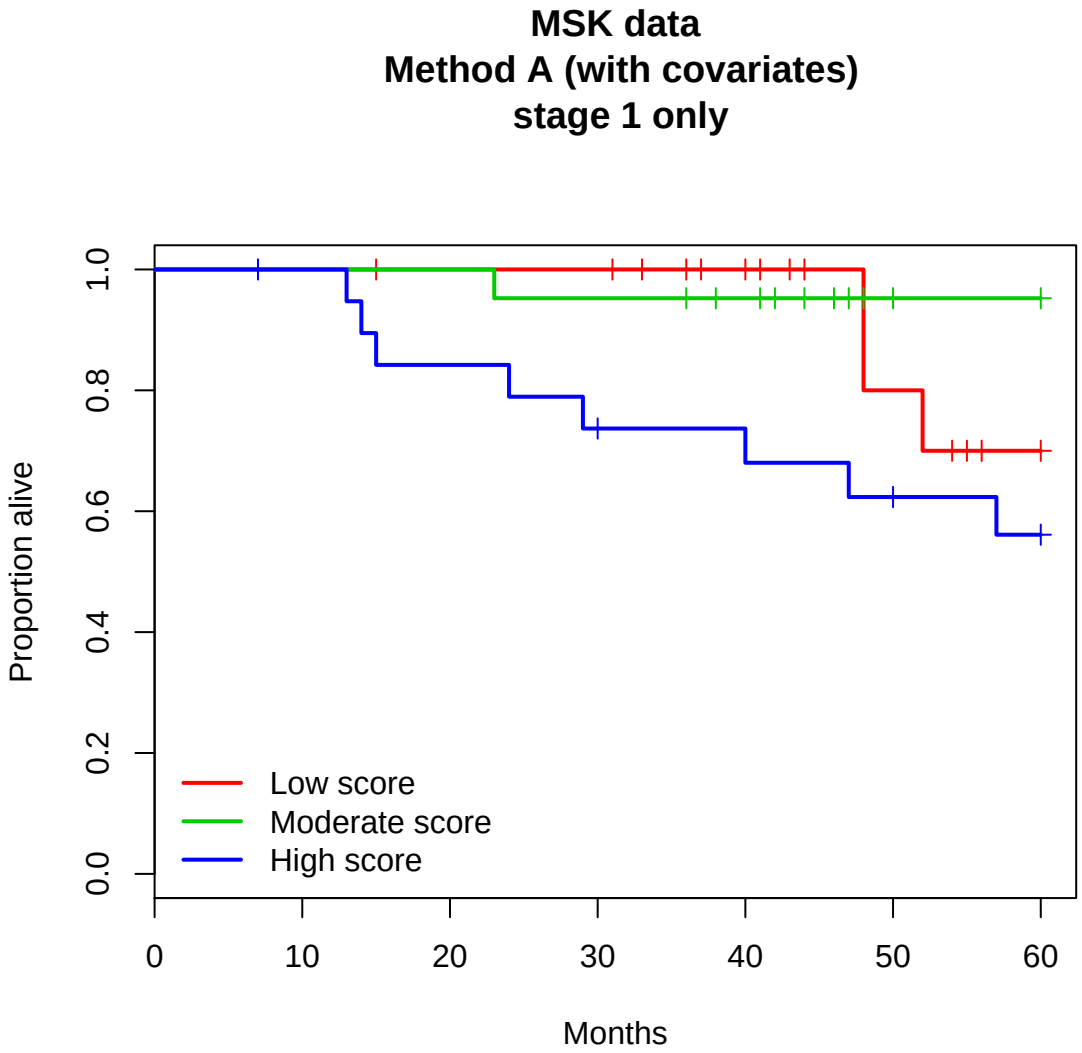
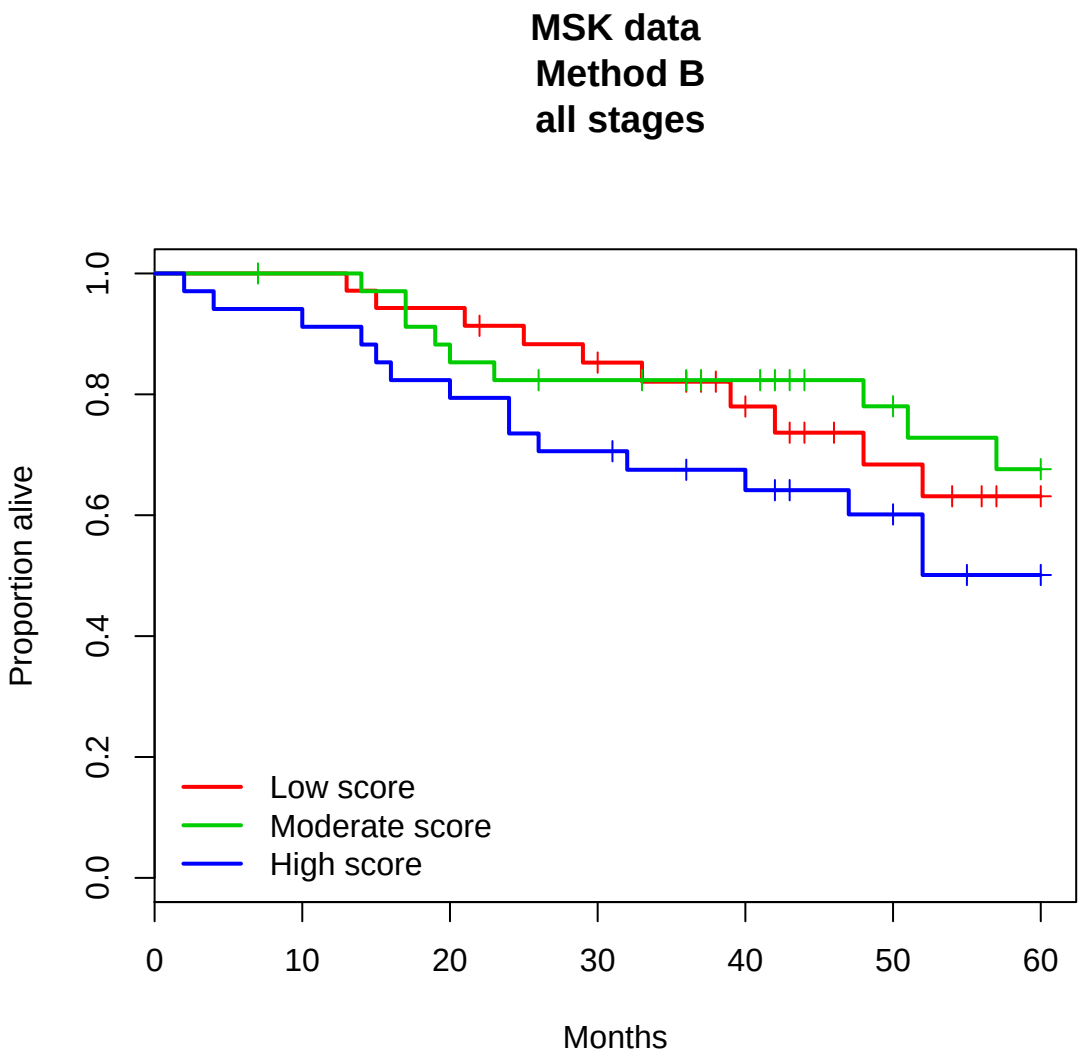
Supp fig. 1

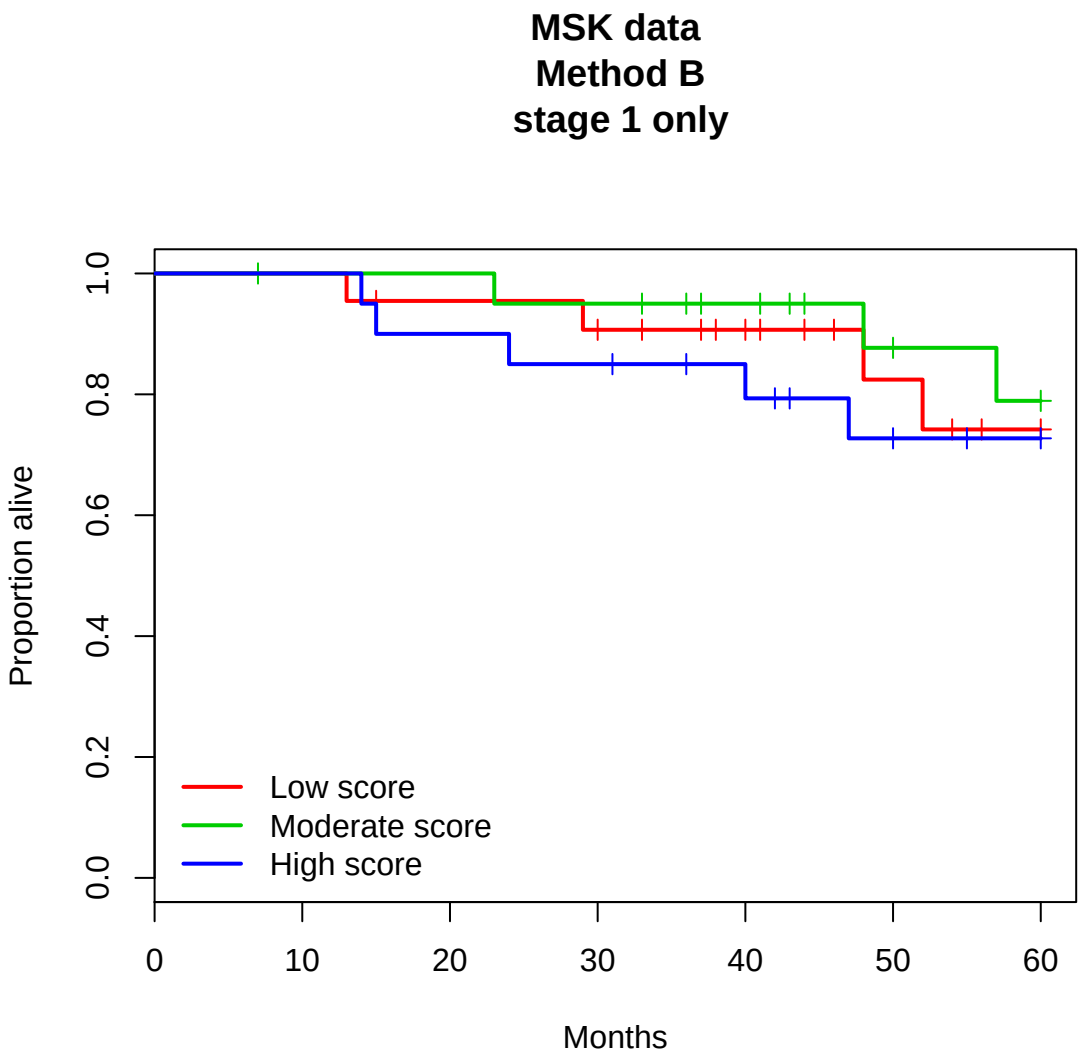

MSK data
Method A
all stages
Proportion alive
0.0
0.2
0.4
0.6
0.8
1.0
0
10
20
30
40
50
60
Months
Low score
Moderate score
High score

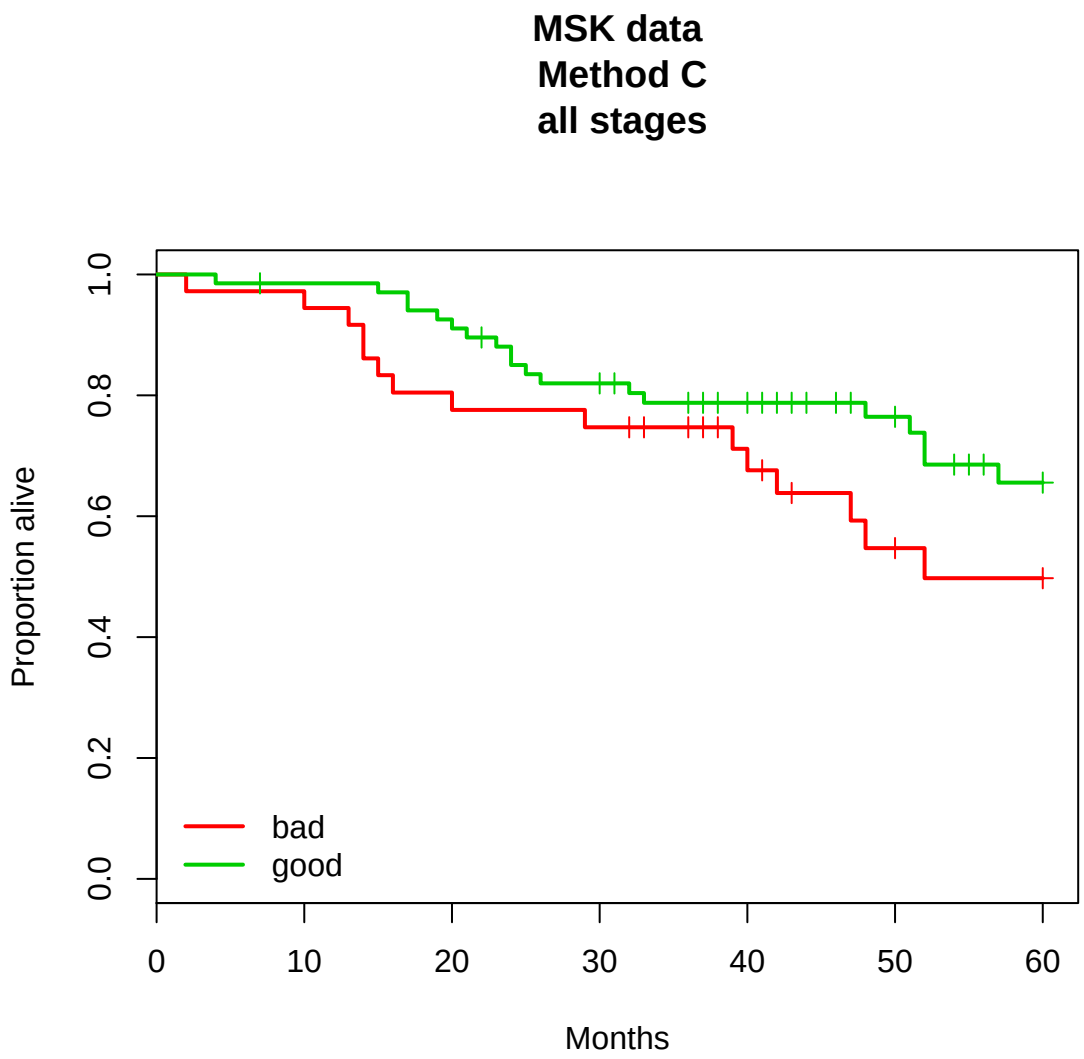

MSK data
Method A
stage 1 only
Proportion alive
0.0
0.2
0.4
0.6
0.8
1.0
0
10
20
30
40
50
60
Months
Low score
Moderate score
High score

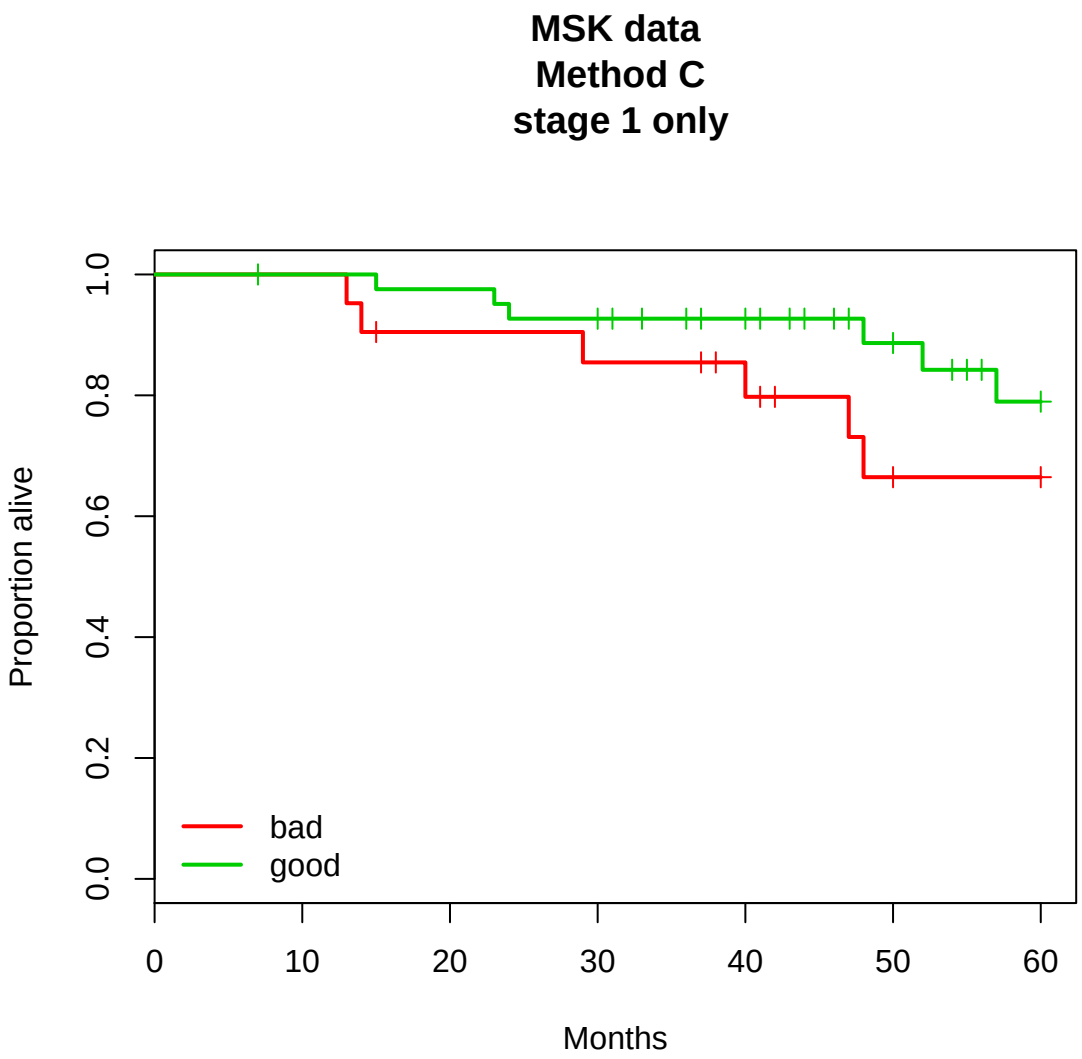

MSK data
Method A (with covariates)
all stages
Proportion alive
0.0
0.2
0.4
0.6
0.8
1.0
0
10
20
30
40
50
60
Months
Low score
Moderate score
High score

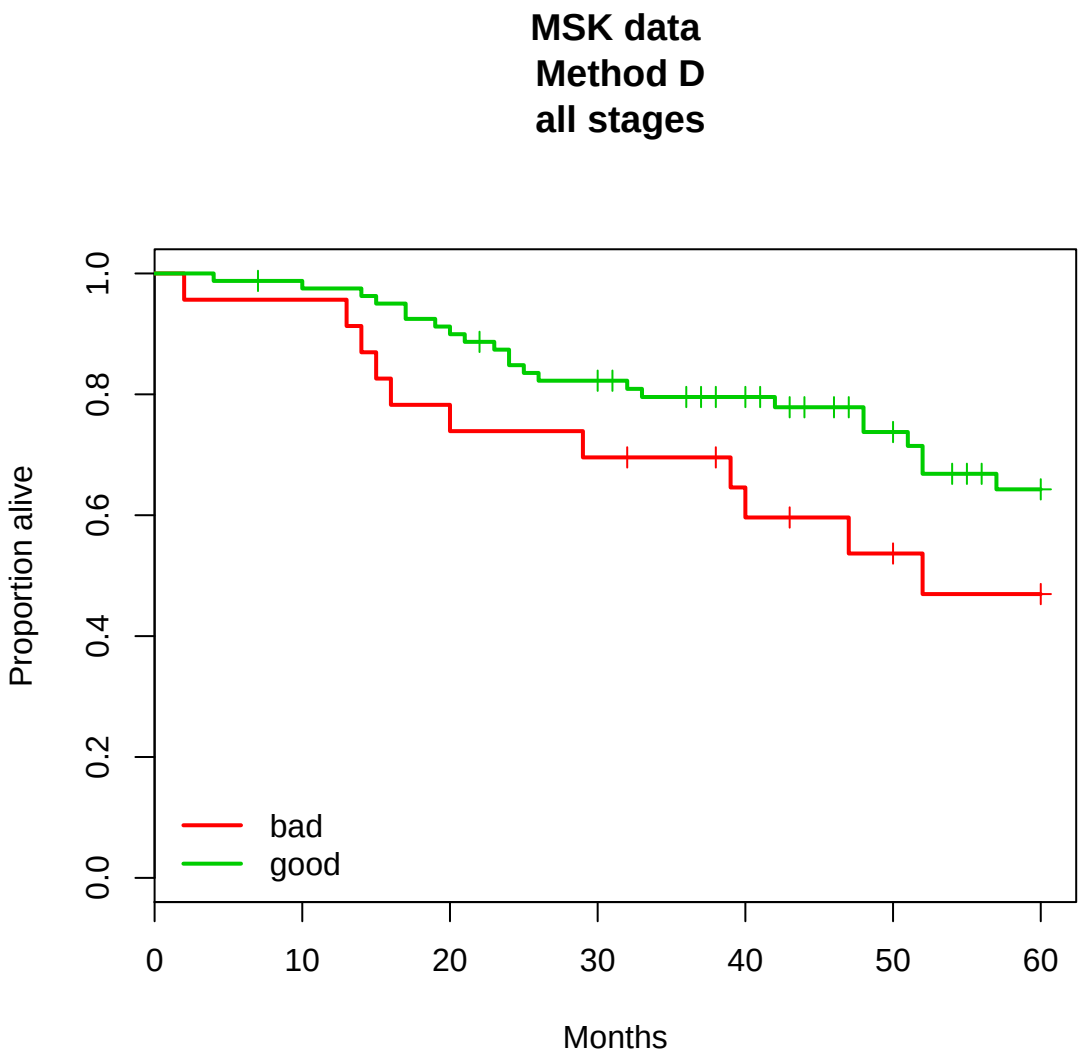

MSK data
Method A (with covariates)
stage 1 only
Proportion alive
0.0
0.2
0.4
0.6
0.8
1.0
0
10
20
30
40
50
60
Months
Low score
Moderate score
High score

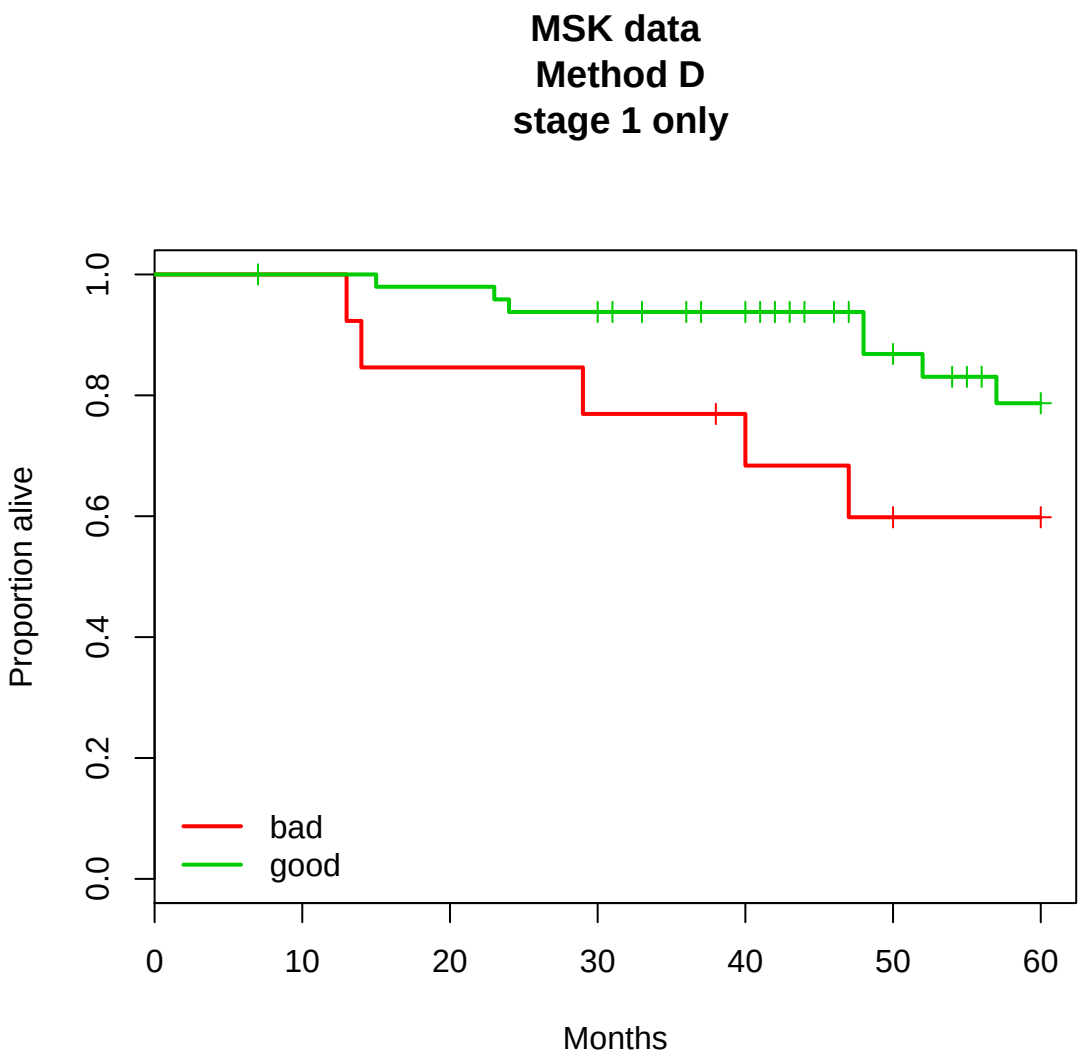

MSK data
Method B
all stages
Proportion alive
0.0
0.2
0.4
0.6
0.8
1.0
0
10
20
30
40
50
60
Months
Low score
Moderate score
High score

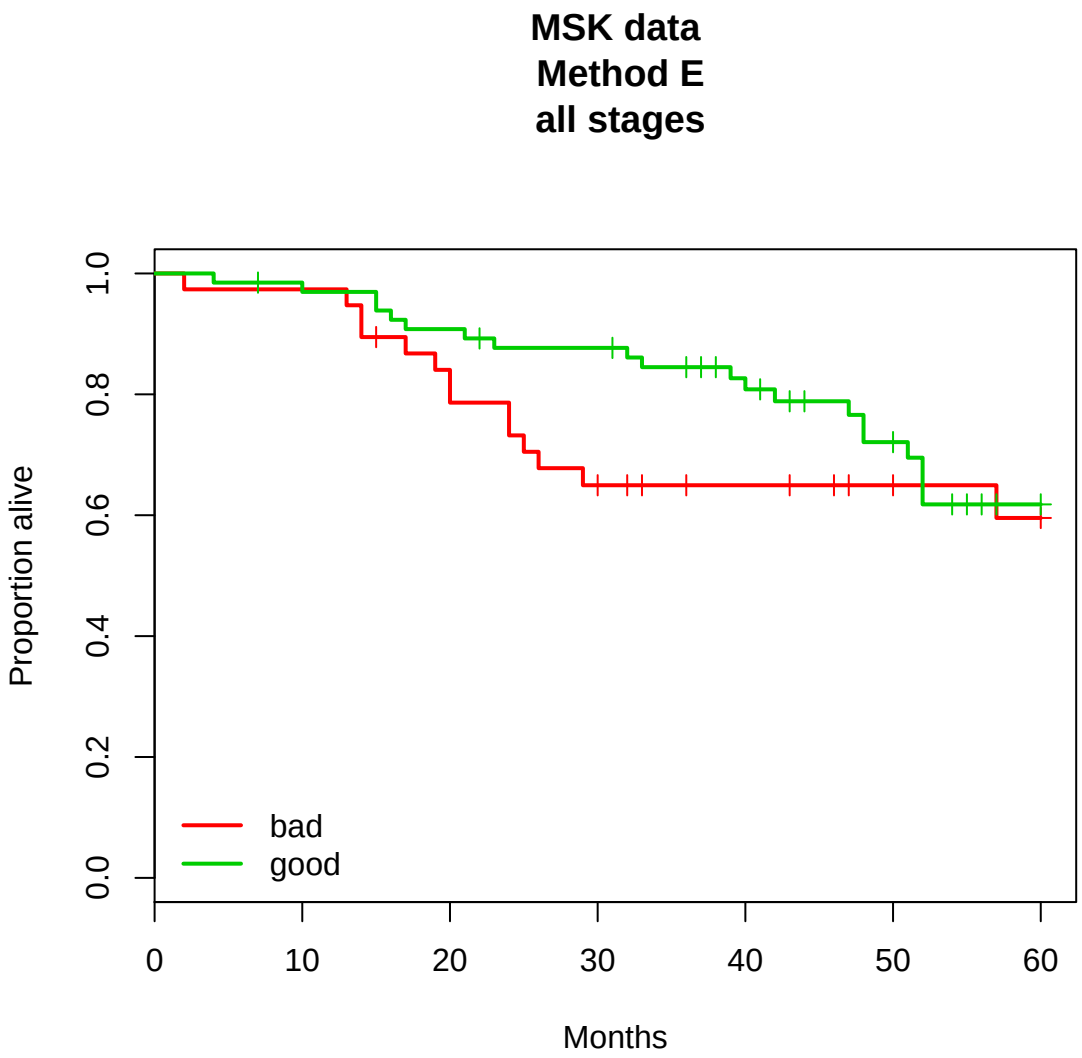

MSK data
Method B
stage 1 only
Proportion alive
0.0
0.2
0.4
0.6
0.8
1.0
0
10
20
30
40
50
60
Months
Low score
Moderate score
High score

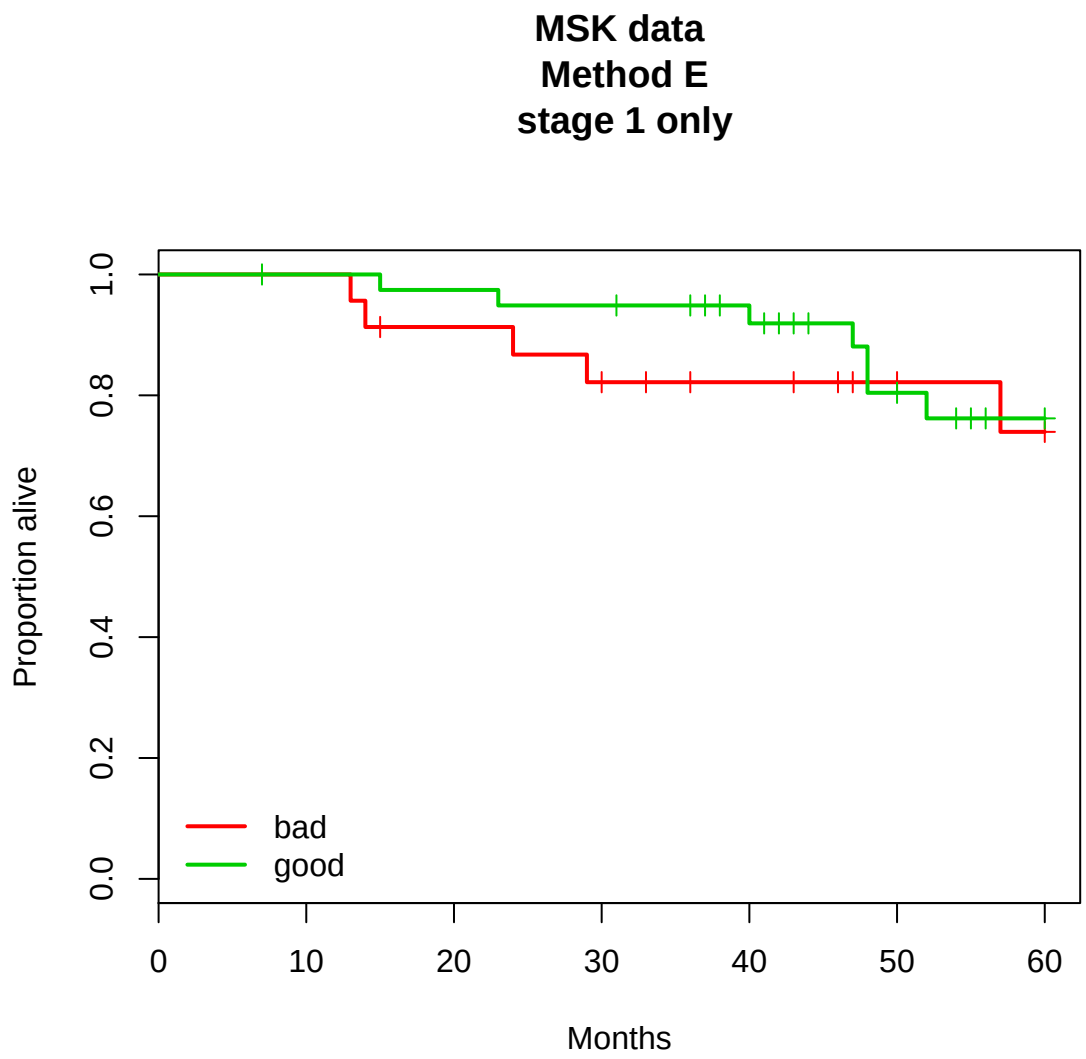

MSK data
Method C
all stages
Proportion alive
0.0
0.2
0.4
0.6
0.8
1.0
0
10
20
30
40
50
60
Months
bad
good

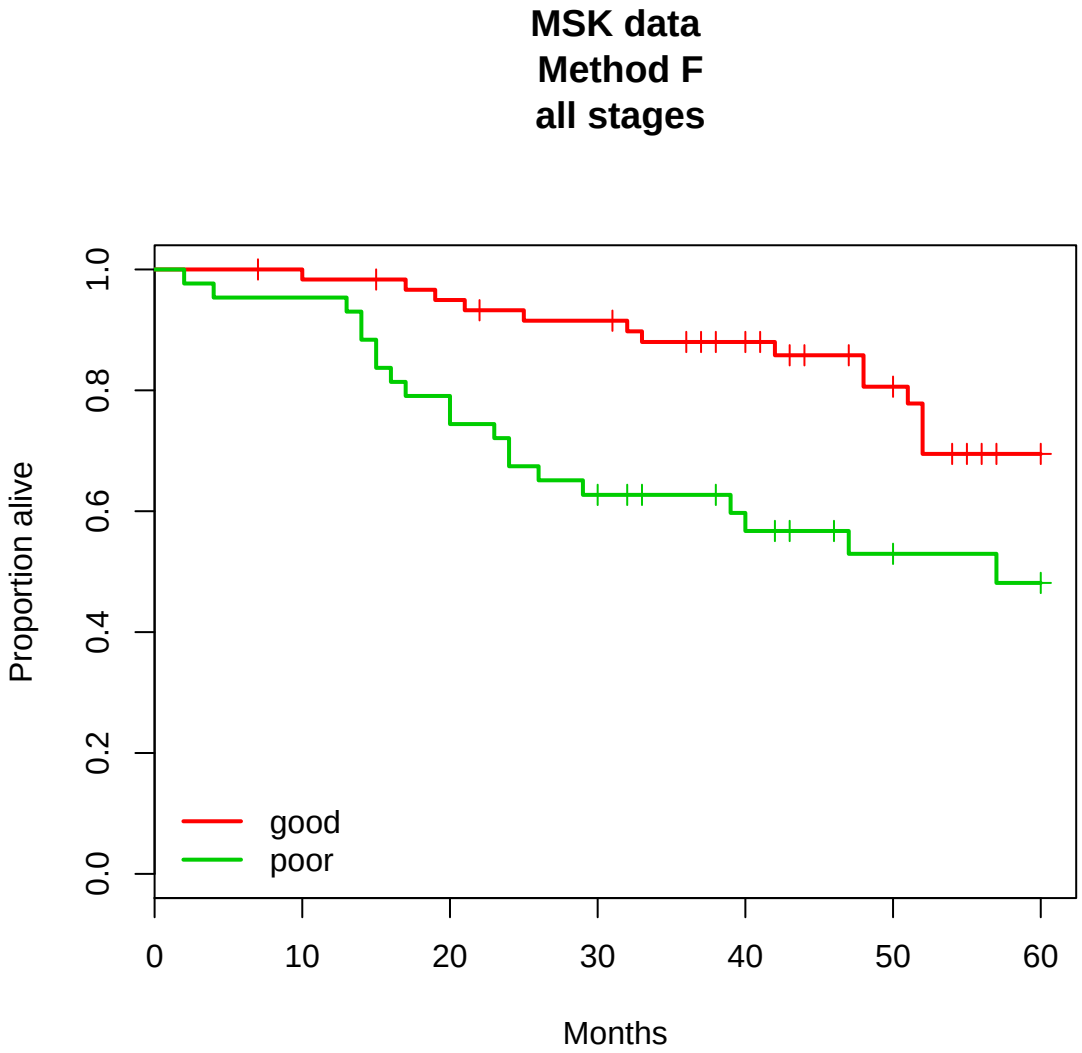

MSK data
Method C
stage 1 only
Proportion alive
0.0
0.2
0.4
0.6
0.8
1.0
0
10
20
30
40
50
60
Months
bad
good

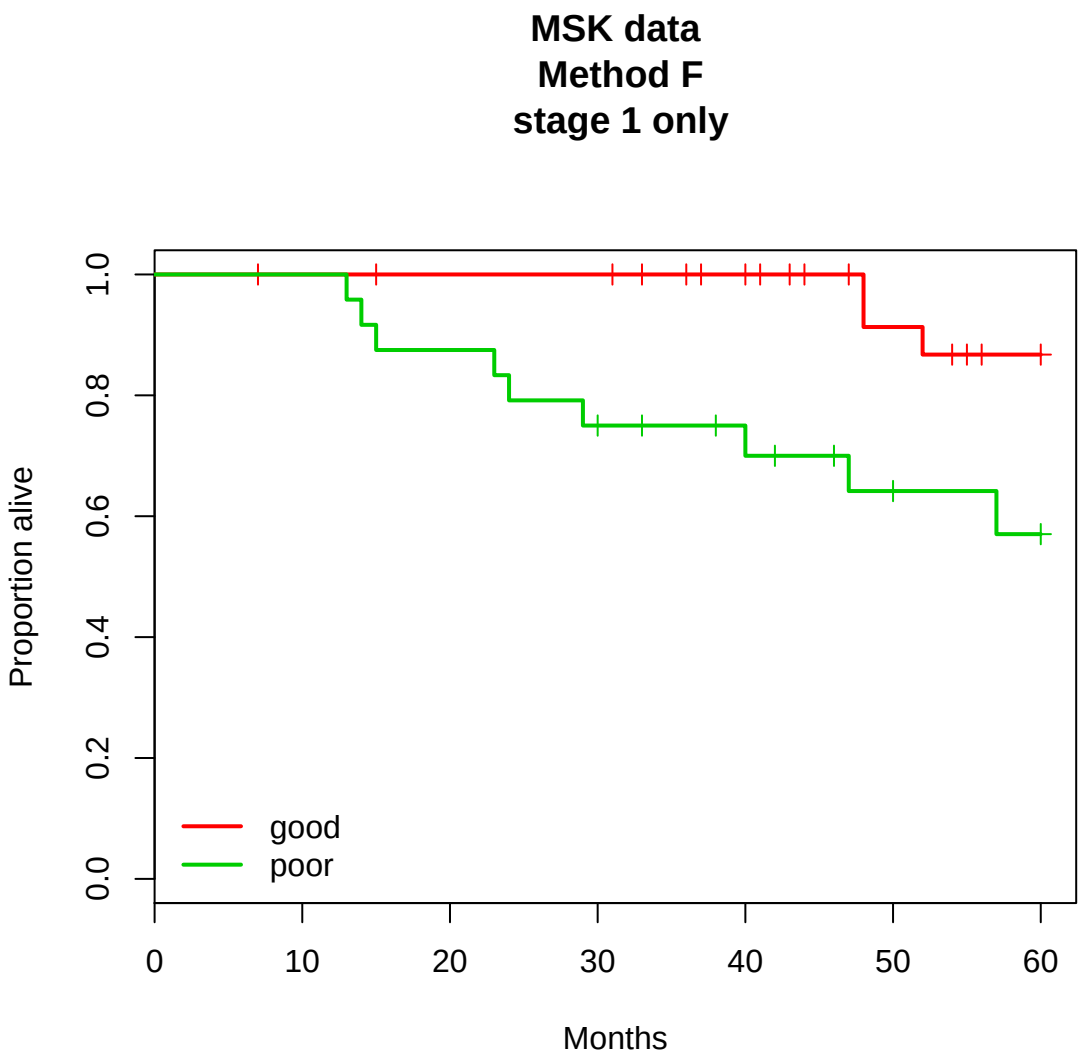

MSK data
Method D
all stages
Proportion alive
0.0
0.2
0.4
0.6
0.8
1.0
0
10
20
30
40
50
60
Months
bad
good

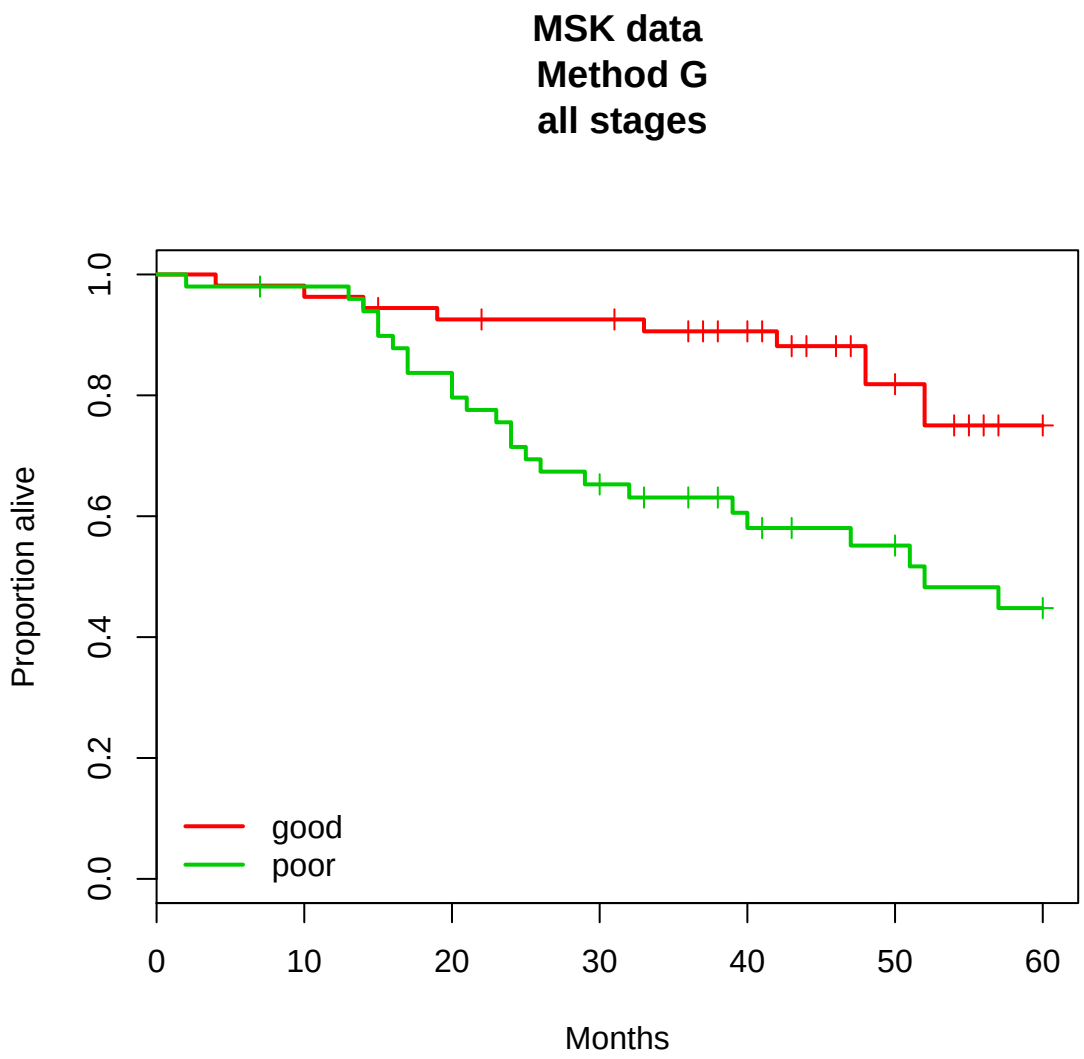

MSK data
Method D
stage 1 only
Proportion alive
0.0
0.2
0.4
0.6
0.8
1.0
0
10
20
30
40
50
60
Months
bad
good

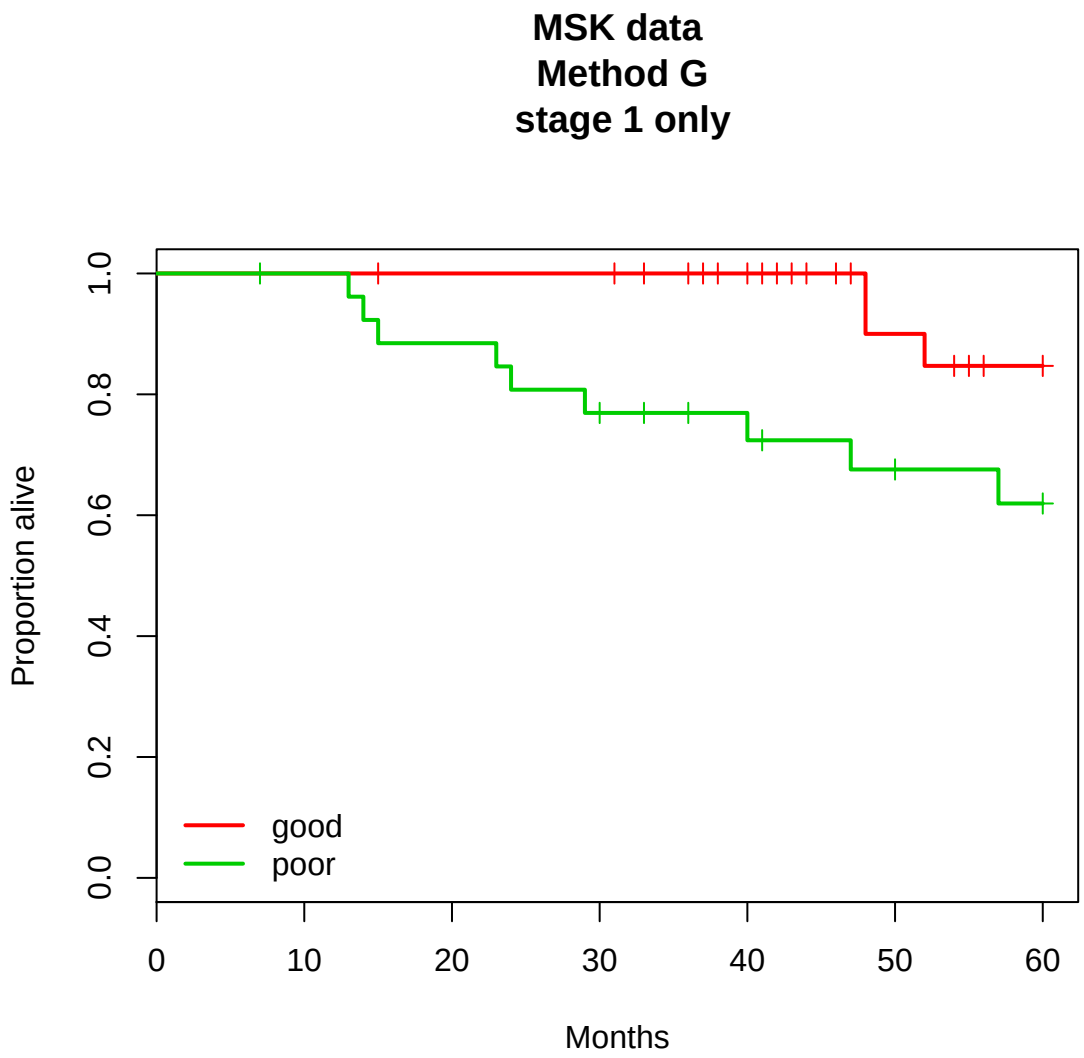

MSK data
Method E
all stages
Proportion alive
0.0
0.2
0.4
0.6
0.8
1.0
0
10
20
30
40
50
60
Months
bad
good

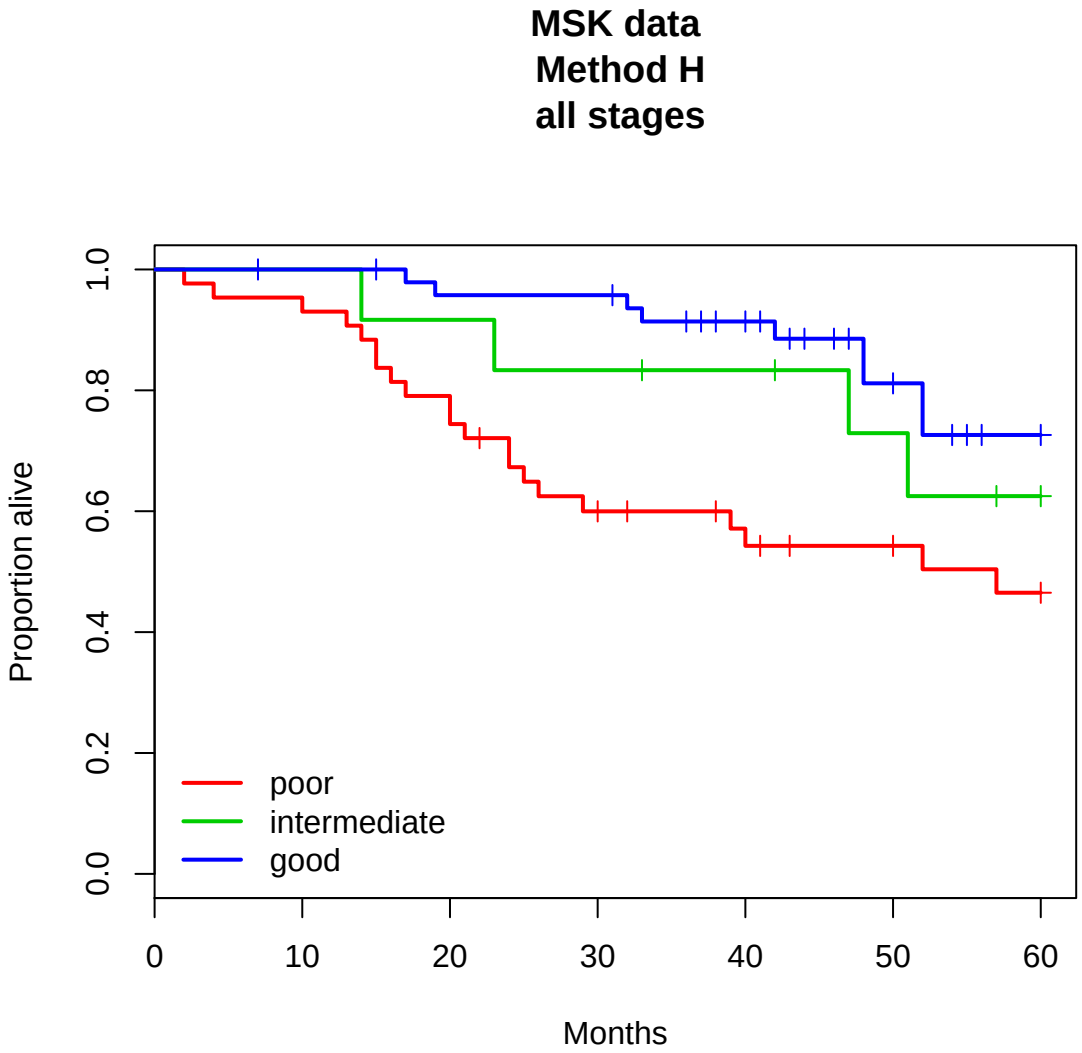

MSK data
Method E
stage 1 only
Proportion alive
0.0
0.2
0.4
0.6
0.8
1.0
0
10
20
30
40
50
60
Months
bad
good

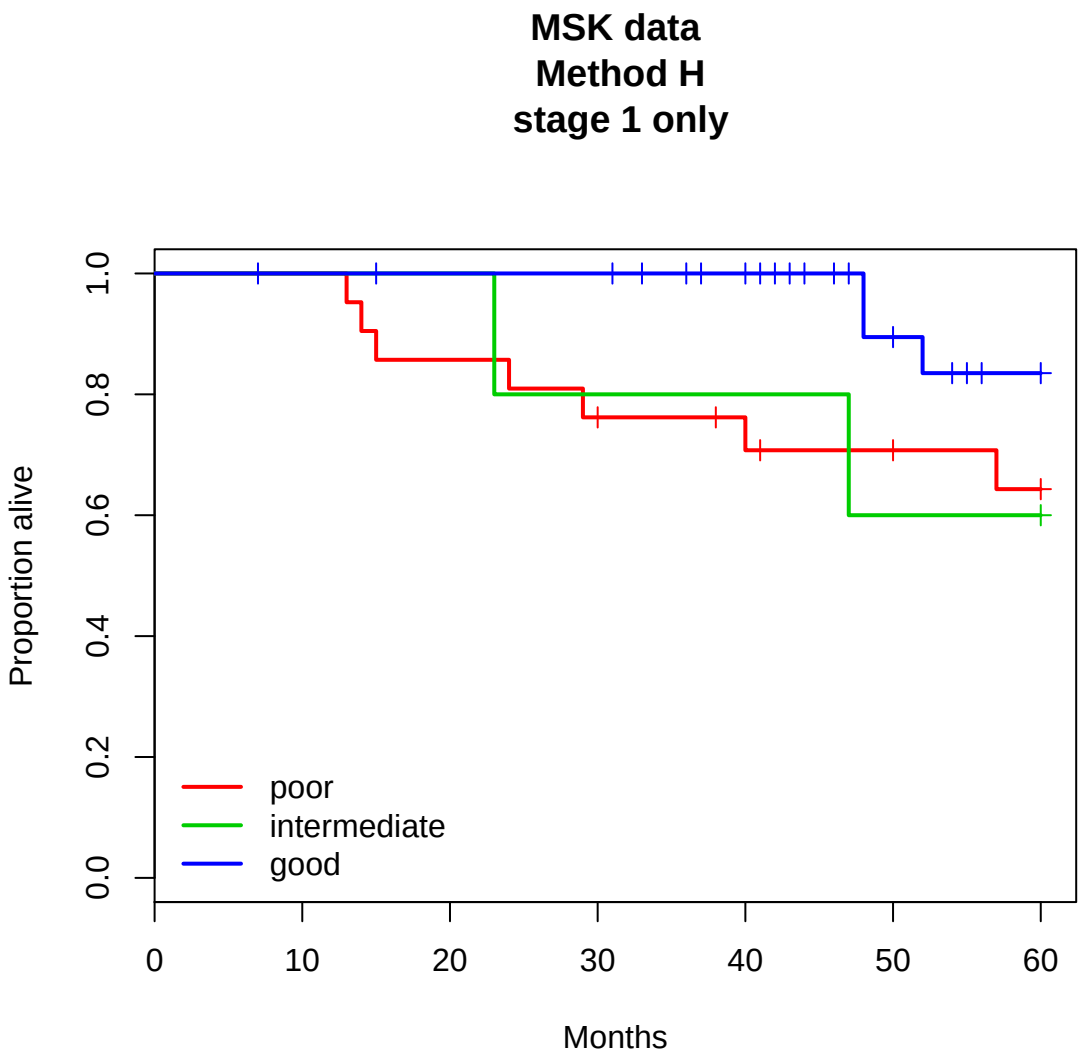

MSK data
Method F
all stages
Proportion alive
0.0
0.2
0.4
0.6
0.8
1.0
0
10
20
30
40
50
60
Months
good
poor

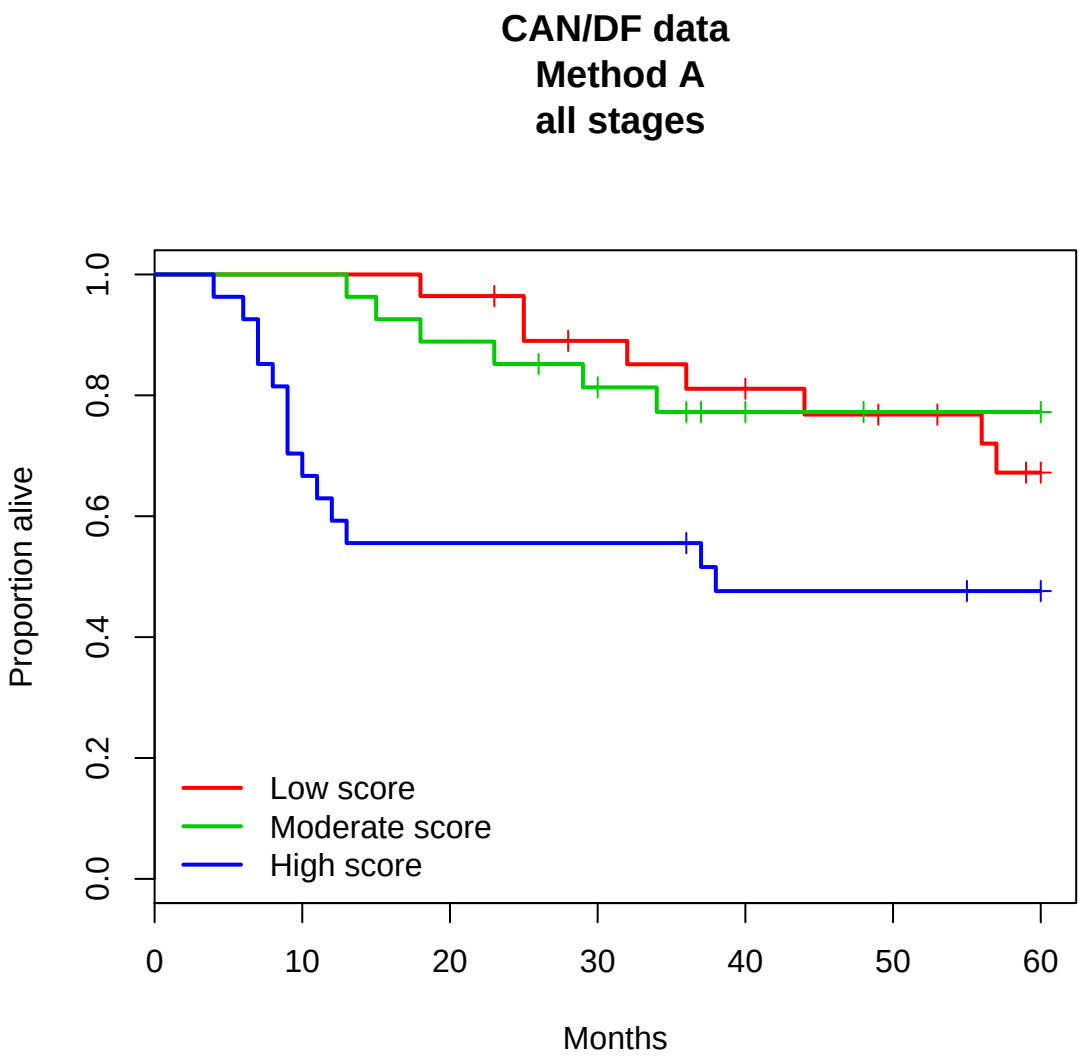

MSK data
Method F
stage 1 only
Proportion alive
0.0
0.2
0.4
0.6
0.8
1.0
0
10
20
30
40
50
60
Months
good
poor

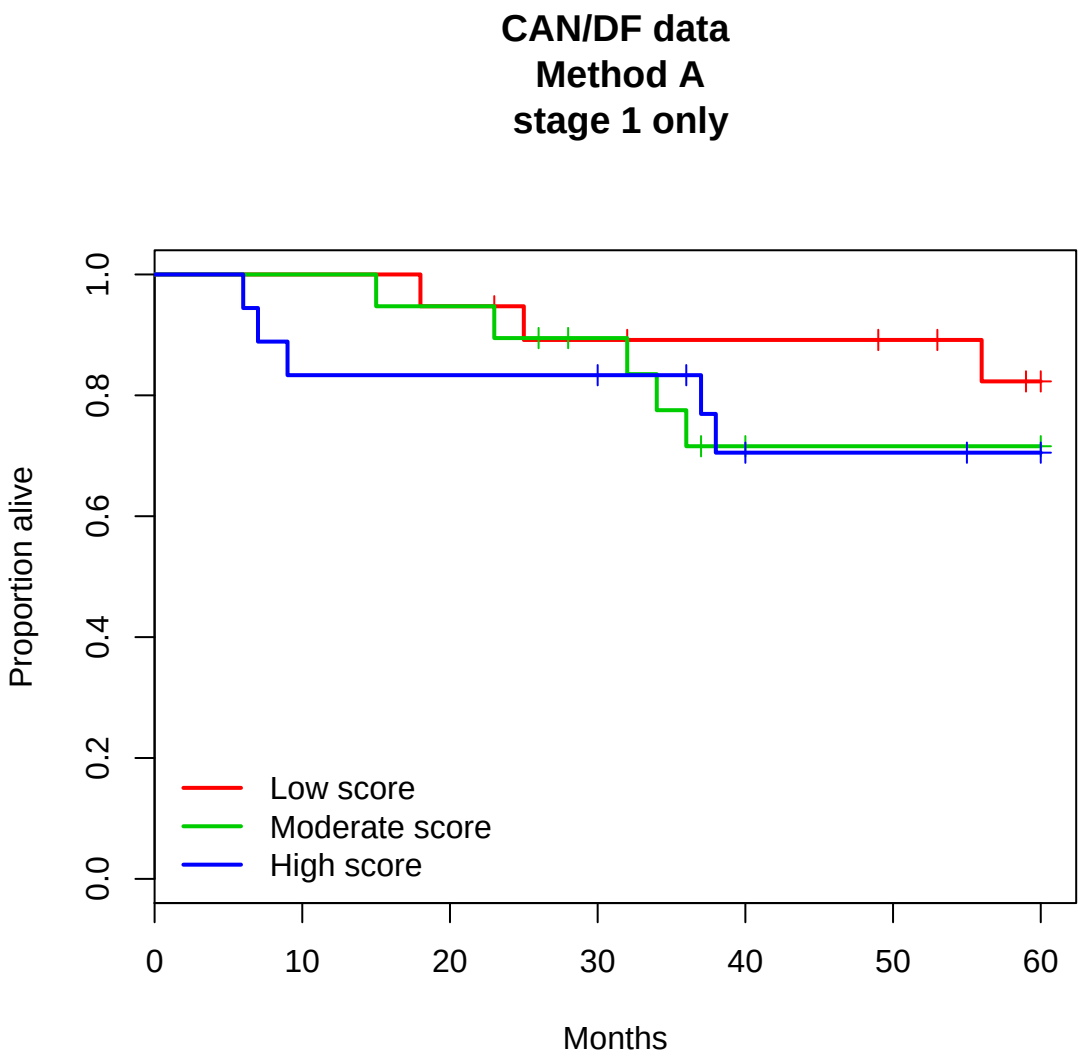

MSK data
Method G
all stages
Proportion alive
0.0
0.2
0.4
0.6
0.8
1.0
0
10
20
30
40
50
60
Months
good
poor

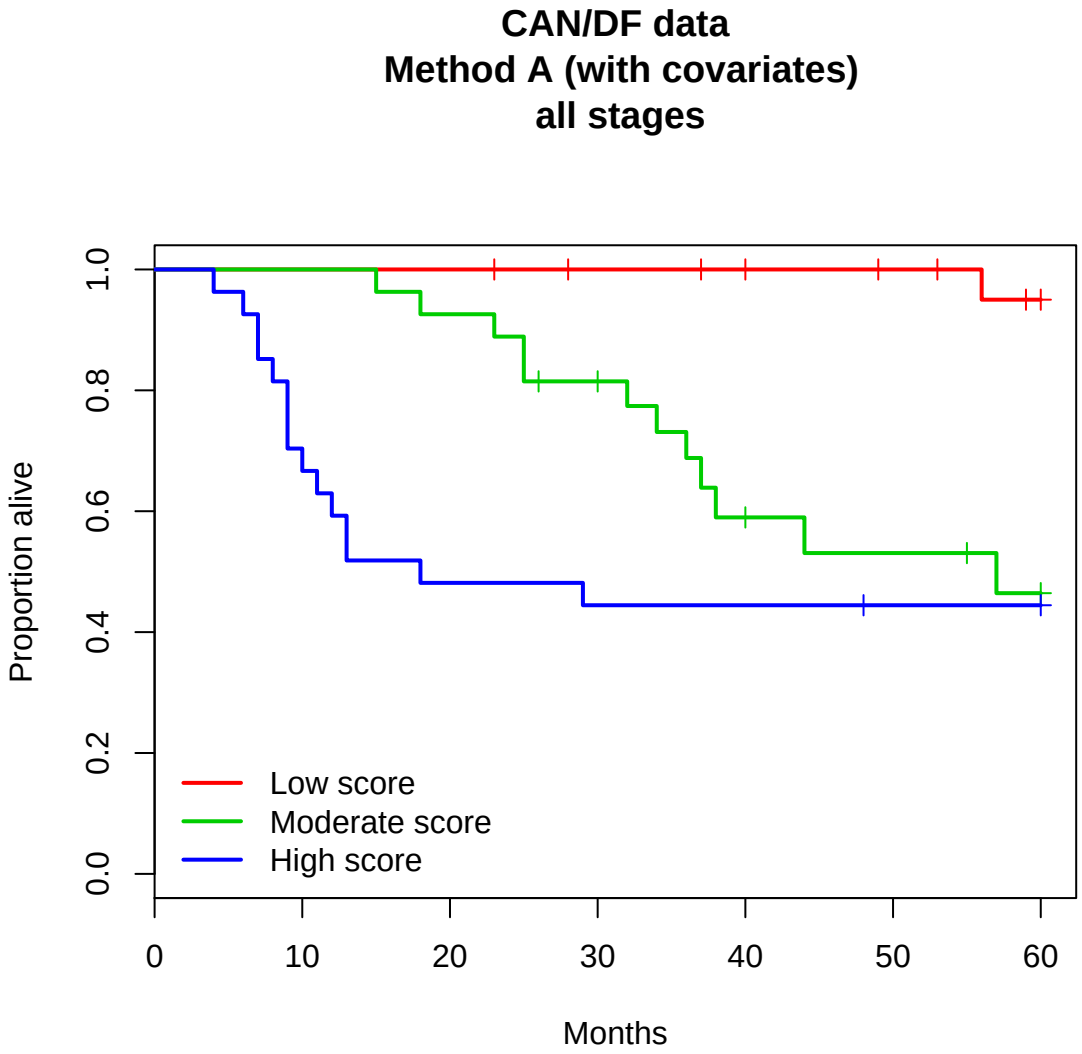

MSK data
Method G
stage 1 only
Proportion alive
0.0
0.2
0.4
0.6
0.8
1.0
0
10
20
30
40
50
60
Months
good
poor

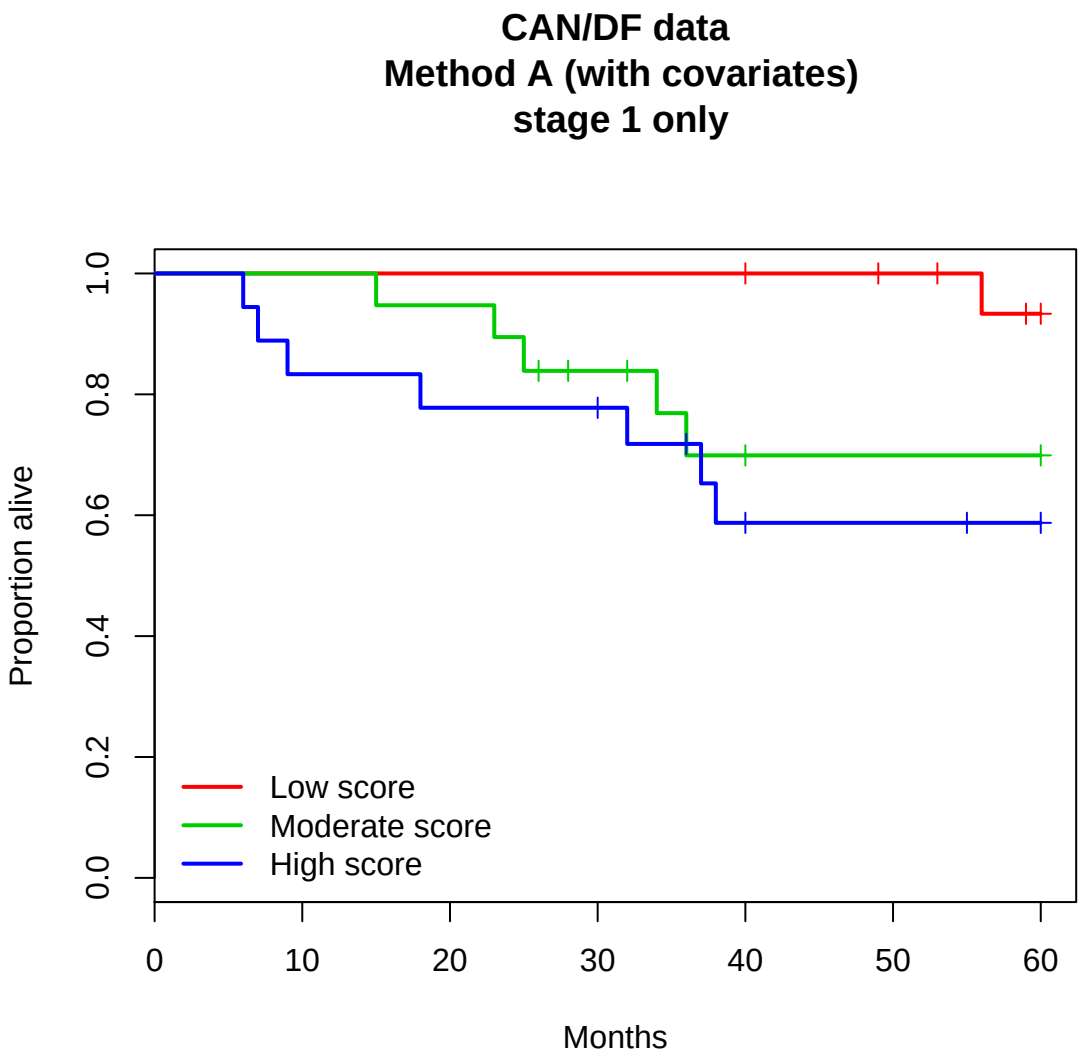

MSK data
Method H
all stages
Proportion alive
0.0
0.2
0.4
0.6
0.8
1.0
0
10
20
30
40
50
60
Months
poor
intermediate
good

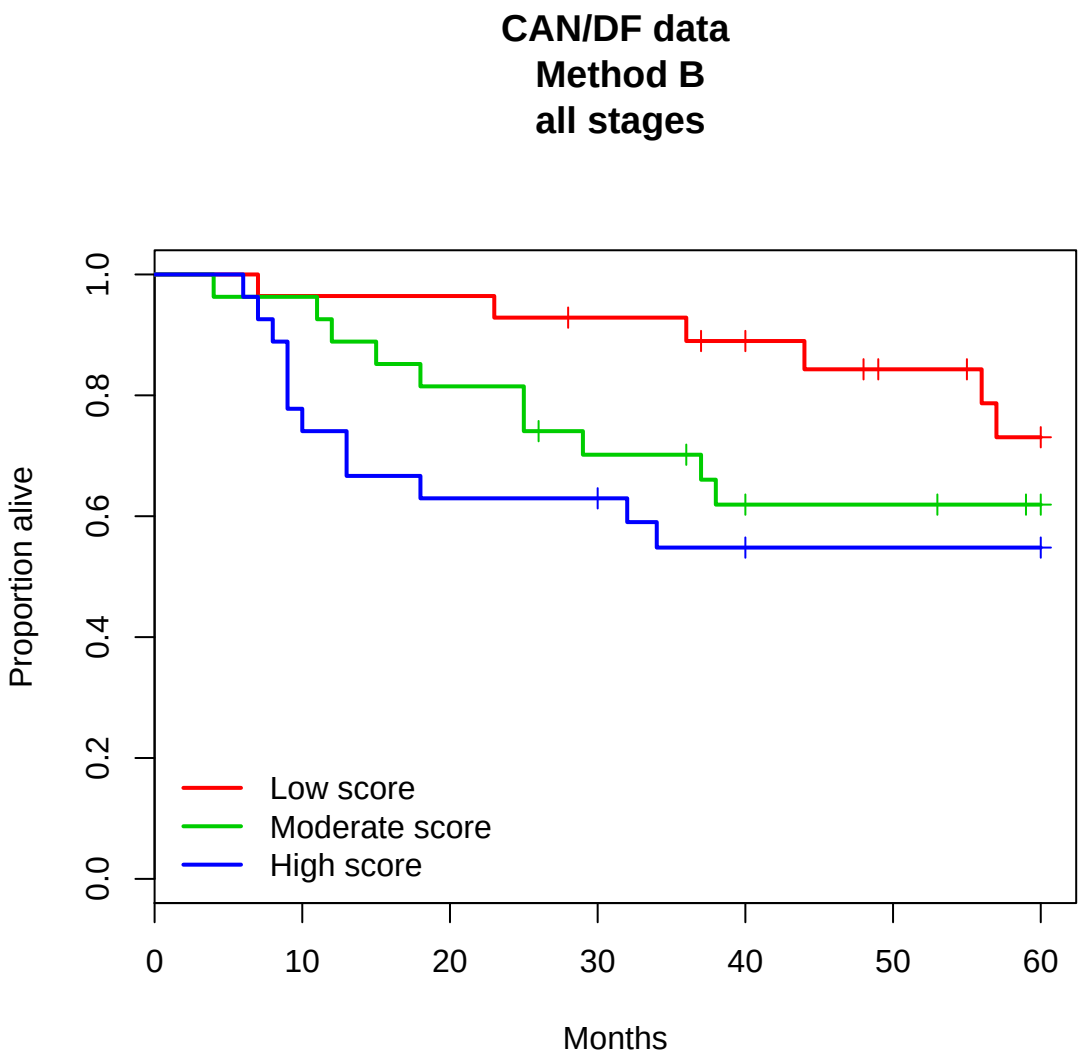

MSK data
Method H
stage 1 only
Proportion alive
0.0
0.2
0.4
0.6
0.8
1.0
0
10
20
30
40
50
60
Months
poor
intermediate
good

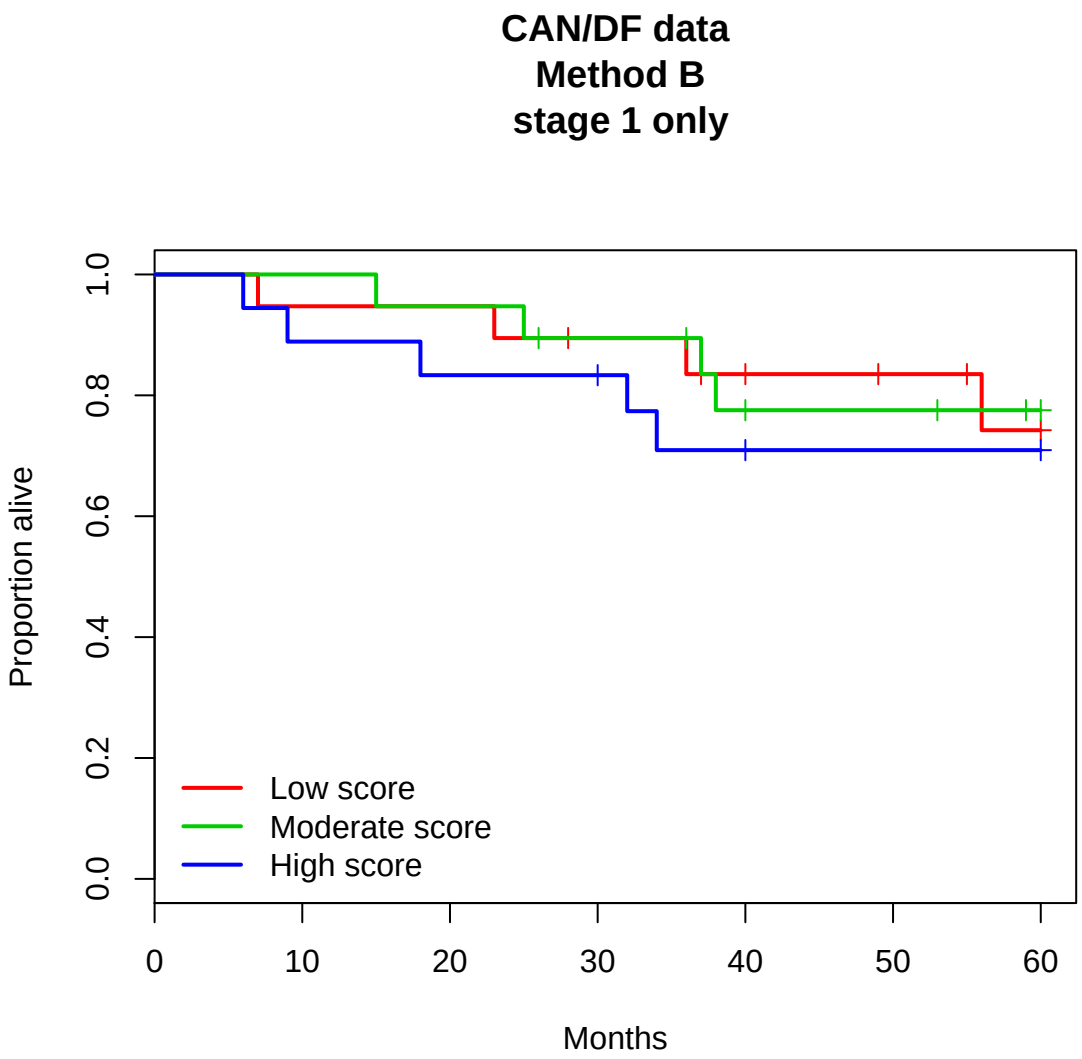

CAN/DF data
Method A
all stages
Proportion alive
0.0
0.2
0.4
0.6
0.8
1.0
0
10
20
30
40
50
60
Months
Low score
Moderate score
High score

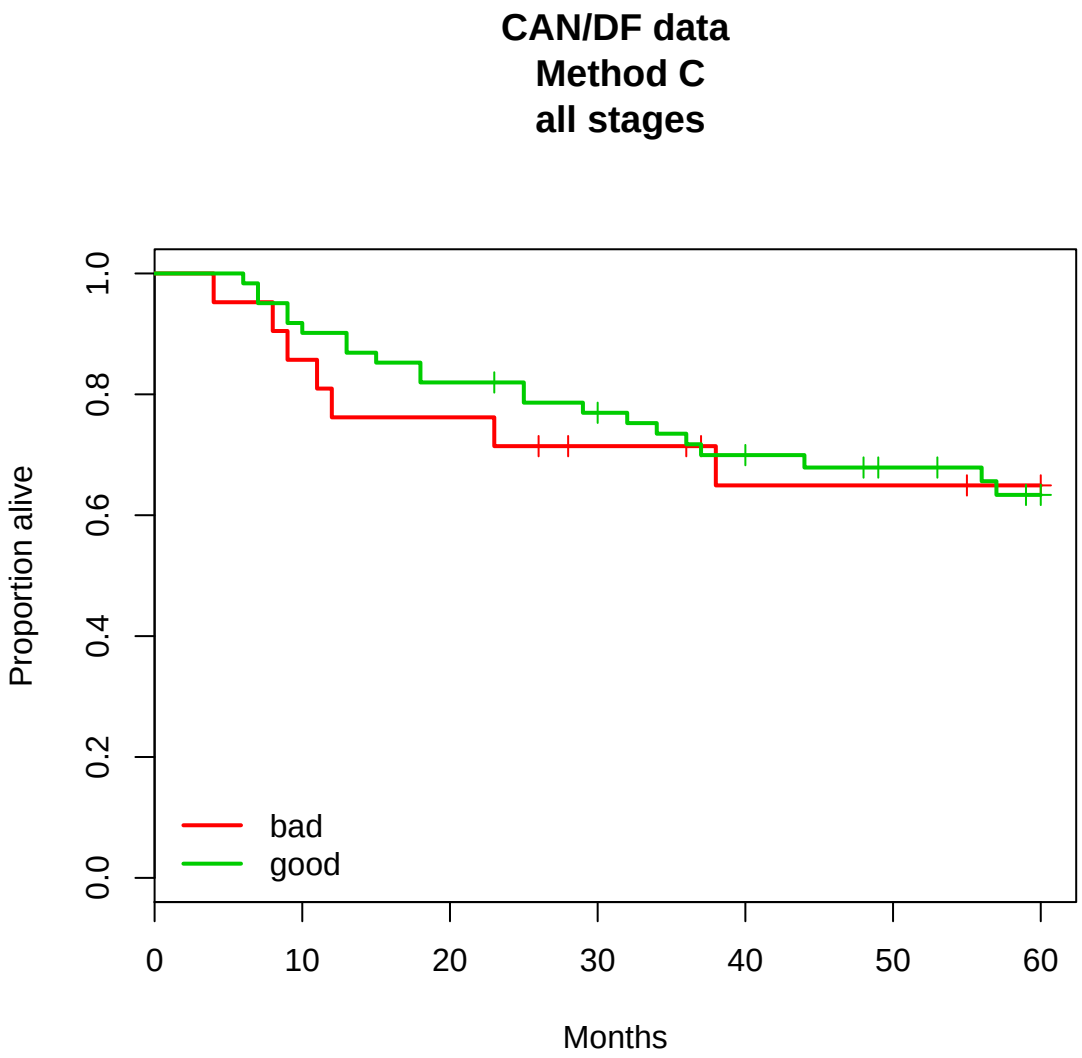

CAN/DF data
Method A
stage 1 only
Proportion alive
0.0
0.2
0.4
0.6
0.8
1.0
0
10
20
30
40
50
60
Months
Low score
Moderate score
High score

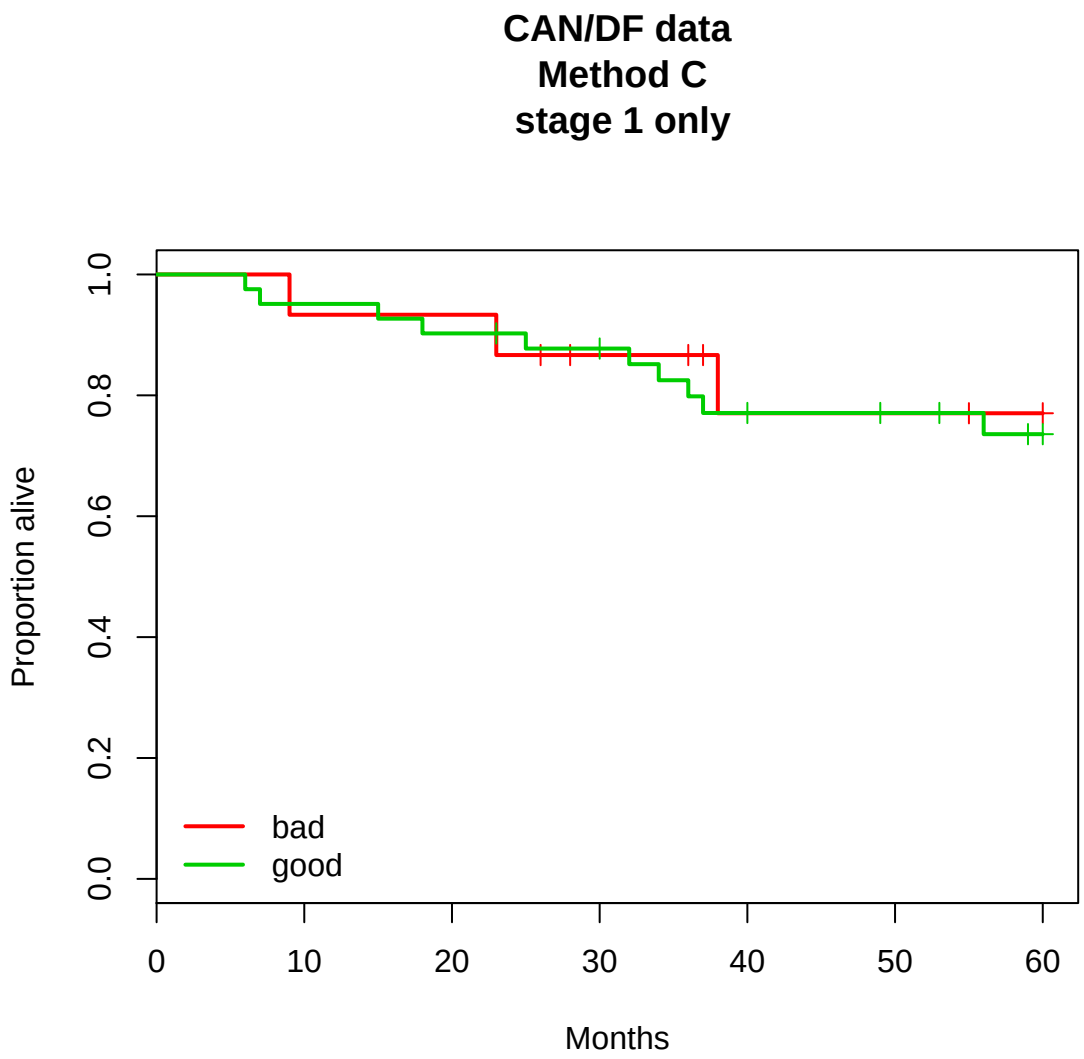

CAN/DF data
Method A (with covariates)
all stages
Proportion alive
0.0
0.2
0.4
0.6
0.8
1.0
0
10
20
30
40
50
60
Months
Low score
Moderate score
High score

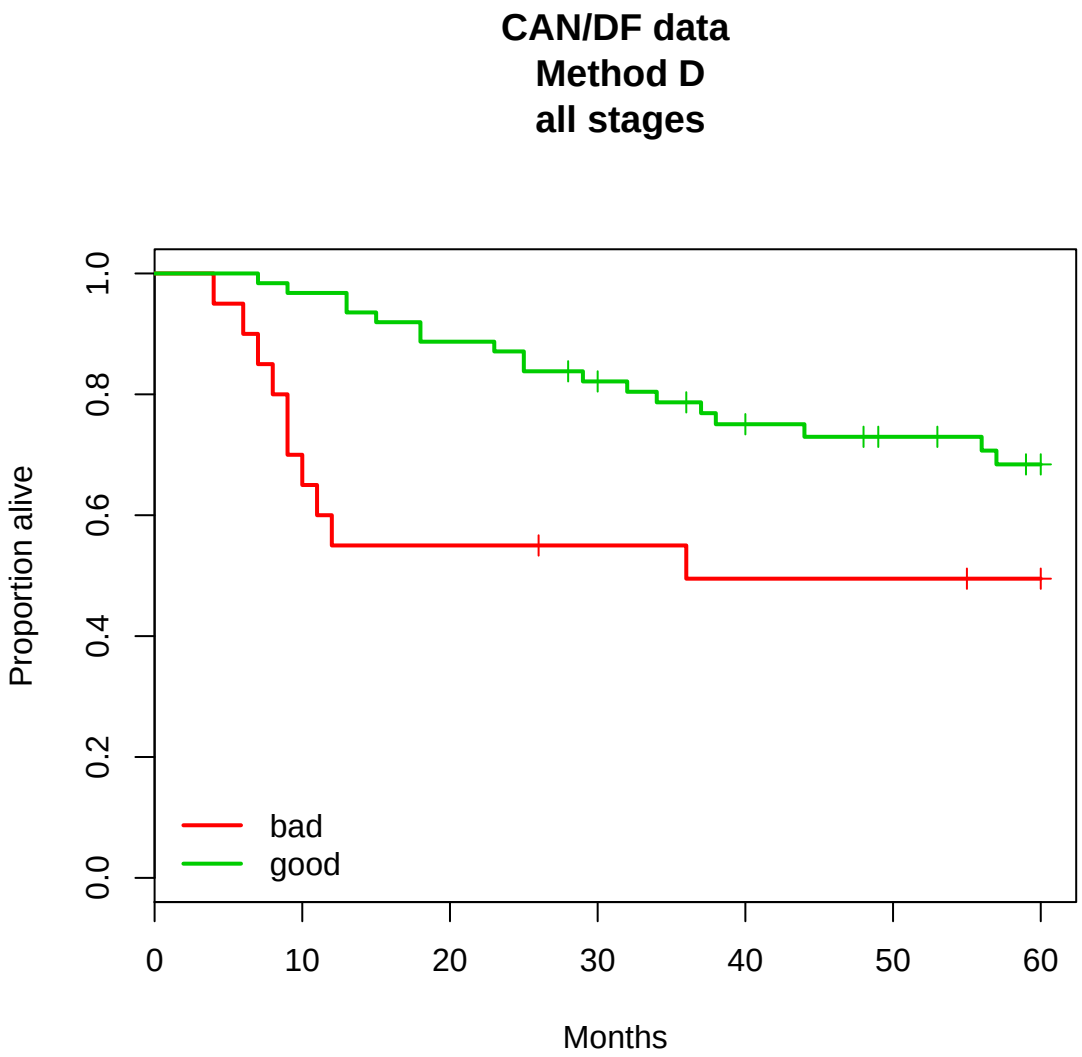

CAN/DF data
Method A (with covariates)
stage 1 only
Proportion alive
0.0
0.2
0.4
0.6
0.8
1.0
0
10
20
30
40
50
60
Months
Low score
Moderate score
High score

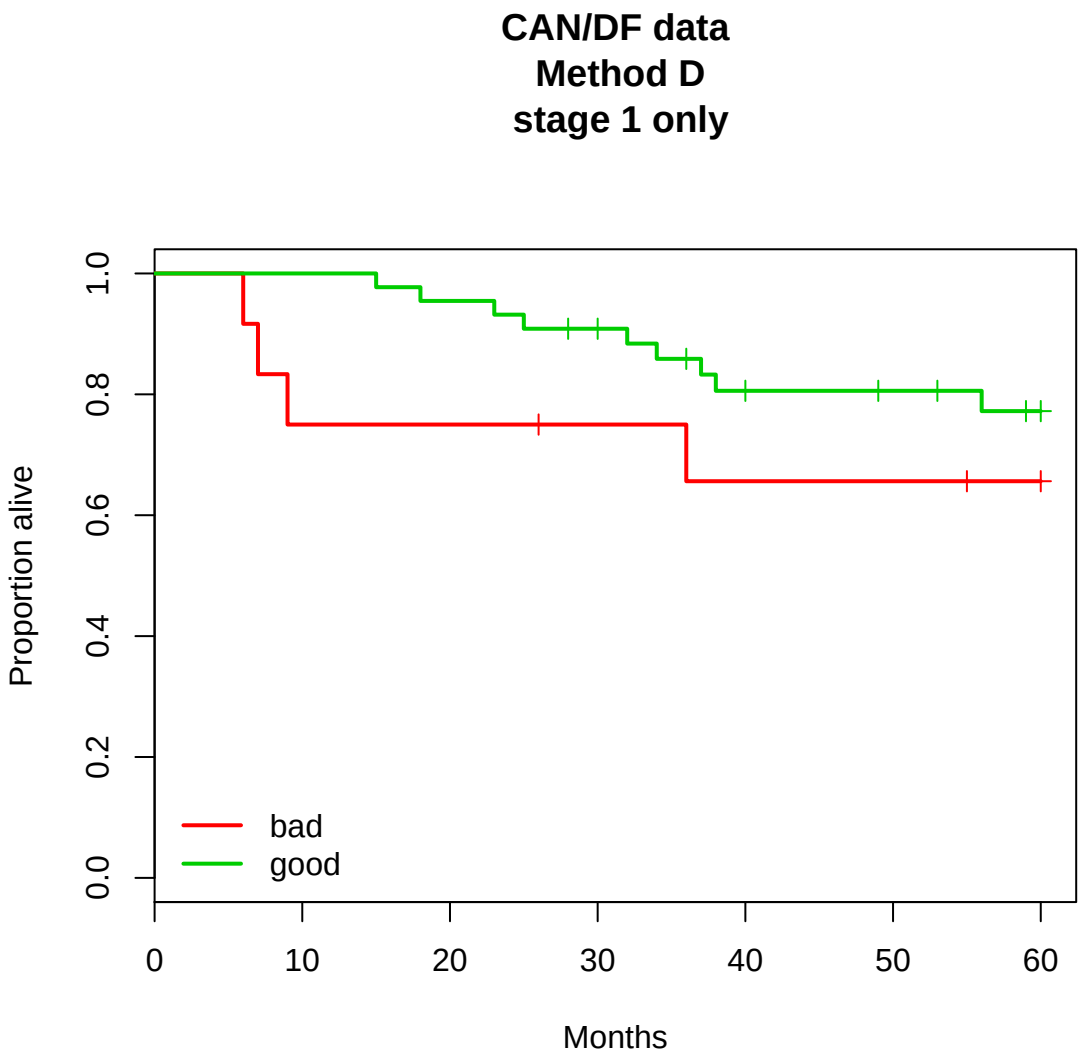

CAN/DF data
Method B
all stages
Proportion alive
0.0
0.2
0.4
0.6
0.8
1.0
0
10
20
30
40
50
60
Months
Low score
Moderate score
High score

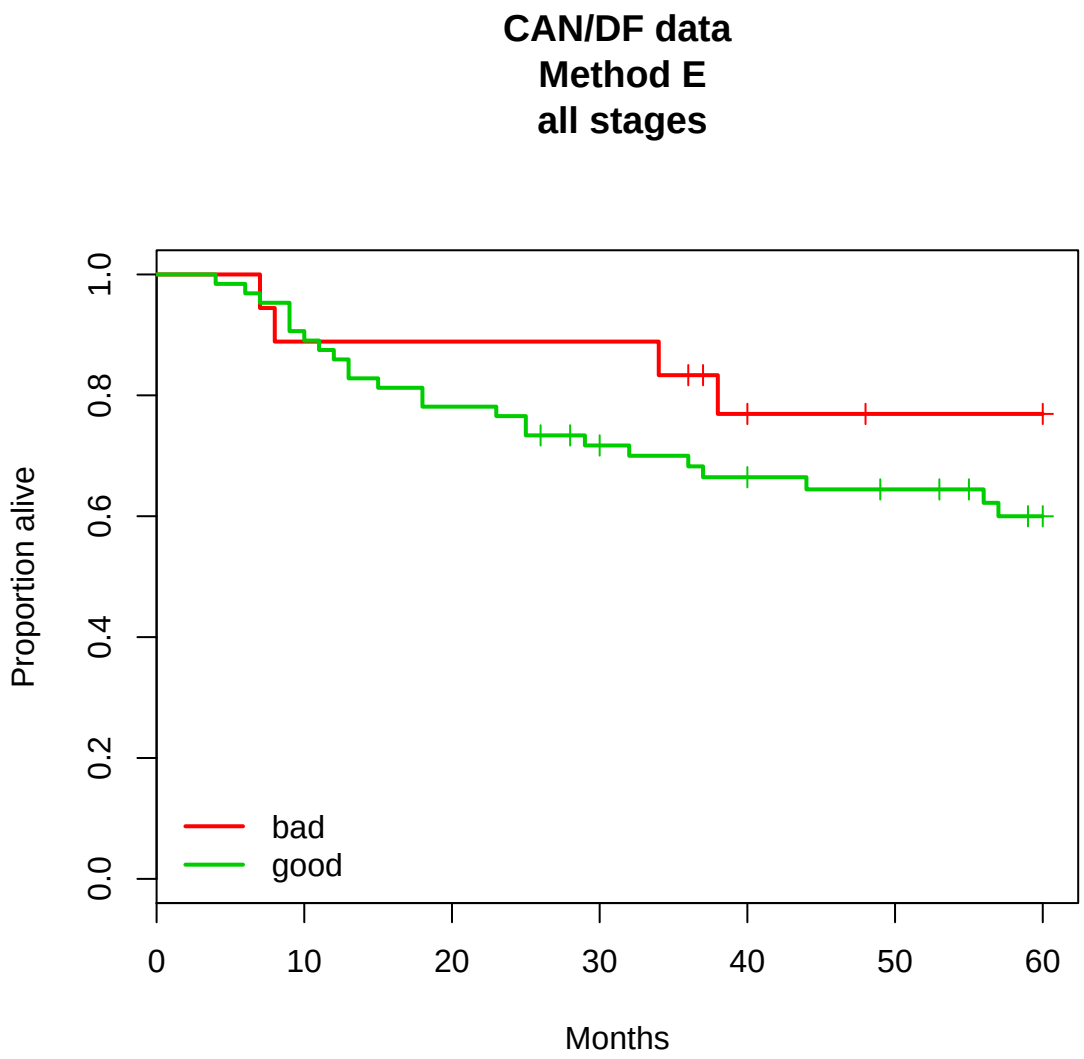

CAN/DF data
Method B
stage 1 only
Proportion alive
0.0
0.2
0.4
0.6
0.8
1.0
0
10
20
30
40
50
60
Months
Low score
Moderate score
High score

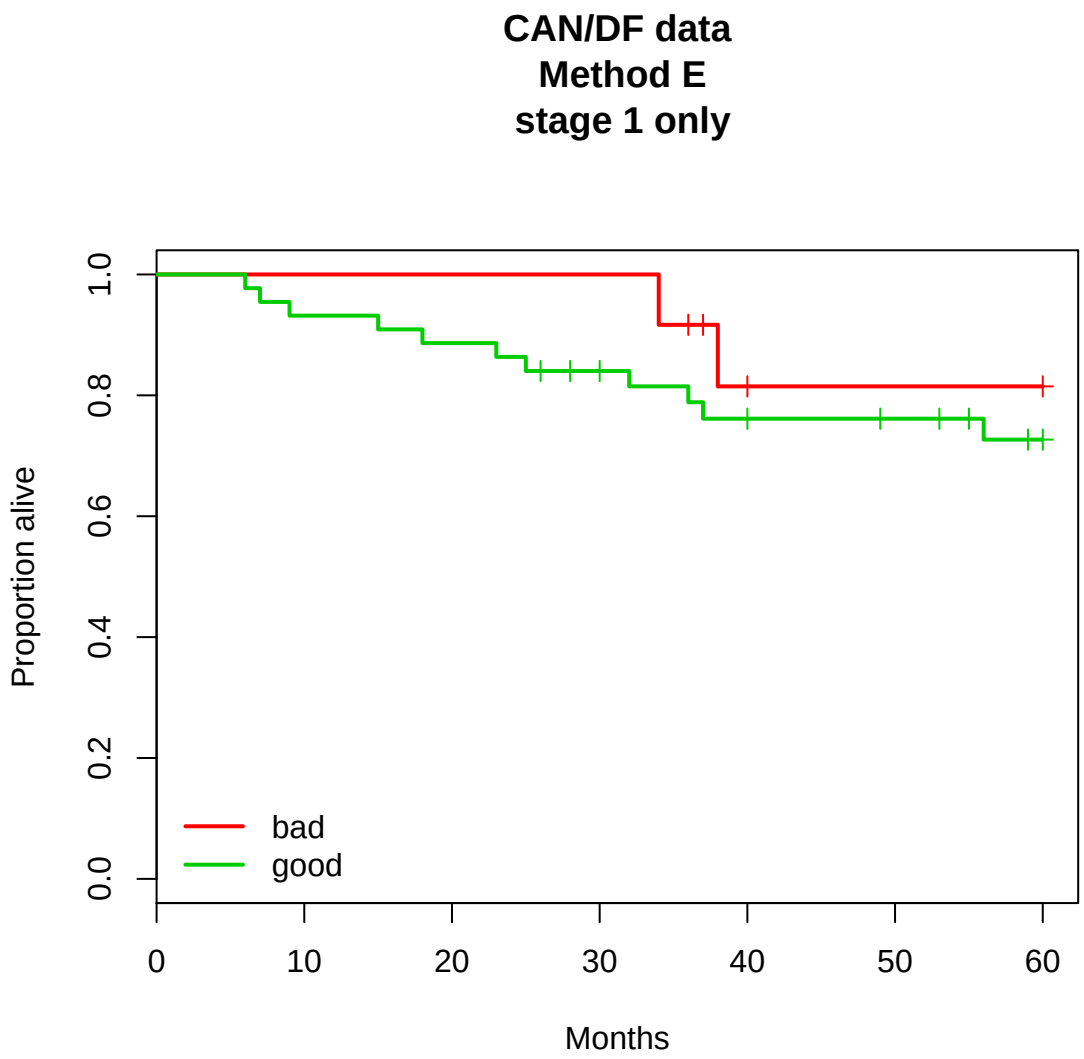

CAN/DF data
Method C
all stages
Proportion alive
0.0
0.2
0.4
0.6
0.8
1.0
0
10
20
30
40
50
60
Months
bad
good

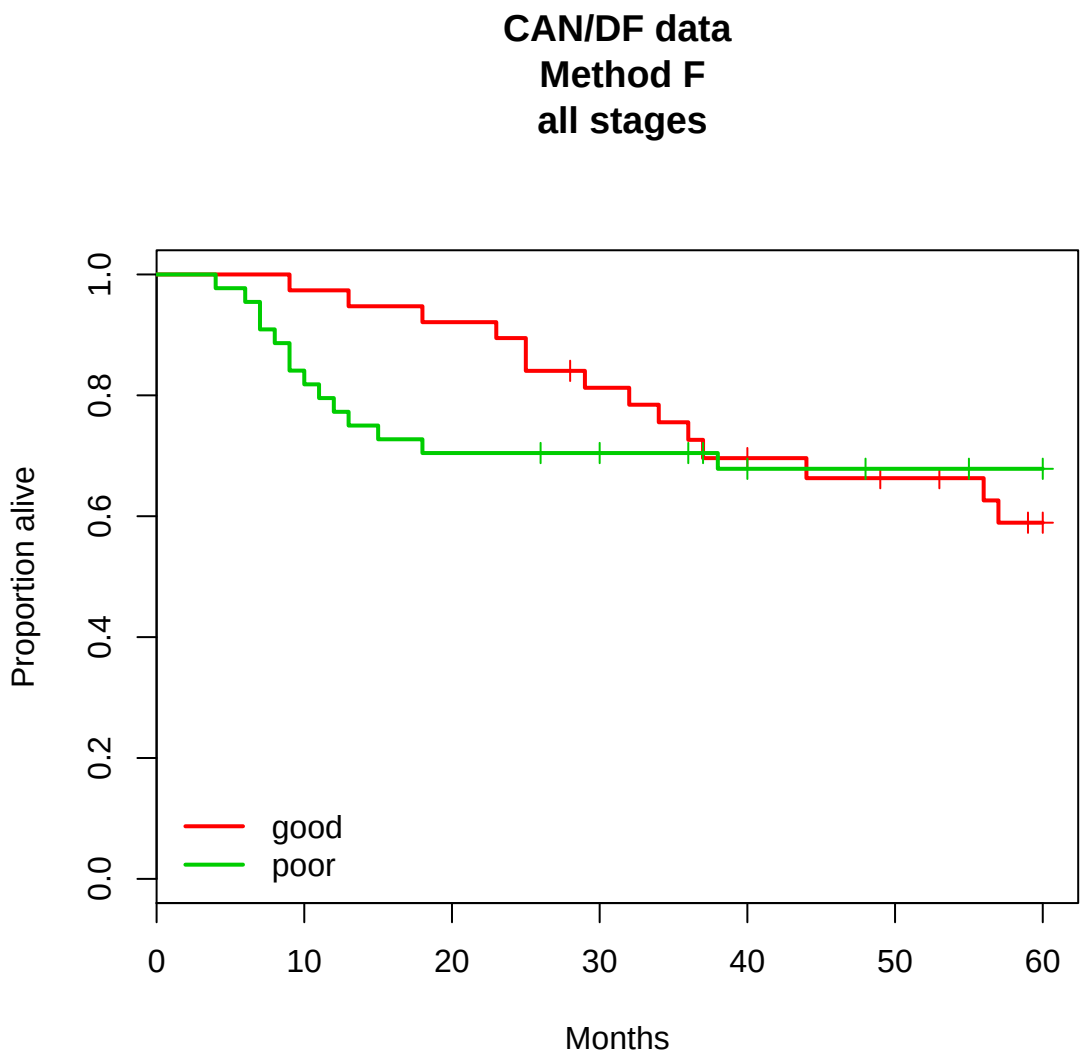

CAN/DF data
Method C
stage 1 only
Proportion alive
0.0
0.2
0.4
0.6
0.8
1.0
0
10
20
30
40
50
60
Months
bad
good

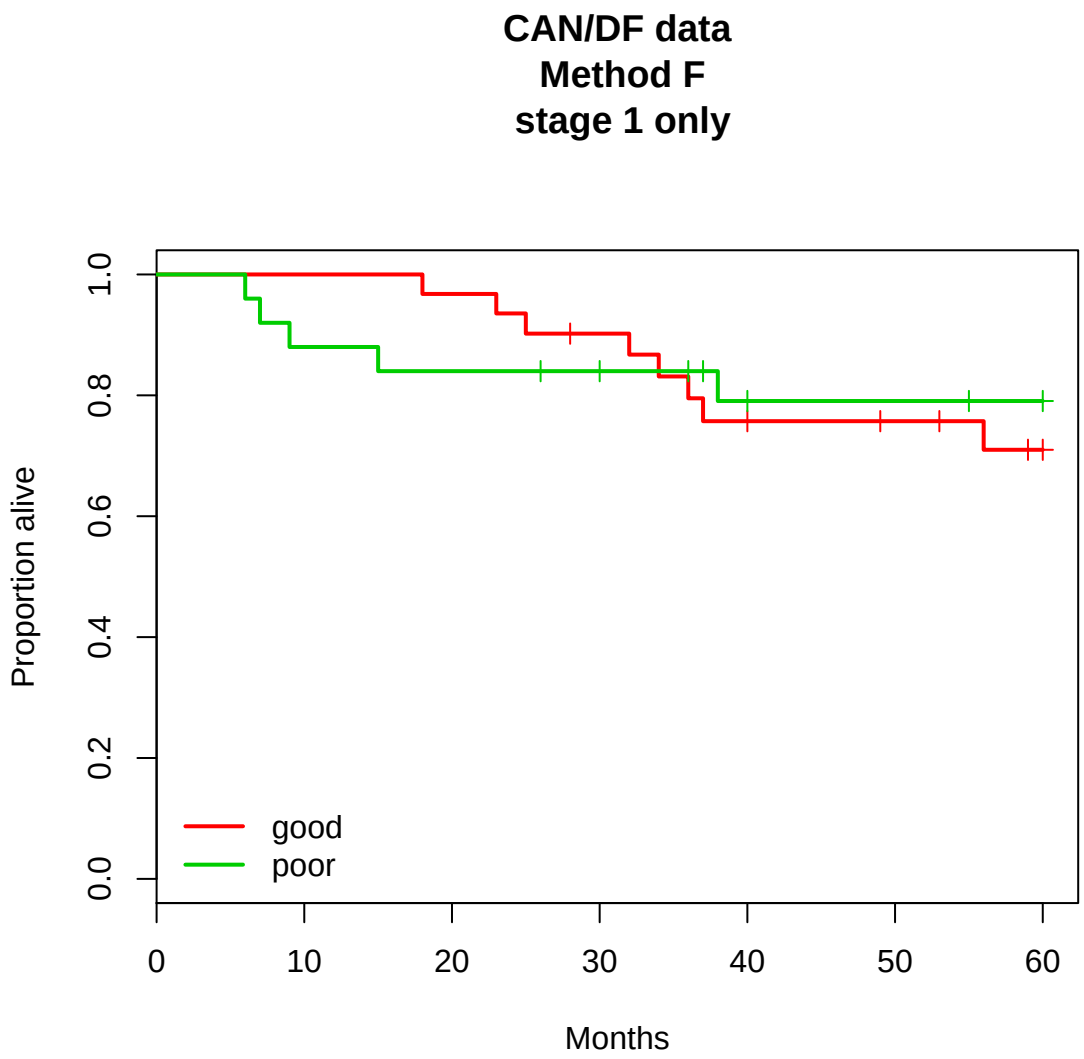

CAN/DF data
Method D
all stages
Proportion alive
0.0
0.2
0.4
0.6
0.8
1.0
0
10
20
30
40
50
60
Months
bad
good

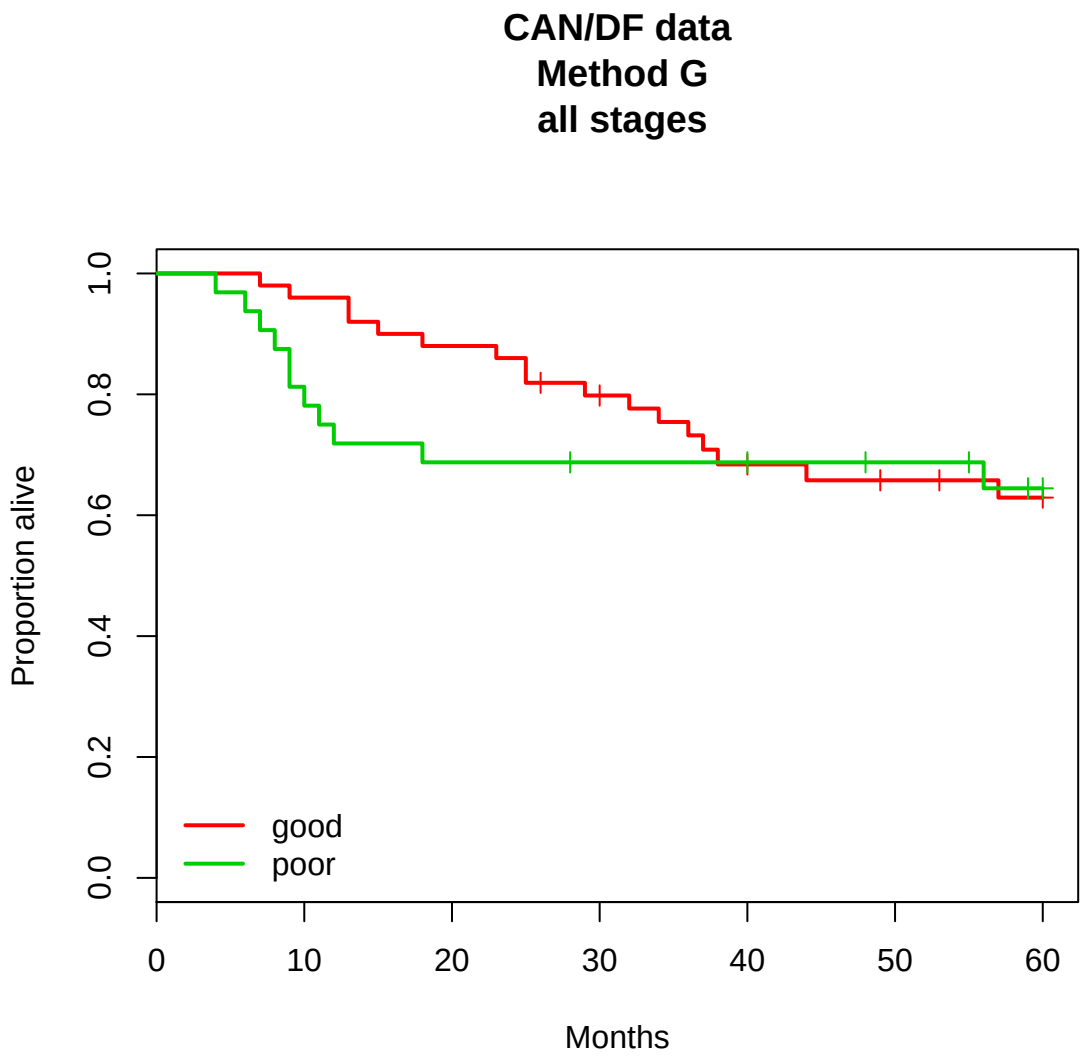

CAN/DF data
Method D
stage 1 only
Proportion alive
0.0
0.2
0.4
0.6
0.8
1.0
0
10
20
30
40
50
60
Months
bad
good

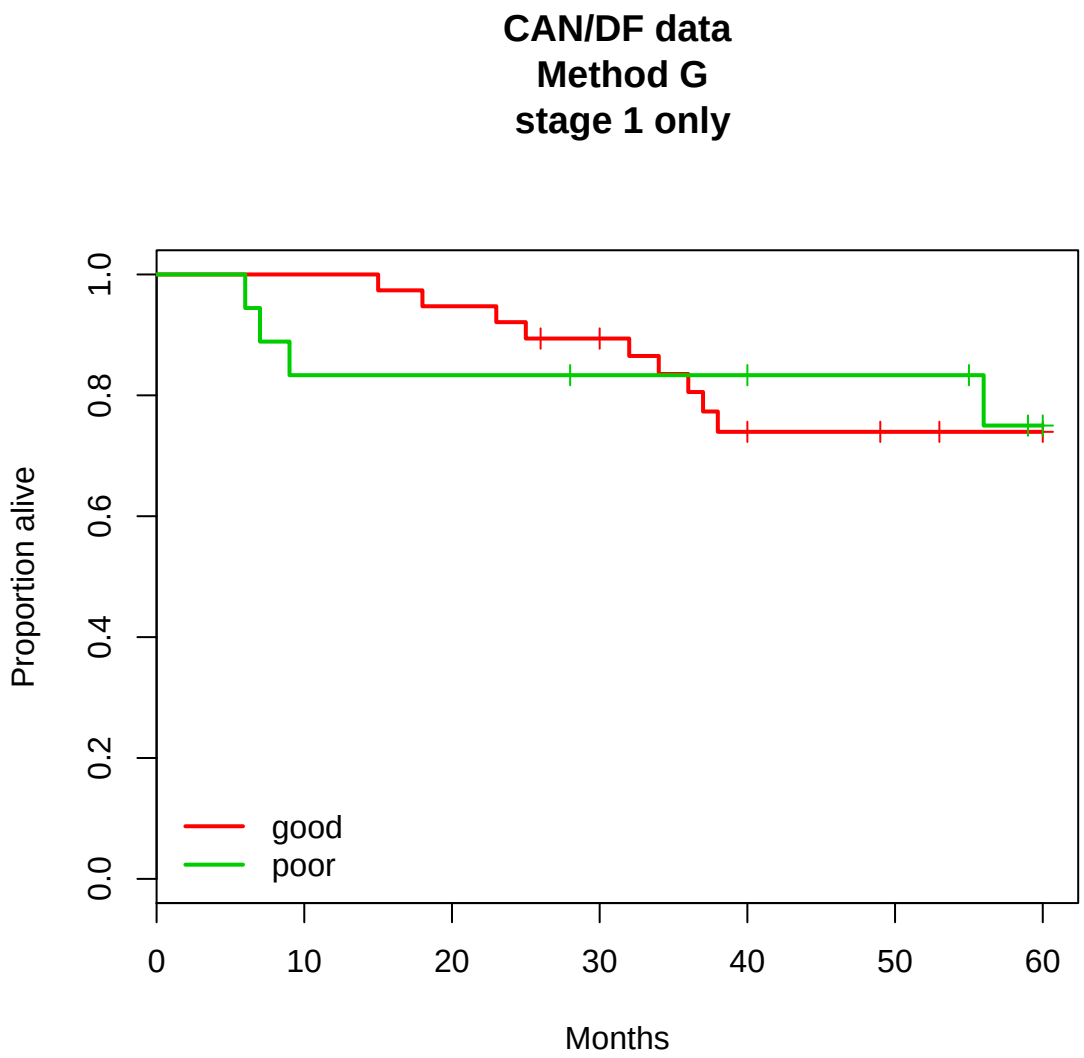

CAN/DF data
Method E
all stages
Proportion alive
0.0
0.2
0.4
0.6
0.8
1.0
0
10
20
30
40
50
60
Months
bad
good

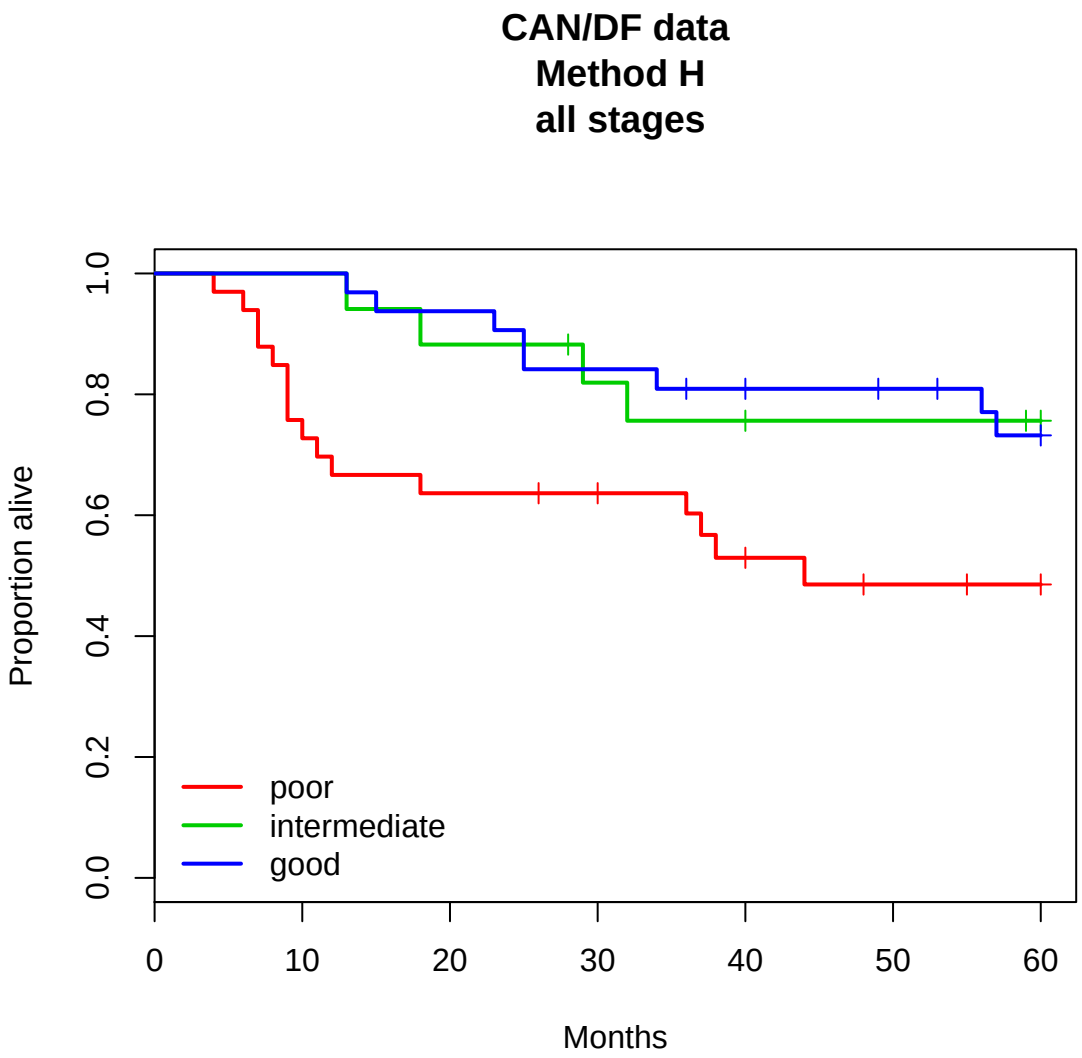

CAN/DF data
Method E
stage 1 only
Proportion alive
0.0
0.2
0.4
0.6
0.8
1.0
0
10
20
30
40
50
60
Months
bad
good

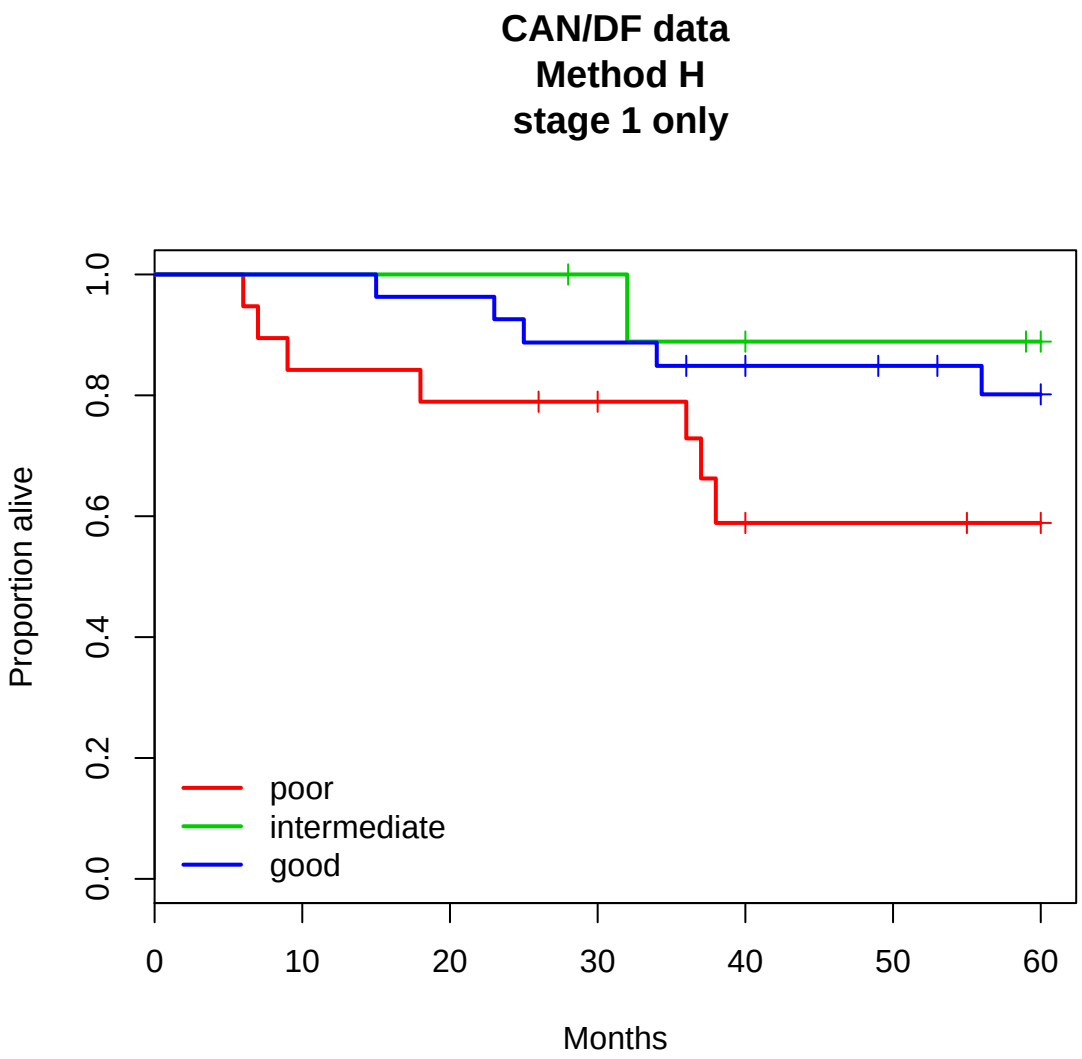

CAN/DF data
Method F
all stages
Proportion alive
0.0
0.2
0.4
0.6
0.8
1.0
0
10
20
30
40
50
60
Months
good
poor


CAN/DF data
Method F
stage 1 only
Proportion alive
0.0
0.2
0.4
0.6
0.8
1.0
0
10
20
30
40
50
60
Months
good
poor

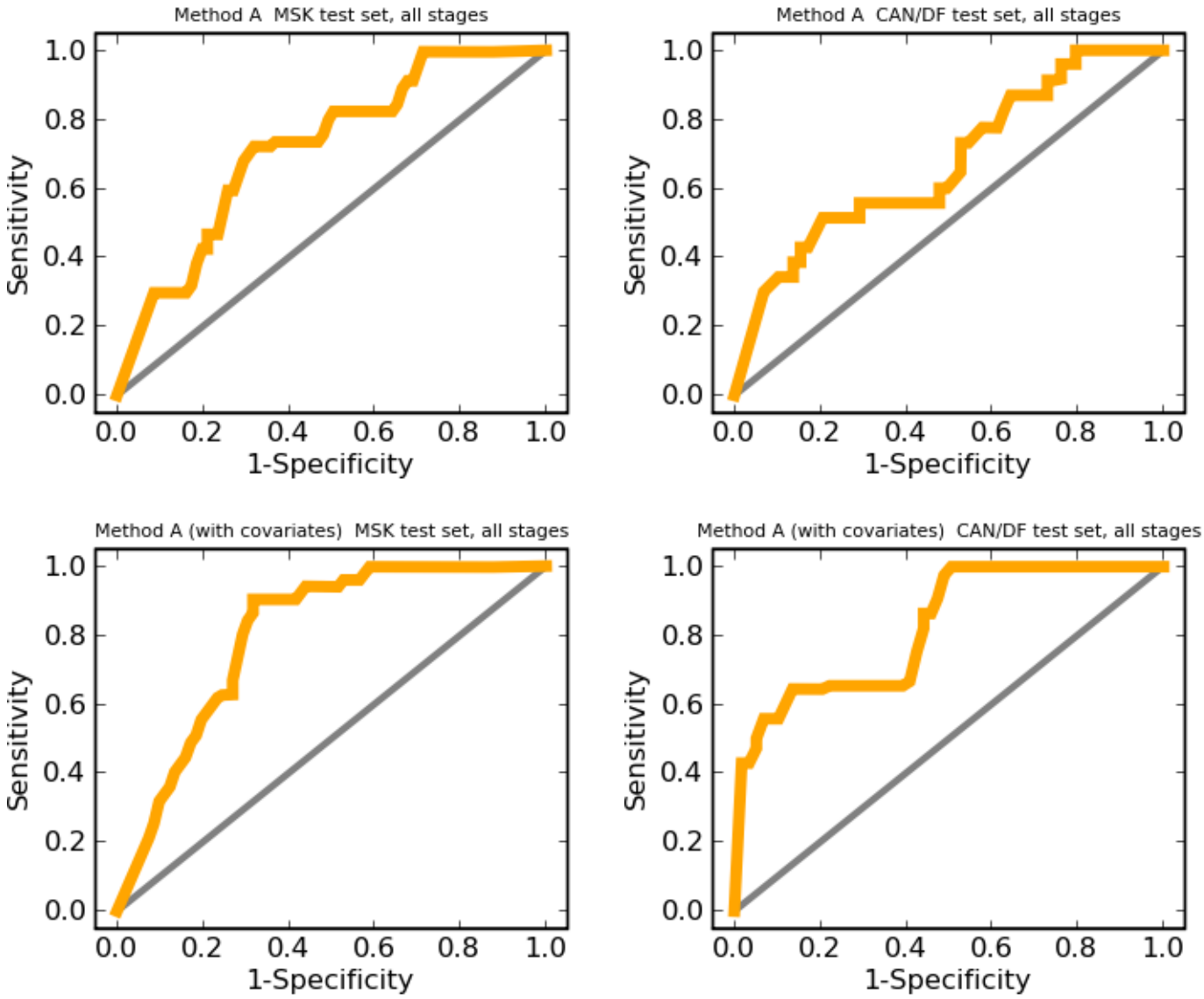

CAN/DF data
Method G
all stages
Proportion alive
0.0
0.2
0.4
0.6
0.8
1.0
0
10
20
30
40
50
60
Months
good
poor

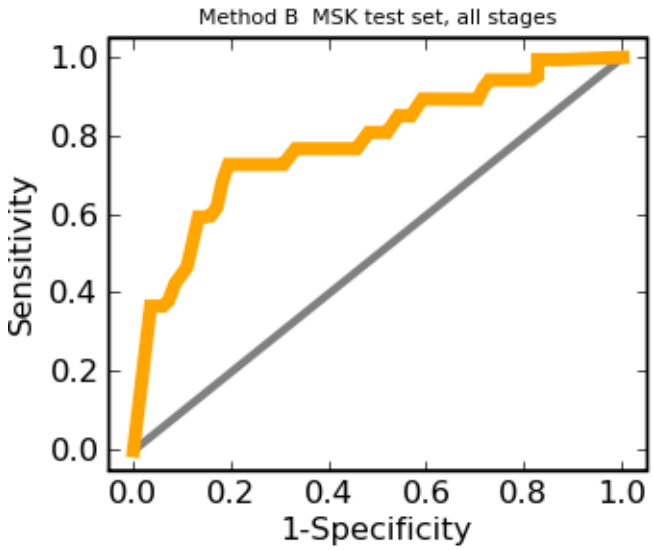

CAN/DF data
Method G
stage 1 only
Proportion alive
0.0
0.2
0.4
0.6
0.8
1.0
0
10
20
30
40
50
60
Months
good
poor

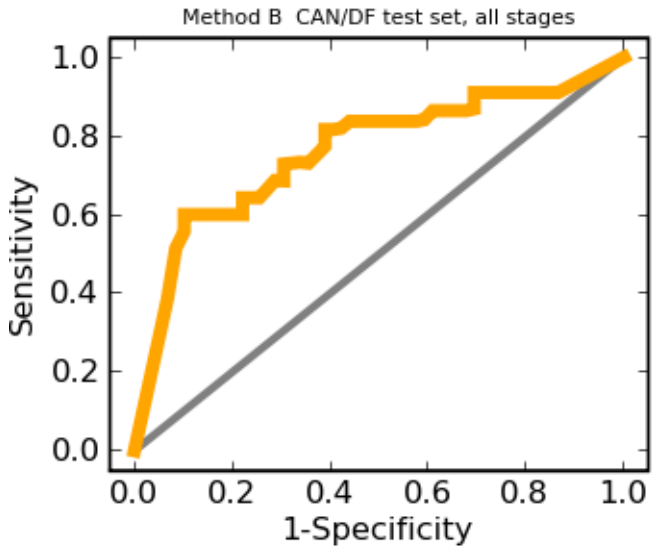

CAN/DF data
Method H
all stages
Proportion alive
0.0
0.2
0.4
0.6
0.8
1.0
0
10
20
30
40
50
60
Months
poor
intermediate
good

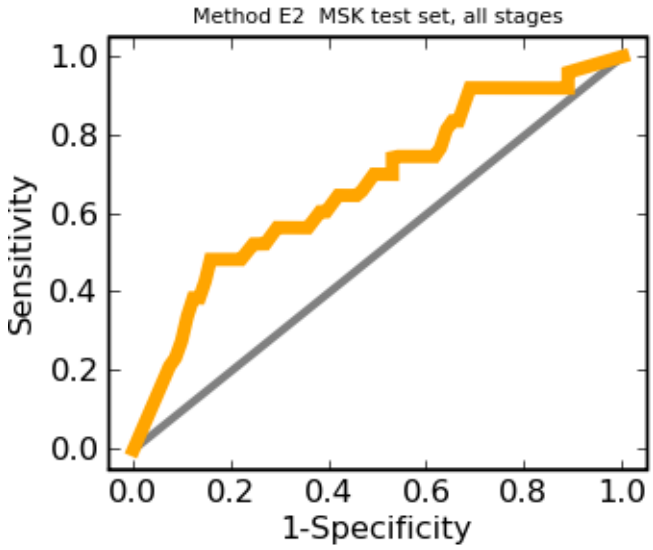

CAN/DF data
Method H
stage 1 only
Proportion alive
0.0
0.2
0.4
0.6
0.8
1.0
0
10
20
30
40
50
60
Months
poor
intermediate
good

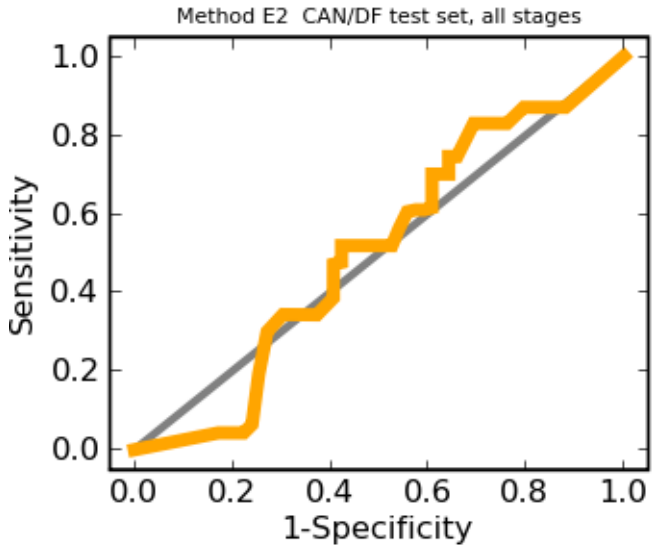
Supp. Fig. 2

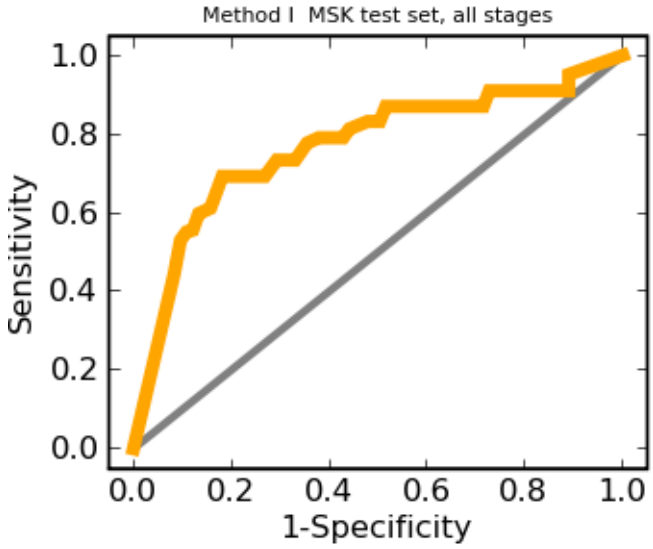

Method A MSK test set, all stages
Method A CAN/DF test set, all stages
Method A (with covariates) MSK test set, all stages
Method A (with covariates) CAN/DF test set, all stages
Sensitivity
1-Specificity
0.0
0.2
0.4
0.6
0.8
1.0

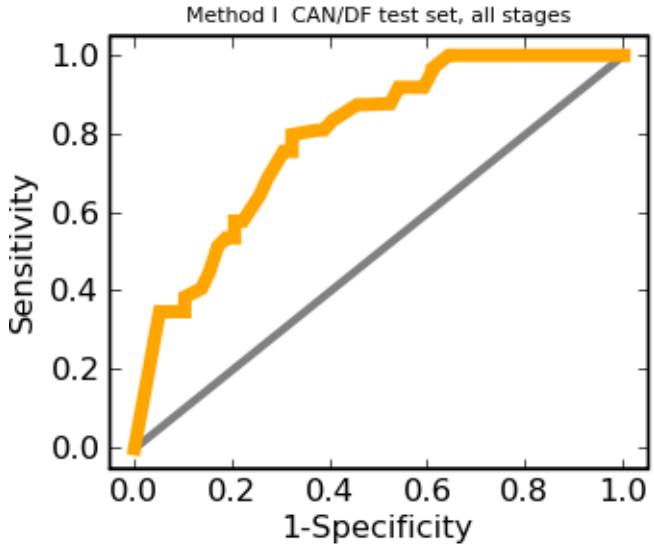

Method B MSK test set, all stages
Sensitivity
1-Specificity
0.0
0.2
0.4
0.6
0.8
1.0

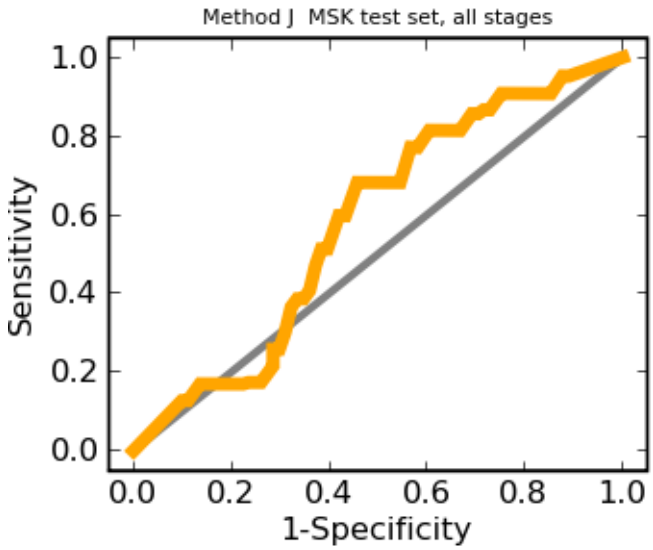

Method B CAN/DF test set, all stages
Sensitivity
1-Specificity
0.0
0.2
0.4
0.6
0.8
1.0

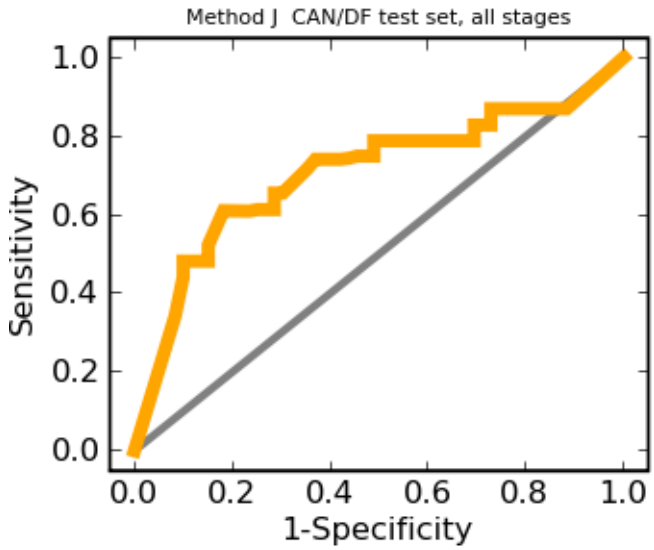

Method E2 MSK test set, all stages
Sensitivity
1-Specificity
0.0
0.2
0.4
0.6
0.8
1.0

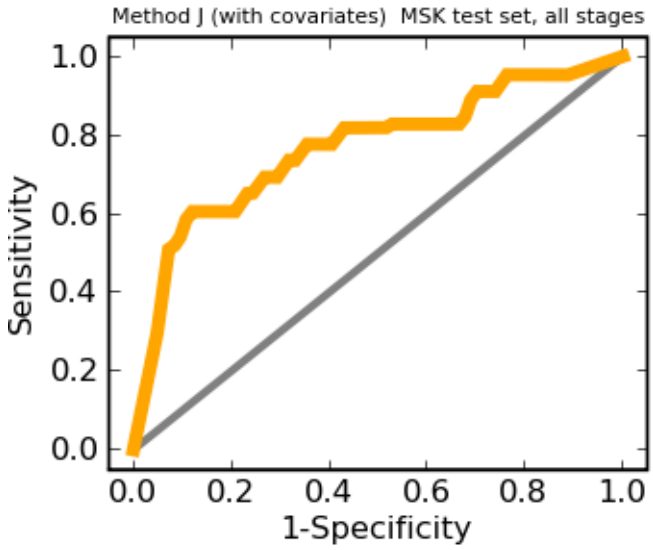

Method E2 CAN/DF test set, all stages
Sensitivity
1-Specificity
0.0
0.2
0.4
0.6
0.8
1.0

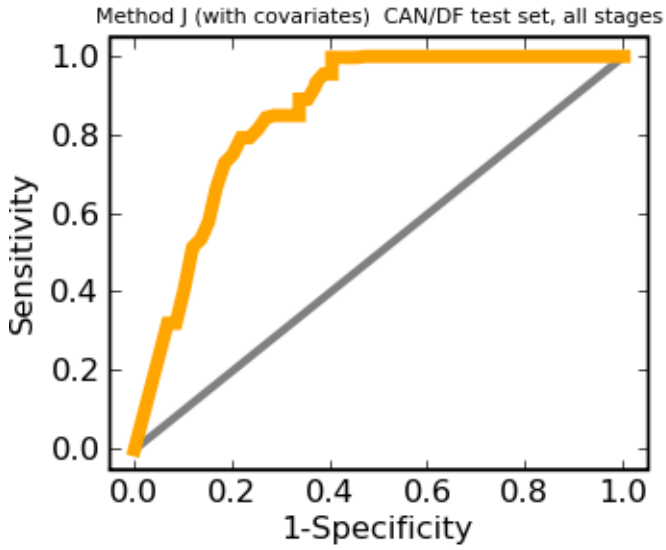

Method I MSK test set, all stages
Sensitivity
1-Specificity
0.0
0.2
0.4
0.6
0.8
1.0

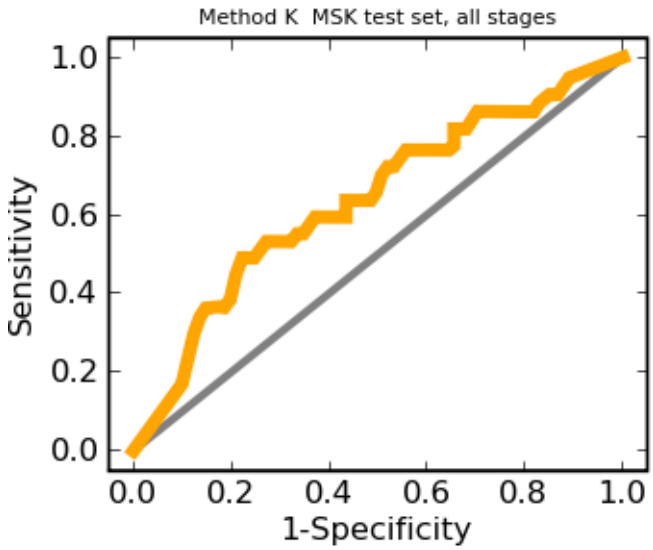

Method I CAN/DF test set, all stages
Sensitivity
1-Specificity
0.0
0.2
0.4
0.6
0.8
1.0

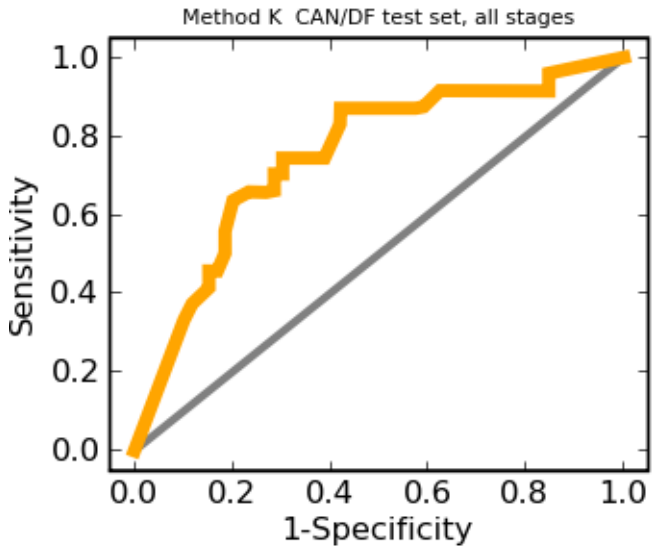

Method J MSK test set, all stages
Sensitivity
1-Specificity
0.0
0.2
0.4
0.6
0.8
1.0

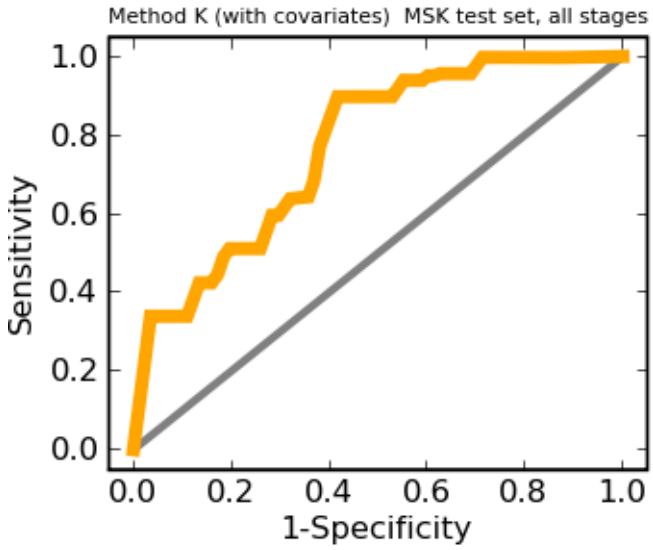

Method J CAN/DF test set, all stages
Sensitivity
1-Specificity
0.0
0.2
0.4
0.6
0.8
1.0

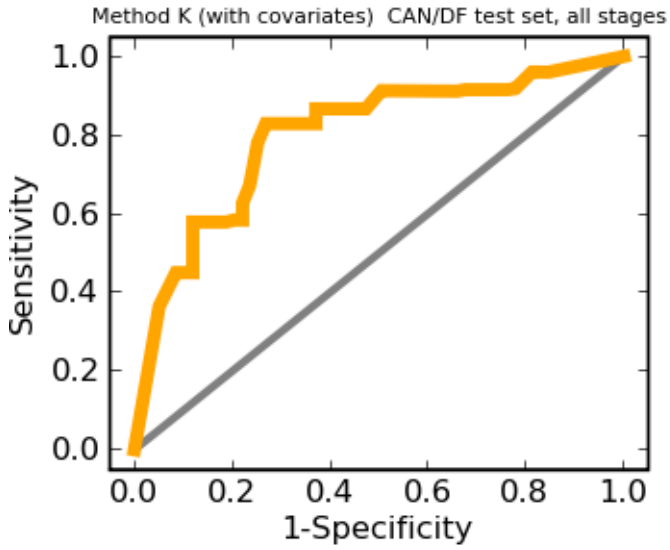

Method J (with covariates) MSK test set, all stages
Sensitivity
1-Specificity
0.0
0.2
0.4
0.6
0.8
1.0

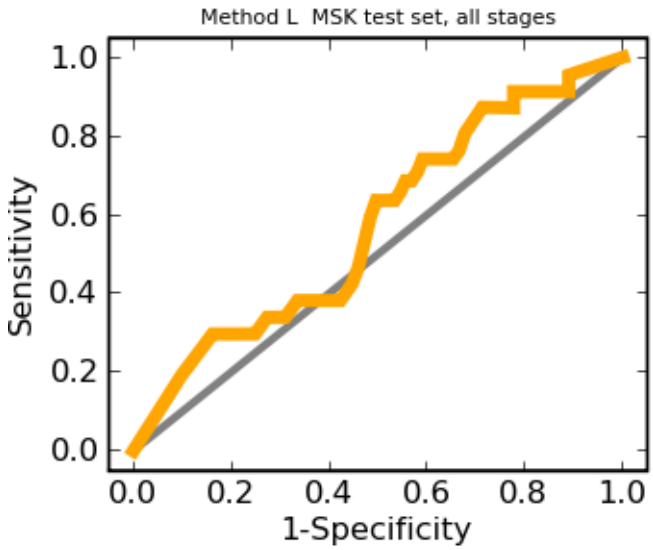

Method J (with covariates) CAN/DF test set, all stages
Sensitivity
1-Specificity
0.0
0.2
0.4
0.6
0.8
1.0

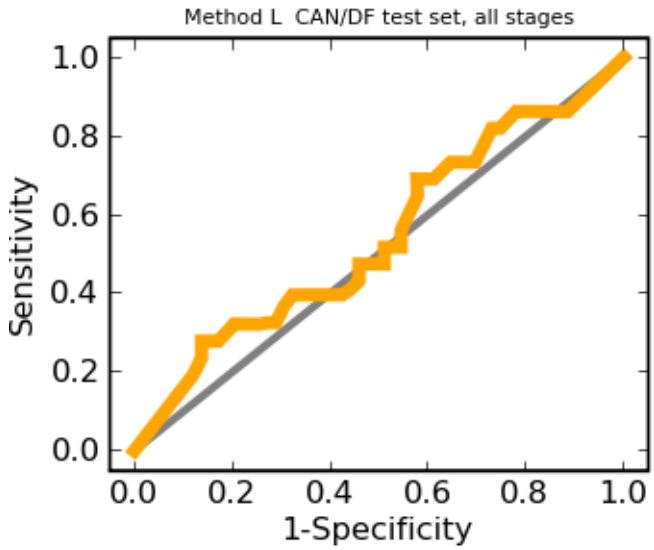

Method K MSK test set, all stages
Sensitivity
1-Specificity
0.0
0.2
0.4
0.6
0.8
1.0

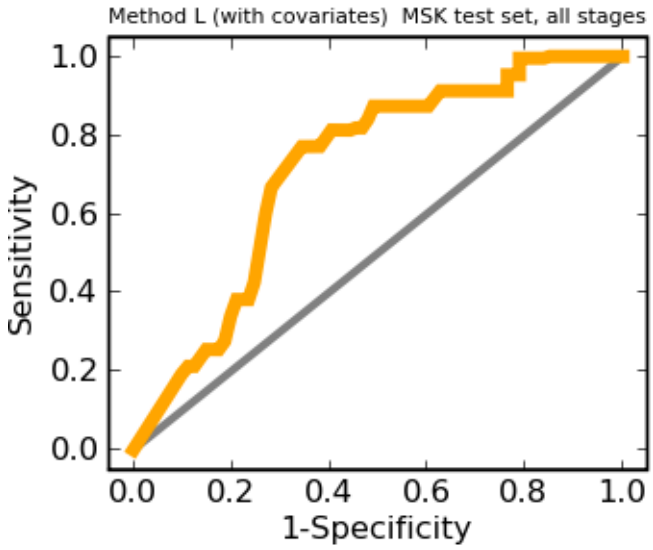

Method K CAN/DF test set, all stages
Sensitivity
1-Specificity
0.0
0.2
0.4
0.6
0.8
1.0

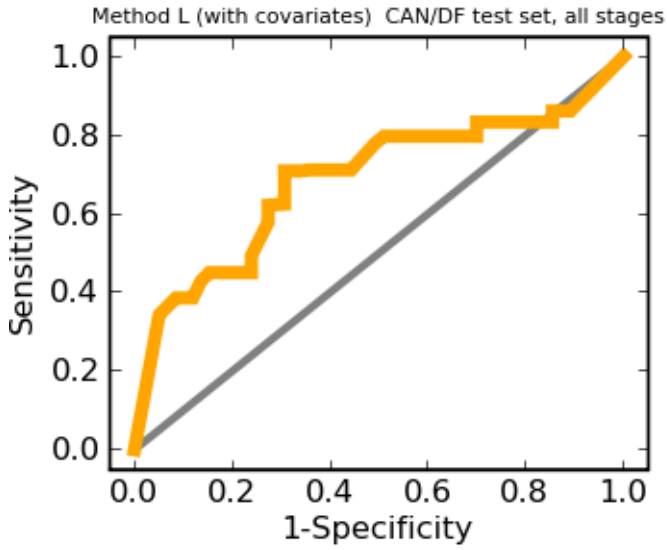

Method K (with covariates) MSK test set, all stages
Sensitivity
1-Specificity
0.0
0.2
0.4
0.6
0.8
1.0

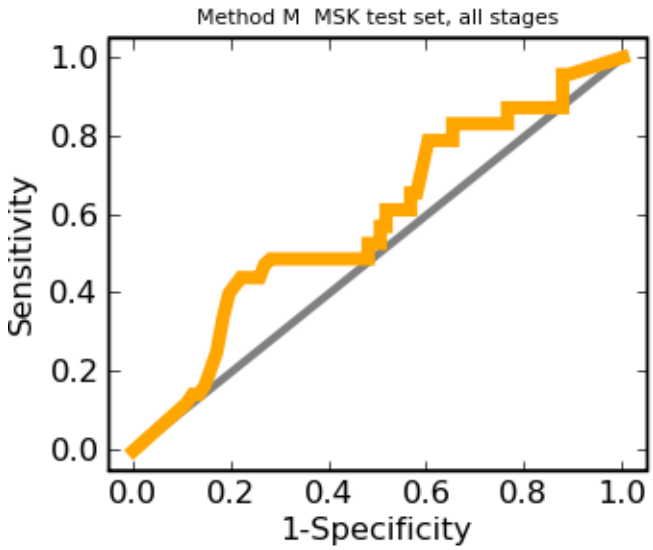

Method K (with covariates) CAN/DF test set, all stages
Sensitivity
1-Specificity
0.0
0.2
0.4
0.6
0.8
1.0

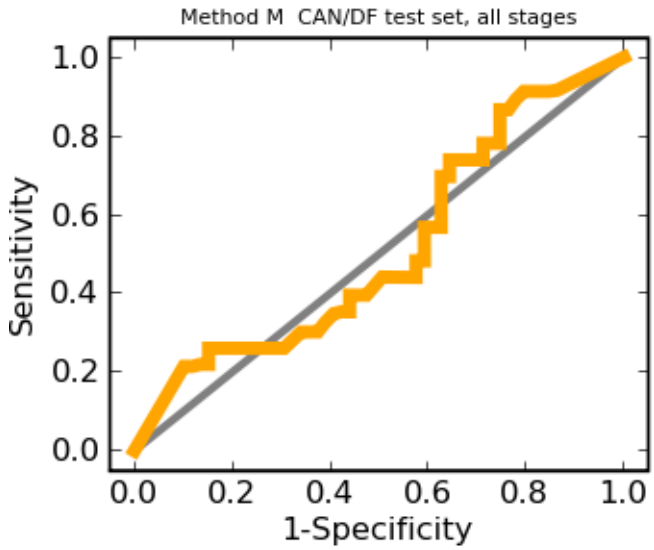

Method L MSK test set, all stages
Sensitivity
1-Specificity
0.0
0.2
0.4
0.6
0.8
1.0

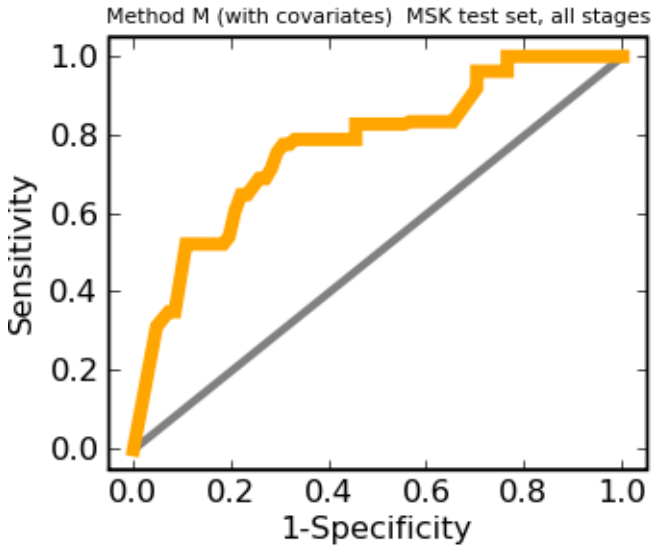

Method L CAN/DF test set, all stages
Sensitivity
1-Specificity
0.0
0.2
0.4
0.6
0.8
1.0

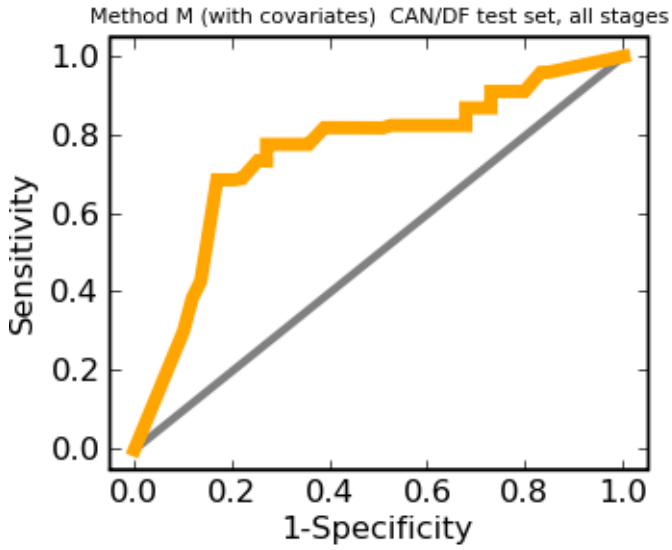

Method L (with covariates) MSK test set, all stages
Sensitivity
1-Specificity
0.0
0.2
0.4
0.6
0.8
1.0

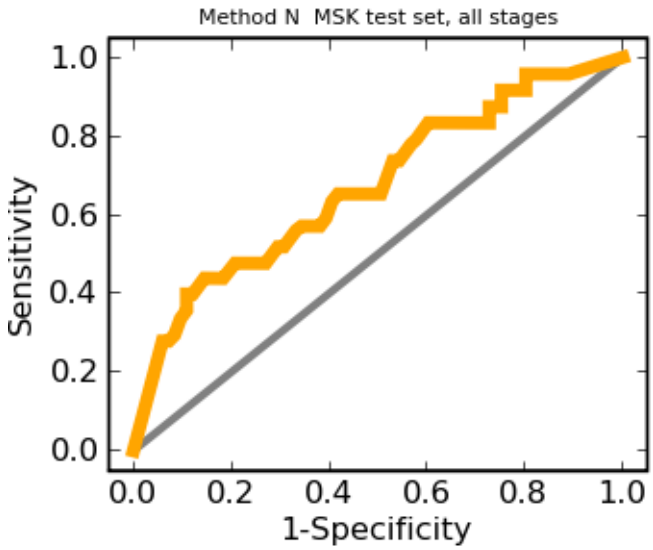

Method L (with covariates) CAN/DF test set, all stages
Sensitivity
1-Specificity
0.0
0.2
0.4
0.6
0.8
1.0

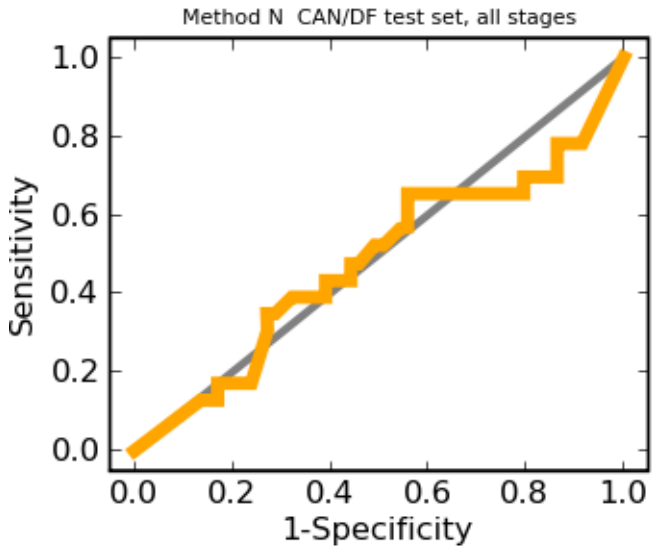

Method M MSK test set, all stages
Sensitivity
1-Specificity
0.0
0.2
0.4
0.6
0.8
1.0

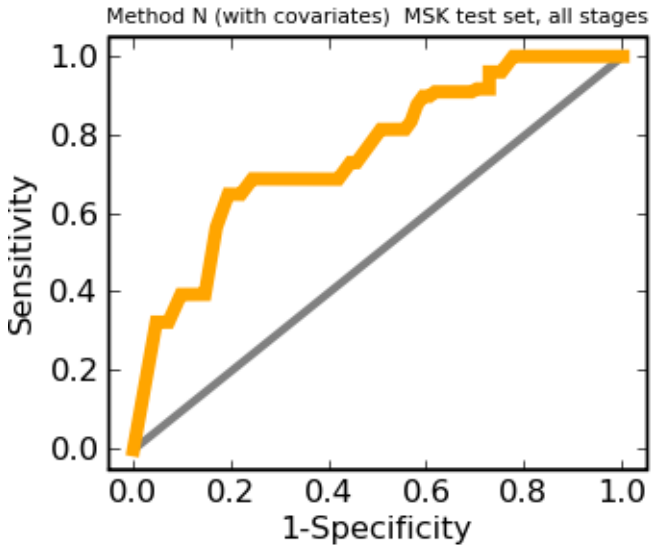

Method M CAN/DF test set, all stages
Sensitivity
1-Specificity
0.0
0.2
0.4
0.6
0.8
1.0

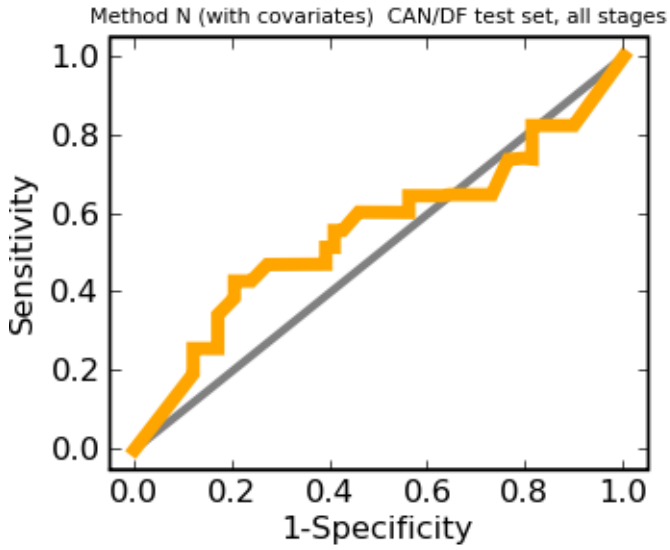

Method M (with covariates) MSK test set, all stages
Sensitivity
1-Specificity
0.0
0.2
0.4
0.6
0.8
1.0

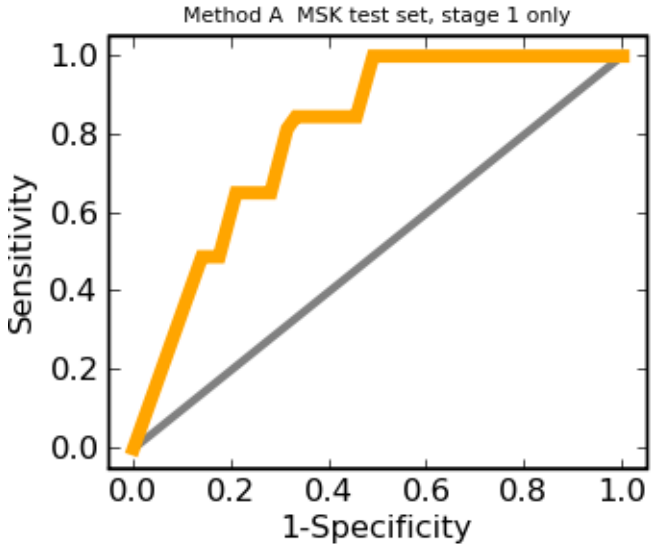

Method M (with covariates) CAN/DF test set, all stages
Sensitivity
1-Specificity
0.0
0.2
0.4
0.6
0.8
1.0

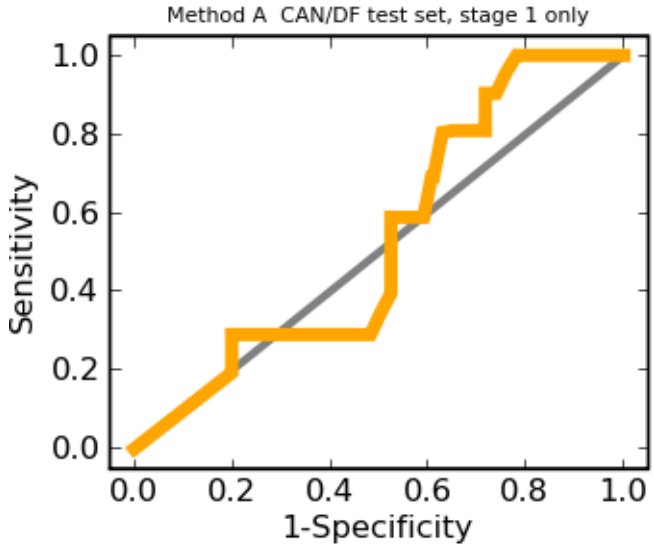

Method N MSK test set, all stages
Sensitivity
1-Specificity
0.0
0.2
0.4
0.6
0.8
1.0

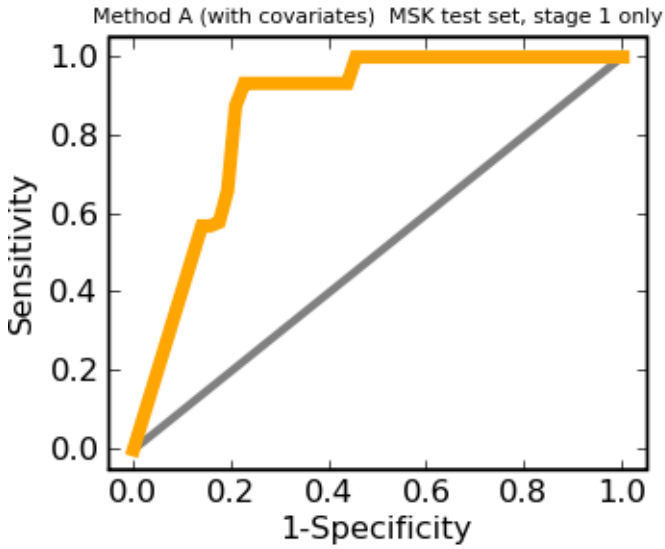

Method N CAN/DF test set, all stages
Sensitivity
1-Specificity
0.0
0.2
0.4
0.6
0.8
1.0

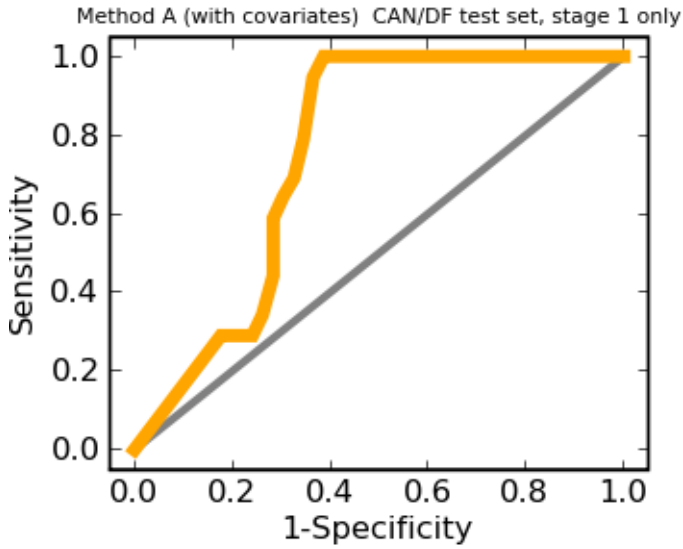

Method N (with covariates) MSK test set, all stages
Sensitivity
1-Specificity
0.0
0.2
0.4
0.6
0.8
1.0

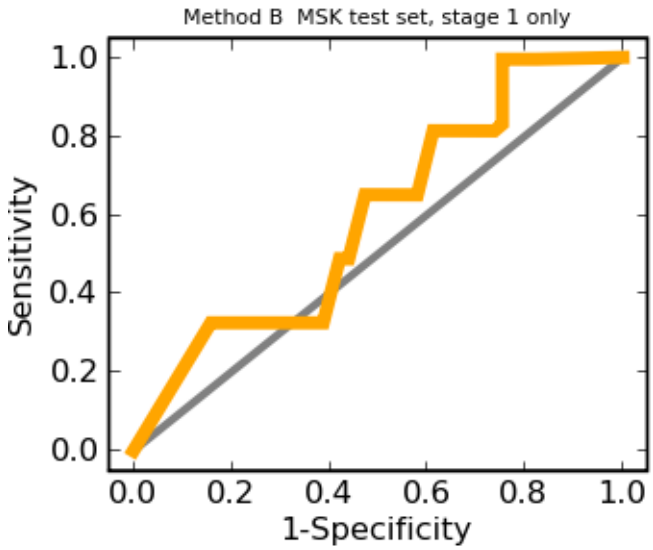

Method N (with covariates) CAN/DF test set, all stages
Sensitivity
1-Specificity
0.0
0.2
0.4
0.6
0.8
1.0

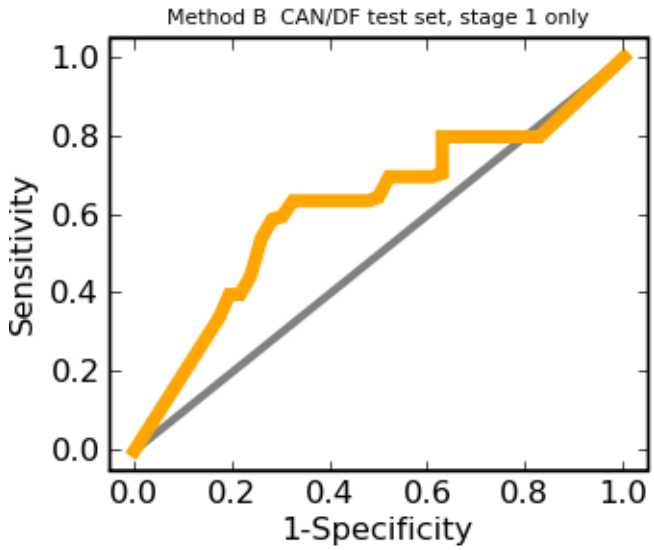

Method A MSK test set, stage 1 only
Sensitivity
1-Specificity
0.0
0.2
0.4
0.6
0.8
1.0

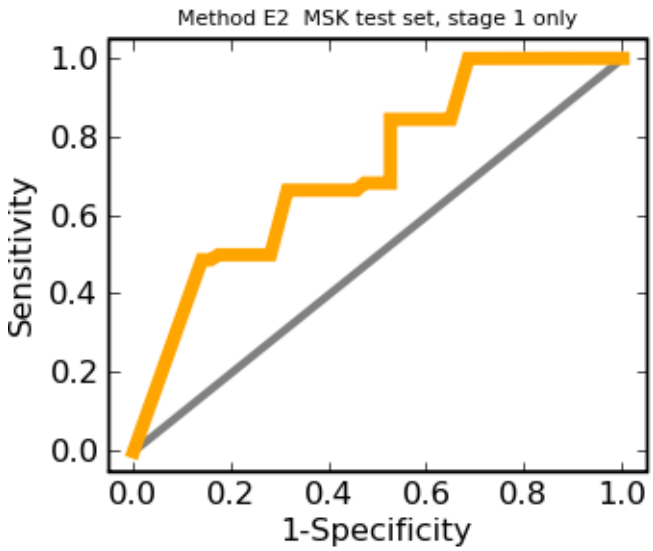

Method A CAN/DF test set, stage 1 only
Sensitivity
1-Specificity
0.0
0.2
0.4
0.6
0.8
1.0

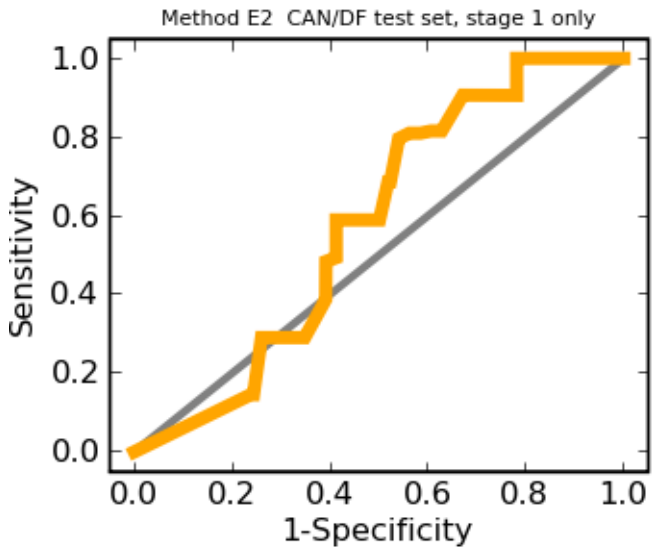

Method A (with covariates) MSK test set, stage 1 only
Sensitivity
1-Specificity
0.0
0.2
0.4
0.6
0.8
1.0

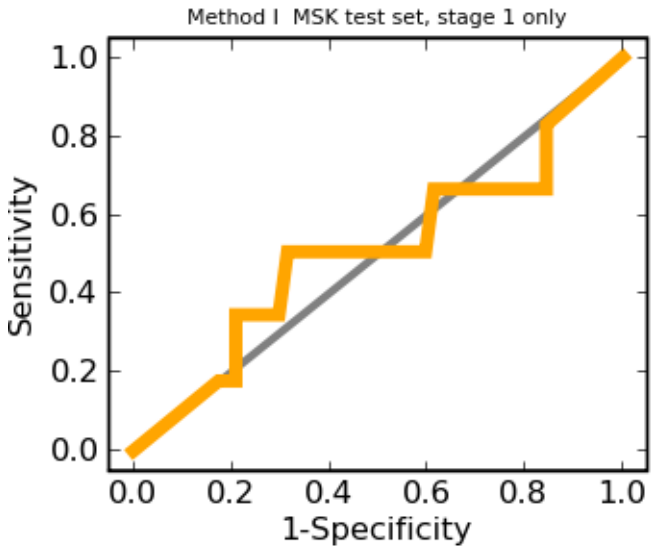

Method A (with covariates) CAN/DF test set, stage 1 only
Sensitivity
1-Specificity
0.0
0.2
0.4
0.6
0.8
1.0

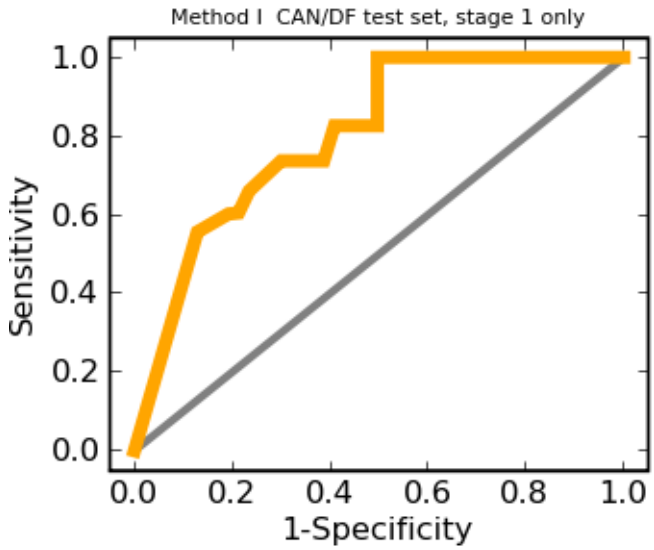

Method B MSK test set, stage 1 only
Sensitivity
1-Specificity
0.0
0.2
0.4
0.6
0.8
1.0

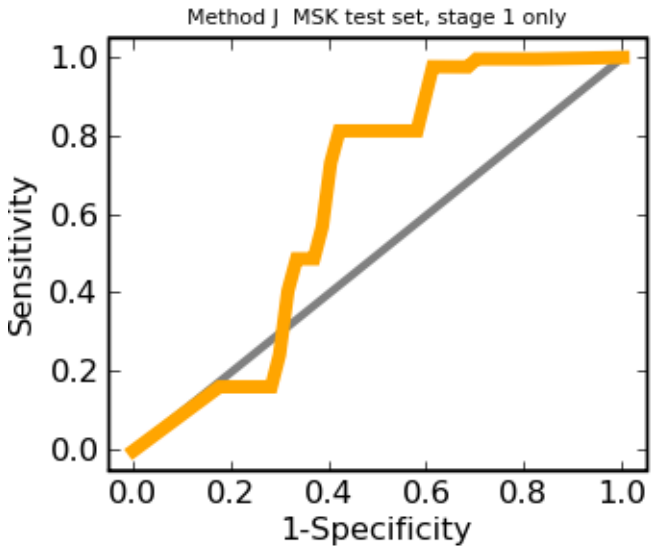

Method B CAN/DF test set, stage 1 only
Sensitivity
1-Specificity
0.0
0.2
0.4
0.6
0.8
1.0

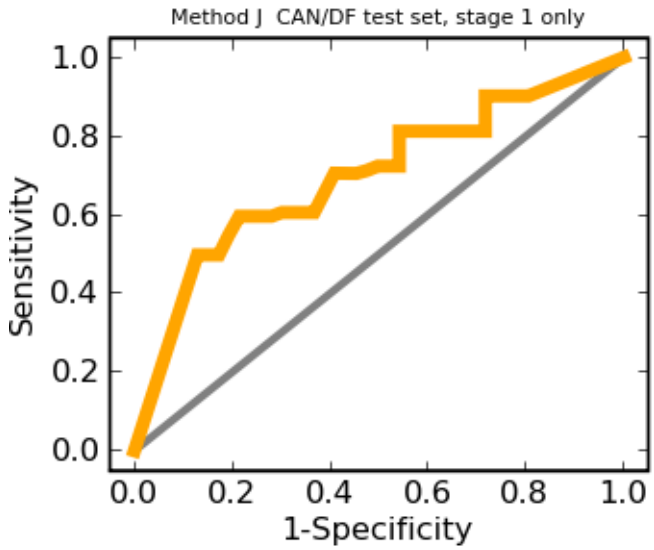

Method E2 MSK test set, stage 1 only
Sensitivity
1-Specificity
0.0
0.2
0.4
0.6
0.8
1.0

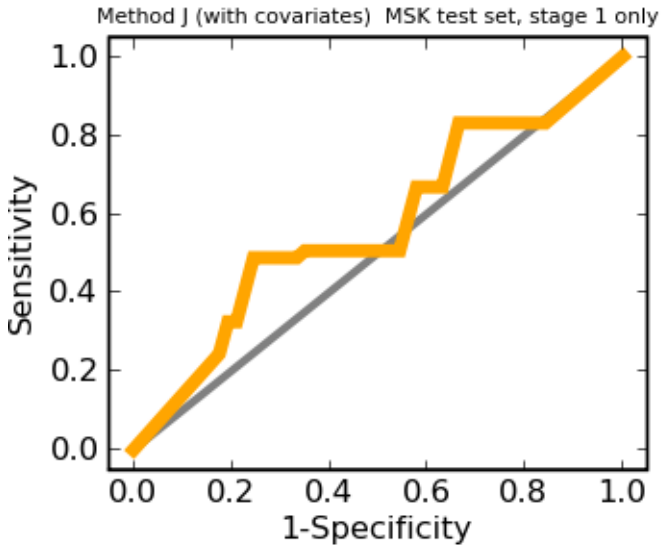

Method E2 CAN/DF test set, stage 1 only
Sensitivity
1-Specificity
0.0
0.2
0.4
0.6
0.8
1.0

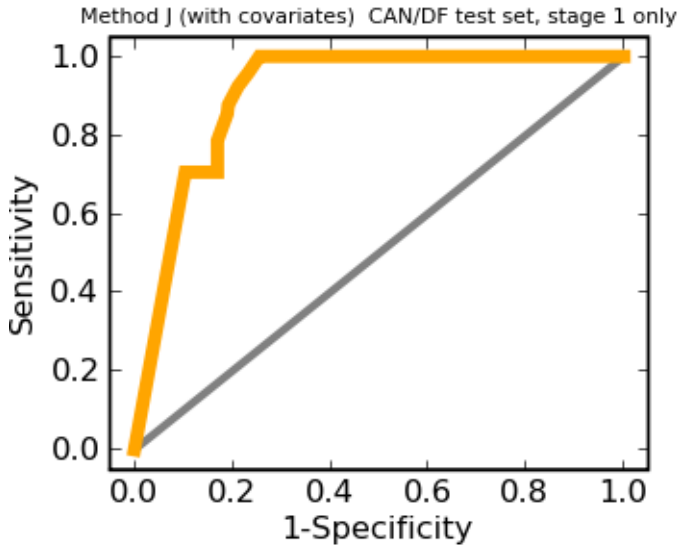

Method I MSK test set, stage 1 only
Sensitivity
1-Specificity
0.0
0.2
0.4
0.6
0.8
1.0

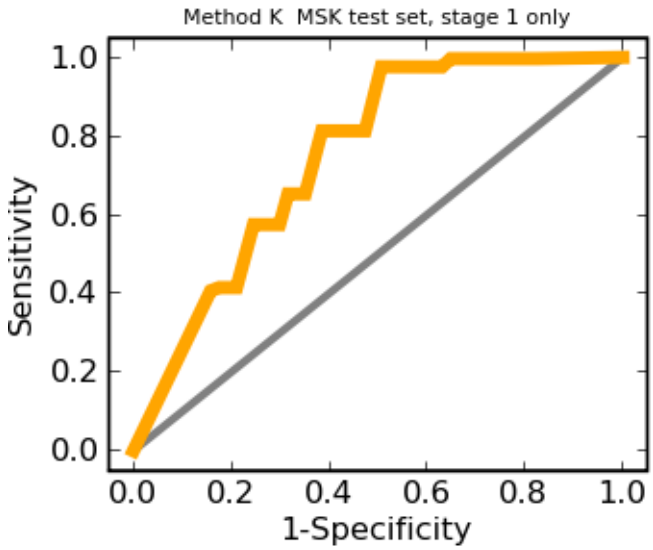

Method I CAN/DF test set, stage 1 only
Sensitivity
1-Specificity
0.0
0.2
0.4
0.6
0.8
1.0

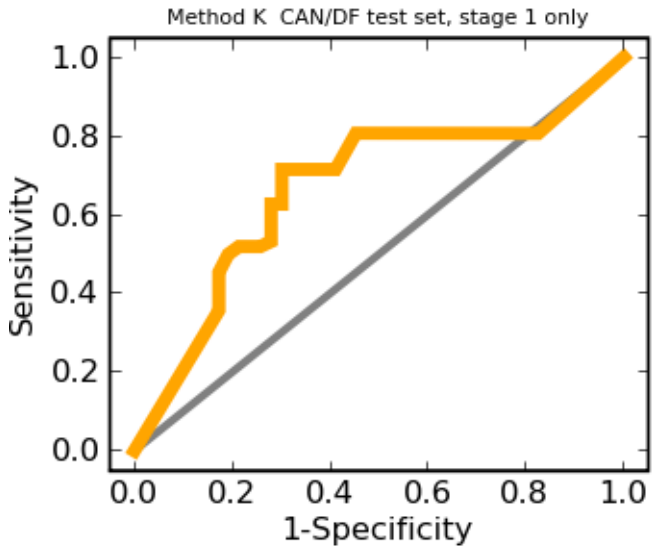

Method J MSK test set, stage 1 only
Sensitivity
1-Specificity
0.0
0.2
0.4
0.6
0.8
1.0

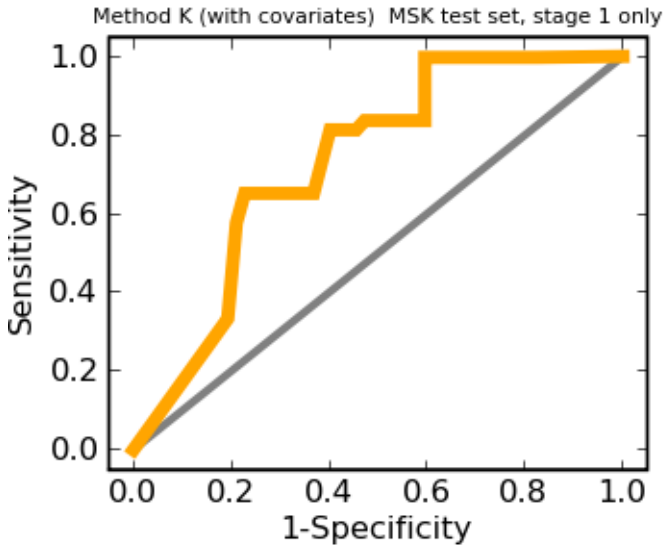

Method J CAN/DF test set, stage 1 only
Sensitivity
1-Specificity
0.0
0.2
0.4
0.6
0.8
1.0

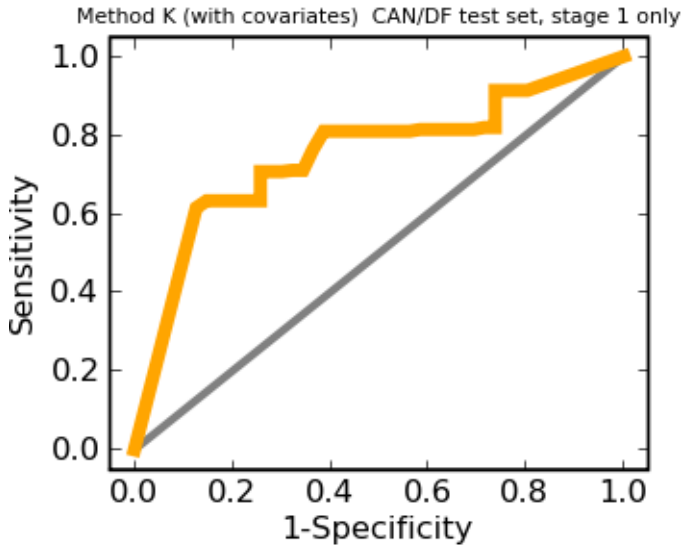

Method J (with covariates) MSK test set, stage 1 only
Sensitivity
1-Specificity
0.0
0.2
0.4
0.6
0.8
1.0

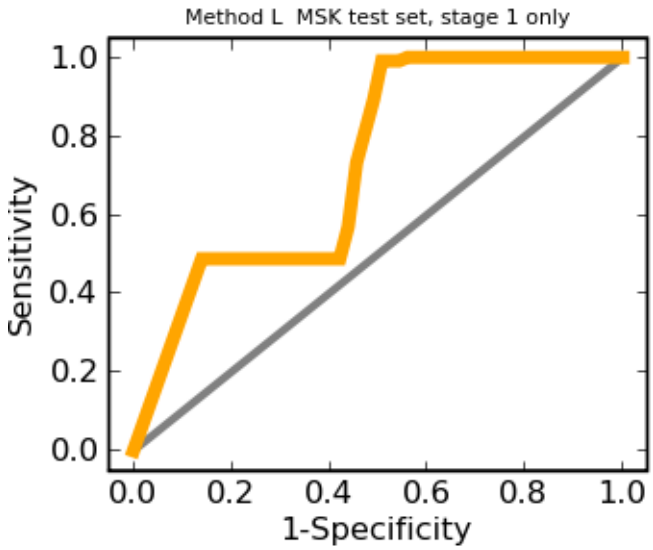

Method J (with covariates) CAN/DF test set, stage 1 only
Sensitivity
1-Specificity
0.0
0.2
0.4
0.6
0.8
1.0

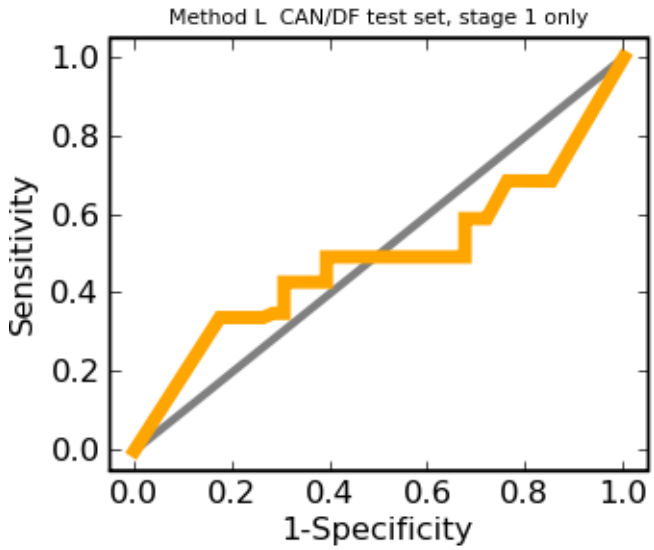

Method K MSK test set, stage 1 only
Sensitivity
1-Specificity
0.0
0.2
0.4
0.6
0.8
1.0

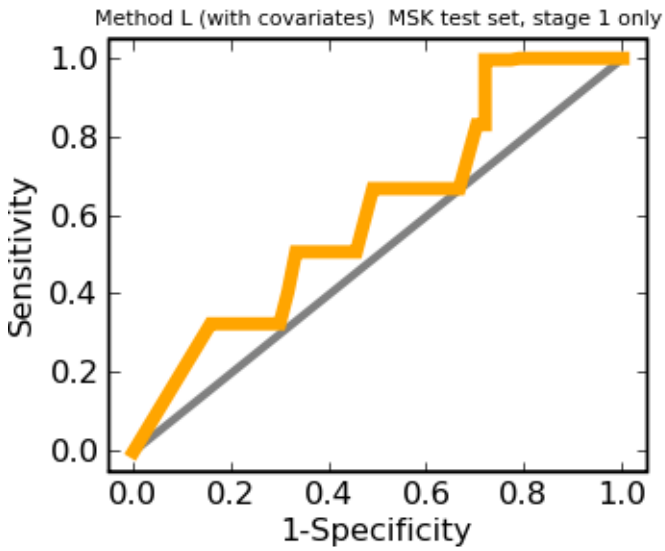

Method K CAN/DF test set, stage 1 only
Sensitivity
1-Specificity
0.0
0.2
0.4
0.6
0.8
1.0

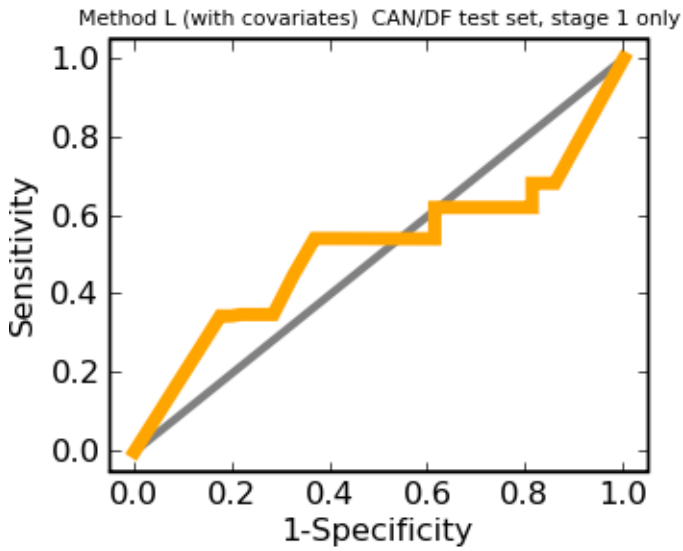

Method K (with covariates) MSK test set, stage 1 only
Sensitivity
1-Specificity
0.0
0.2
0.4
0.6
0.8
1.0

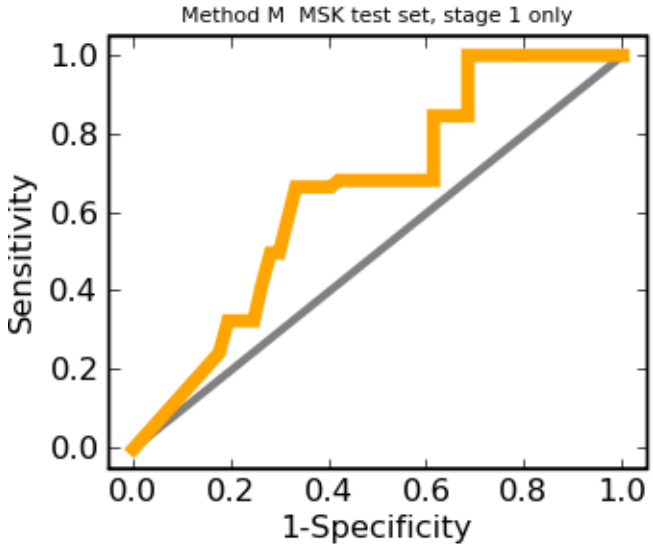

Method K (with covariates) CAN/DF test set, stage 1 only
Sensitivity
1-Specificity
0.0
0.2
0.4
0.6
0.8
1.0

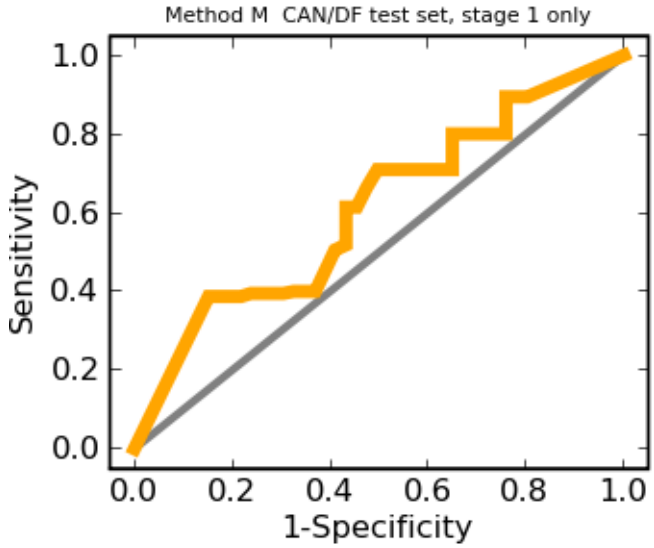

Method L MSK test set, stage 1 only
Sensitivity
1-Specificity
0.0
0.2
0.4
0.6
0.8
1.0

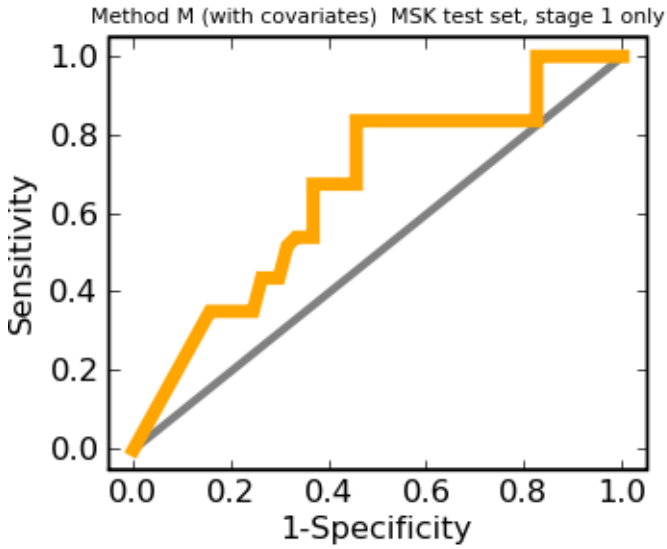

Method L CAN/DF test set, stage 1 only
Sensitivity
1-Specificity
0.0
0.2
0.4
0.6
0.8
1.0

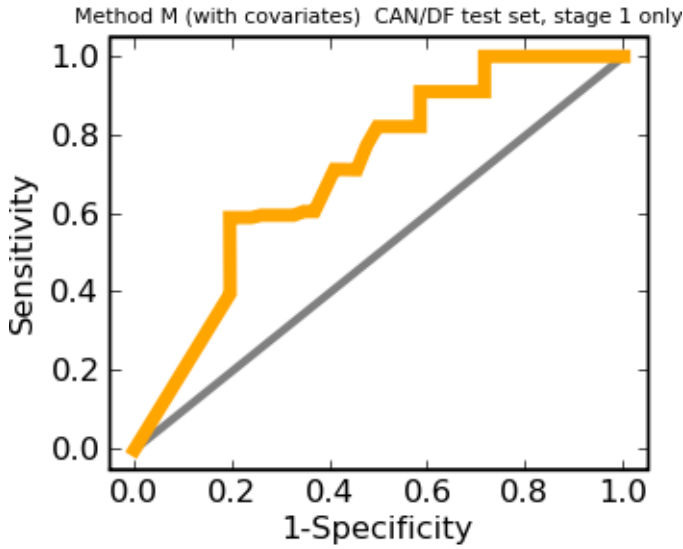

Method L (with covariates) MSK test set, stage 1 only
Sensitivity
1-Specificity
0.0
0.2
0.4
0.6
0.8
1.0

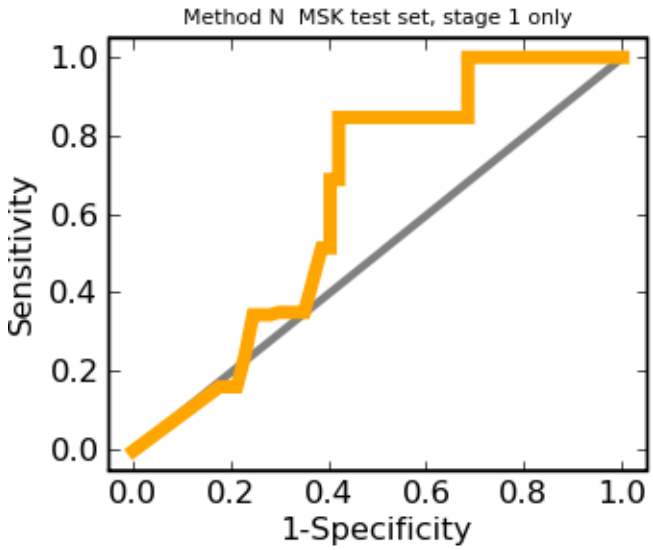

Method L (with covariates) CAN/DF test set, stage 1 only
Sensitivity
1-Specificity
0.0
0.2
0.4
0.6
0.8
1.0

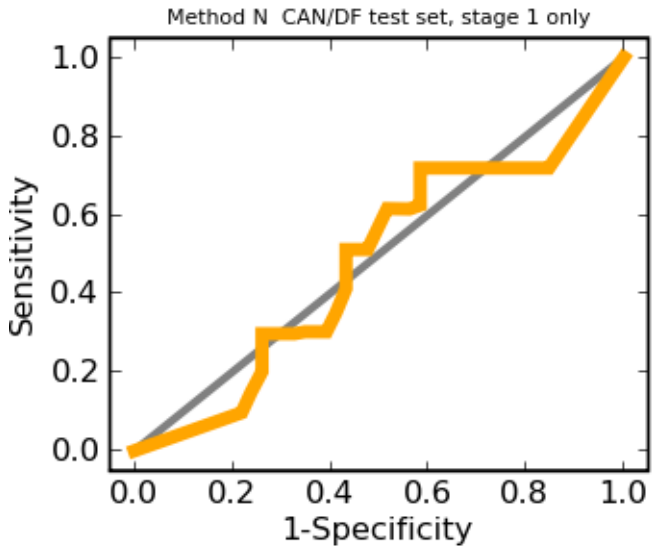

Method M MSK test set, stage 1 only
Sensitivity
1-Specificity
0.0
0.2
0.4
0.6
0.8
1.0

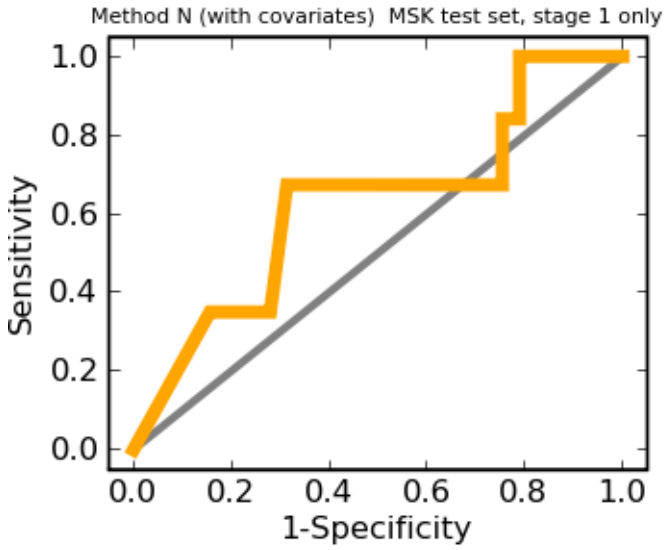

Method M CAN/DF test set, stage 1 only
Sensitivity
1-Specificity
0.0
0.2
0.4
0.6
0.8
1.0

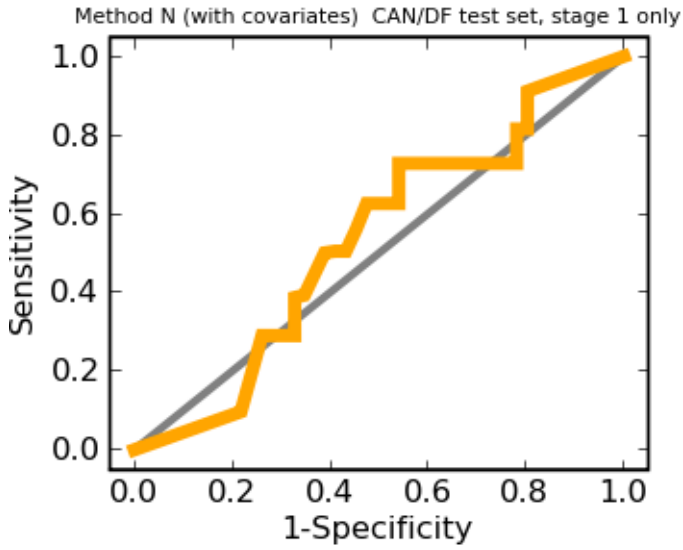

Method M (with covariates) MSK test set, stage 1 only
Sensitivity
1-Specificity
0.0
0.2
0.4
0.6
0.8
1.0


Method M (with covariates) CAN/DF test set, stage 1 only
Sensitivity
1-Specificity
0.0
0.2
0.4
0.6
0.8
1.0


Method N MSK test set, stage 1 only
Sensitivity
1-Specificity
0.0
0.2
0.4
0.6
0.8
1.0


Method N CAN/DF test set, stage 1 only
Sensitivity
1-Specificity
0.0
0.2
0.4
0.6
0.8
1.0


Method N (with covariates) MSK test set, stage 1 only
Sensitivity
1-Specificity
0.0
0.2
0.4
0.6
0.8
1.0


Method N (with covariates) CAN/DF test set, stage 1 only
Sensitivity
1-Specificity
0.0
0.2
0.4
0.6
0.8
1.0

Supplementary Materials 1: Detailed Materials and Methods

Investigator consortium: Four institutions (University of Michigan Cancer Center (UM), Moffitt Cancer Center (HLM), Memorial Sloan-Kettering Cancer Center (MSK) and the Dana-Farber Cancer Institute (DFCI)) formed a consortium with support and collaboration of NCI investigators to develop and validate gene expression signatures of lung adenocarcinomas. Lung adenocarcinoma specimens were obtained from the following sources: UM (100 samples), University of Minnesota VA/CALGB (77 samples), HLM (79 samples), MSK (104 samples), and Toronto/DFCI (82 samples). The Toronto/DFCI 82 samples included 43 samples from University Health Network (UHN) and 39 samples from National Cancer Institute of Canada Clinical Trials Group (NCIC CTG). A common protocol and common reagents were previously developed and used for blinded data comparability study that has been published (Dobbin et al, 2005). Using this validated protocol, samples were processed and analyzed at four sites: UM analyzed samples from UM and VA/CALGB, the HLM and MSK analyzed samples from their own institutions, and UHN/NCIC CTG samples processed at DFCI (from hereon referred as CAN/DF samples).

Hypotheses: Prior to the analyses or having the clinical, pathology or outcome data available, we developed an approach for unbiased evaluation of the data to address four specific hypotheses. The specific hypotheses are that overall survival can be predicted in four settings: 1) using gene expression data for subjects of all AJCC stages, 2) using gene expression and clinical data for subjects of all AJCC stages, 3) using gene expression data for stage 1 AJCC subjects, and 4) using gene expression and clinical data for stage 1 AJCC subjects. Note that for hypotheses where evaluation is on stage 1 subjects, there was no restriction on the stages included in the training set (methods could use only stage 1 samples for training, or all samples for training).

Sample criteria and clinical data: Nearly 505 adenocarcinomas of stages I, II and III were collected with approximately 60% of the samples represented by stage I tumors. Tumors were collected by surgical resection from patients who have provided consent and protocols were approved by the Institutional Review Boards (IRB-Med) of the respective institutions. None of the patients received preoperative chemotherapy or radiation and least two years of follow-up information was available. Study pathologists at each of the four sites reviewed both the tumor permanent sections and the frozen sections of the samples considered for inclusion in this study. An initial review was performed to identify stage, diagnosis, and the regions of the frozen section containing the maximum tumor cellularity for macrodissection. Regions containing a minimum of 60% tumor cellularity were required, and in most instances tumor cellularity of at least 70-90% was identified for inclusion in the sample for RNA isolation. A second pathological review of each case was performed to coordinate the pathological information across all tumor collection sites. This collaborative re-evaluation of each tumor sample was used to verify the initial pathological findings and where the second pathological review disagreed with the first pathological diagnosis the samples were excluded. The extensive pathological review is also being used to produce an extensive pathology database for each sample that includes information regarding tumor

differentiation, subtype, tumor size, and presence of lymphocytic infiltration, angio-lymphatic invasion, pleural surface involvement, and positive bronchiolar margins. This database, including representative digital photographs taken of each specimen's paraffin section and the frozen section used for RNA isolation will be made available pending final review. Clinical information was also collected about tumor staging, history of prior cancers, lymph node involvement by lymph node sampling/dissection, smoking history, age, gender, operation type, last follow-up date, and patient's status at last follow-up. Additional valuable information, such as CXR and CT scans, pulmonary function tests, time and site of recurrence, and adjuvant treatment information was collected when available. Survival outcome was determined and most patients have reliable long-term follow-up. For the analyses in this study patients were censored after 60 months of follow-up and others excluded for analysis if death occurred within 1 month of operation. Three subjects, all in the UM/HLM training set, were excluded from the analysis as they died within 1 month of operation. The clinical data collected from each individual site underwent two forms of review. A core set of critical covariates were identified for thorough quality assessment. These were reviewed by an NCI statistician to evaluate discrepancies in coding patterns between sites, logical inconsistencies in the data, and missing data. The complete dataset underwent a second round of quality review by the CALGB Statistical Center Data Operations. The clinical dataset was finalized and locked in December, 2006.

The planned total sample size was 600 tumors. Sample size determination was based on obtaining 95% power to identify a gene with hazard ratio of 1.5 associated with unit change in standardized gene expression (defined as expression values that have been standardized to have variance 1 for that gene). Assumes one-sided significance level of 0.001 used to select genes for inclusion in the predictor, and 28% event rate in the population. Additional adjustments for 1) a multiple correlation coefficient of 0.15 between the gene of interest and the clinical covariates (using variance inflation factor method of Hsieh and Lavori, (2000), and 2) 17% sample loss due to various causes.

Multiple comparison adjustment in the context of predictor development involves a tradeoff. Very strict Bonferroni-type adjustment may reduce predictor performance by eliminating informative genes. Too lax adjustment may reduce predictor performance by inclusion of too many noise genes. It was thought that the significance level used in the sample size calculation represented a reasonable tradeoff between the two. A total of 486 tumor samples were arrayed and a total of 442 samples were retained after the QC evaluation of the clinical, pathological, and array data.

Isolation of RNA: Frozen dissected tumor tissue was immersed in 1 ml of Trizol Reagent (Invitrogen Corporation, Carlsbad, CA). Tissue sections were disrupted with a glass homogenizer or glass beads to facilitate dissolution in the Trizol reagent as necessary. Purification of the RNA from the Trizol dissolved samples followed the manufacturer's protocol. The Trizol purified RNA was further purified using the RNeasy columns and the manufacturer's cleanup procedure (Qiagen Inc., Valencia, CA). The quality of total RNA was assessed by electrophoretic analysis on the Agilent 2100 Bioanalyzer. The RNA yield was determined by OD_{260} measurement.

Preparation of Labeled RNA Targets for Hybridization: Beginning with total RNA the poly(A) RNA pool was specifically converted to cDNA and then amplified and labeled with biotin following the procedure initially described by Van Gelder *et al.*. First-strand cDNA synthesis was carried out using the Superscript Choice System (Invitrogen Corporation, Carlsbad, CA) and the T7 promoter/oligo (dT) primer (5'-GGCCAGTGAATTGTAATACGACTCACTATAGGGAGGCGG-$(dT)_{24}$-3'), (T7-Oligo(dT) Promoter Primer Kit, Affymetrix Corporation, #900375, Affymetrix Corporation, Santa Clara, CA). Following annealing the rest of the cDNA synthesis reaction was prepared such that the final reaction contains 5 mg RNA, 100 pmol T7-(T)24 primer, 500 mM each dNTP, 10 mM DTT, 50 mM Tris-HCl, pH 8.3, 75 mM KCl, 3 mM $MgCl_2$, and 200 U of Superscript II reverse transcriptase (Invitrogen). The reaction was incubated for 1 hr at 42°C. A second-strand cDNA synthesis was then performed at 16°C for 2 hr in a total volume of 150 mL, using 10U of *E.coli* DNA ligase, 40 U of *E. coli* DNA polymerase I, and 2 U of *E. coli* RNase H in the presence of 200 mM of each dNTP, 10 mM NH_4SO_4, 1.3 mM DTT, 26.7 mM Tris-HCl, pH 7.0, 100 mM KCl, 5 mM $MgCl_2$, and 150 mM b-NAD^+. All second strand components were purchased from the Invitrogen Corporation. Following the second-strand DNA synthesis, 10 U of T4 DNA Polymerase (Invitrogen) was added and the samples were incubated an additional 5 min at 16°C. The reaction was stopped by the addition of 0.5 M EDTA and the nucleic acids were purified using the Affymetrix GeneChip sample clean-up modules (#900371, Affymetrix). The nucleic acids were eluted with 25 ml of DEPC treated water. Twenty-two microliters of the purified cDNA was used as the template for a transcription reaction performed with the BioArray™ HighYield™ RNA Transcript Labeling Kit according to manufacturer's instructions (ENZO Life Sciences, New York, NY) which incorporates biotinylated UTP into the transcripts. The Biotin-labeled RNA was purified using RNeasy columns (Qiagen Inc., Valencia, CA) and fragmented to a size of 35 to 200 bases by incubating at 94^0 C for 35 minutes in fragmentation buffer (40 mM Tris-acetate, pH 8.1/100 mM potassium acetate/30 mM magnesium acetate). The integrity of the starting material and the products of each reaction were monitored on agarose gels to assess the size distribution of the products and compare them to the starting material.

Array hybridization and scanning. The hybridization solutions consisted of 20 µg of fragmented RNA and 0.1 mg/ml sonicated herring sperm DNA, in 1x MES buffer (containing 100 mM MES, 1 M Na^+, 20 mM EDTA, and 0.01% Tween 20). In addition the hybridization solutions were spiked with known concentrations of RNA from the bacterial genes, BioB, BioC, and BioD, and one phage gene, Cre, as hybridization standards. The hybridization mixtures was heated to 99°C for 5 min followed by incubation at 45°C for 5 min before injection of the sample into an Affymetrix GeneChip array cartridge 133A. All hybridizations were carried out at 45°C for 16–17 h with mixing on a rotisserie at 60 rpm. Following hybridization, the solutions were removed and the arrays were rinsed with 1x MES. Subsequent washing and staining of the arrays was carried out using the GeneChip Fluidics station protocol EukGE_WS2: This consists of 10 cycles of 2 mixes per cycle with non-stringent wash buffer (6x SSPE, 0.01% Tween 20) at 25°C followed by 4 cycles of 15 mixes per cycle with stringent wash buffer (100

mM MES, 0.1 M Na^+, and 0.01% Tween 20) at 50°C. The probe arrays were then stained for 10 min in streptavidin-phycoerythrin solution (SAPE) [1x MES solution, 10 µg/ml SAPE (Molecular Probes, Eugene, OR), and 2 µg/µl acetylated BSA (Invitrogen)] at 25°C. The post-stain wash was 10 cycles of 4 mixes per cycle at 25°C. The probe arrays were treated for 10 min with an antibody solution [1x MES solution, 2 µg/µl acetylated BSA, 0.1 µg/µl normal goat IgG (Sigma Chemical, St. Louis, MO), 3 µg/µl biotinylated goat-anti-streptavidin antibody, (Vector Laboratories, Burlingame, CA)] at 25°C followed by a second staining for 10 min in streptavidin-phycoerythrin solution (SAPE) [1x MES solution, 10 µg/ml SAPE (Molecular Probes, Eugene, OR), and 2 µg/µl acetylated BSA (Invitrogen)] at 25°C. The final wash was 15 cycles of 4 mixes per cycle at 30°C with non-stringent wash buffer. Following washing and staining, probe arrays was scanned once at 1.56 µm resolution using the Affymetrix GeneChip Scanner 3000 or at 3 µm resolution using the Affymetrix GeneChip Scanner 2500.

The cel files for the study are available at the following URL:
https://caarraydb.nci.nih.gov/caarray/publicExperimentDetailAction.do?ex
pId=1015945236141280
Links to the pathology and clinical data are also available at this site.

Data Preprocessing: Arrays were visually scanned for any defects or scanning artifacts that might compromise the final results. The CEL file data from each of the four generation sites was then transferred to NCI where an honest broker system was set up to anonymize the data for further work. This CEL file data and the MAS 5.0 generated values calculated from the CEL files were used to evaluate the quality of individual arrays using metrics such as scaling factor, 5/3' ratios, and histograms of spot intensities. From a combined analysis by two of the sites a small number of arrays were excluded under the suspicion of poor quality array data. During this initial evaluation of the array data principal components analysis suggested that the data generated at the UM, HLM and MSK sites were quite similar, but that the data from the DF site was different probably due to slightly weaker average signals across the arrays and slightly greater variation across all samples. It was decided that the combined data from UM and HLM would be used as the training set, with the data from MSK held out as an external validation dataset containing similar microarray data and the data from the DF site held out as a second and more challenging external test set. All CEL files generated at the four sites were quantile normalized as a group using the array NCI_U133A_61L as a reference file. The quantile normalized CEL files were imported into the software dChip (Li et al, 2001) (Build version February 2006) and signal values for each probeset were calculated using the default settings. The calculated microarray data was then combined with the clinical data for each sample and assigned a sample number by the honest broker (one of the NCI collaborators).

Overall, 44 of the 486 arrayed samples (including one Stratagene reference sample) were excluded from the study based on violations of the clinical criteria that were discovered after the initial sample was defined, or due to poor array quality. Eleven of the patient samples, and the single excluded Stratagene sample, were excluded due to quality-control problems with the microarray data, with the remaining 33 samples excluded due to

violations of clinical and pathological criteria.

Protocol for classifier development: Four teams of investigators, corresponding to four of the sites contributing samples (HLM, MSK, CAN, and UM) obtained the training data from all sites for classifier development from the honest broker. Each team was encouraged to develop a single classifier (assigning the samples to discrete risk groups) or risk scoring procedure (assigning the samples to points on a risk continuum) for each of the four study hypotheses (see methods). For simplicity, we often use the term "classifier" to refer to either a categorical or continuous outcome prediction. The categorical procedures are denoted explicitly with subscript "cat" in the main figures. There was no stipulation that all training samples be used equally, so, for example, it was permitted to train only on the stage 1 samples, or only on the samples from a single institution. The teams used only the training data to define the classifiers, and the test data was not incorporated in any way. Note however that we did permit the classifier to be defined in terms of standardized test set expression levels, or in terms of percentile points of test set gene expression features. This was done to reflect a standardization and calibration process that would surely be part of any clinical implementation of this technology. Although some of the teams presumably had access to the test data derived from their own institution (never the complete test data set), all teams were on the honor system to not use this data in training their classifier. By a specified date, each team provided documentation of their proposed approach to the honest broker. At this point, the test set expression data and clinical covariates (age, stage, and gender), recoded with generic identifiers, was released by the honest broker. Each team then applied its method(s) to the test data in order to predict the outcome for each sample. These results were collected by the honest broker, and the results were released together with the actual outcomes to all co-authors for performance evaluation.

References for Supplementary Materials 1:

Dobbin KK, Beer DG, Meyerson M, Yeatman TJ, Gerald WL, Jacobson JW, Conley B, Buetow KH, Heiskanen M, Simon RM, Minna JD, Girard L, Misek DE, Taylor JM, Hanash S, Naoki K, Hayes DN, Ladd-Acosta C, Enkemann SA, Viale A, Giordano TJ. Interlaboratory comparability study of cancer gene expression analysis using oligonucleotide microarrays. *Clin Cancer Res*. 2005; Jan 15;11(2 Pt 1):565-72.

Hsieh FY, Lavori PW. Sample-size calculations for the Cox proportional hazards regression model with nonbinary covariates. Control Clin Trials 2000;21(6):552-60.

Van Gelder RN, von Zastrow ME, Yool A, Dement WC, Barchas JD, Eberwine JH. Amplified RNA synthesized from limited quantities of heterogeneous cDNA. Proc Natl Acad Sci U S A 1990;87(5):1663-7.

Li C, Wong W. Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. Proc Natl Acad Sci USA. 2001;98:31–6.

Supplementary Materials 2: Classifier Development and Applications

Method A (Gene clusters and ridge regression)

Samples used in developing the classifier:

All UM and HLM samples in the training set, for both the prediction of all stages, and the prediction of stage 1 samples

Initial data processing and filtering:

All probe sets for which fewer than five samples showed greater than 50 raw expression units were removed. For the remaining probesets all values less than 0 were set to zero, all values were increased by 1, and then all the values were $\log_2$ transformed. Standard deviations were calculated for each probe set and one quarter of the probe sets with the smallest standard deviations were removed. An additional 274 probesets were removed that had a greater than 2 fold difference in expression between the UM samples and the HLM samples. This resulted in 13,838 probe sets for subsequent analysis. Expression levels within each of the four sample sets (UM, HLM, MSK, and CAN/DF) were mean centered at the probe-set level.

Gene Selection or feature reduction:

The gene expression data was reduced to 100 features per sample with a clustering algorithm. This algorithm selected a set of 100 "central probe sets" from among the actual probe sets, with the goal of maximizing the correlation between each individual probe set and its closest central probe set. Specifically, a gain function was constructed as follows: first, for each probeset, the central probeset to which it is most correlated was identified. Then, these maximizing correlation coefficients were summed over all genes. An iterative stochastic search was used to maximize this gain function. At each round of optimization a random subset of the current central probesets were reassigned to different probesets. The size of the subset is Poisson distributed with mean 5, and the reassignment was done by sampling uniformly and without replacement from the non-central probesets. If the new configuration has higher value under the gain function than the current probeset, the new configuration becomes the current configuration. Approximately 400 iterations of this process were sufficient for the search to approximately stabilize at a local solution. Once the central probe sets were identified, each of the 13,838 probe sets was assigned once more to its closest central probe set, producing a set of 100 probe set clusters. The expression levels within each probe set were then transformed across samples to have unit variance. Finally, for each sample the average of these standardized expression levels within each probe set cluster was calculated to be used as a vector of 100 risk-scoring features. C code implementing this algorithm is available by e-mail from one of the authors (kshedden@umich.edu).

Classifier:

Ridged Cox proportional hazards models were used to define linear risk scores based on the training data. A linear risk score is an expression $R = \beta_0 + \Sigma \beta_k X_k$, where X_k is the value of gene expression feature *k* in a given sample. To construct the risk scores, a penalized likelihood $L(\beta) - \lambda \|\beta_{-0}\|^2$ was numerically maximized, where $L(\beta)$ is the partial likelihood for the Cox model and β_{-0} is the vector of Cox slopes with the intercept omitted. A ridging weight (λ) of 35 was used based on a modified Akaike Information Criterion that accounts for the degrees of freedom lost by ridging. Risk scores were defined either by gene expression alone, or by a combination of gene expression and clinical measures (age, sex, and AJCC stage), all appearing as additive terms in the ridged Cox model. Stage was coded as a 3-level factor (1, 2, and 3/4 together). A ridging weight of 35 was also used in this case.

Class prediction or risk calculation for the test data:

The risk scoring procedures developed on the training data were applied to both test sets. First, the same probe sets filtered in the training data were removed from the test data. Probe set expression levels were then centered separately within each of the two test sets. Expression summaries at the cluster level were calculated for the test samples based on the clusters derived from the training set (probe sets were standardized relative to the standard deviation on the training data, so as to limit biases resulting from compositional differences between the training and testing data). Risk scores were then constructed for each test sample, using the Cox model coefficients determined from the training data. This classifier gives a continuous risk score.

Method B (Stratified Cox model on univariately selected genes)

Samples used in developing the classifier:

All UM and HLM samples in the training set, for both the prediction of all stages, and the prediction of stage 1 samples

Initial data processing and filtering:

All Affymetrix control probesets were removed from consideration. The gene expression values (g) for each probeset were transformed to a standardized scale by subtracting the median (μ) across the samples and dividing by the MAD (τ). This data transformation is described by the formula $(g-\mu)/\tau$. The transformation accounts for variability in the gene expression values from batch to batch across the different sets of data.

Gene Selection or feature reduction:

A stratified Cox model was developed with the training data using stratification based on AJCC stage. Four stages (Ia, Ib, II & III) and gene expression (g) were used in modeling survival (Sg) using the equation $Sg(t, stage, g) = Sg0(t, stage)*e^{(g*beta)}$, where t is patient survival time and Sg0 is baseline survival based on stage. Due to the weak association between gene expressions and survival the decision was made to select the top 50 genes (and ties) based on the model likelihood ratio statistic. Fifty-two probesets were selected for the final predictor.

Classifier:
The predictor, based on the selected genes, involved two computations. A risk score is first calculated for a sample using the log hazard ratio parameters β_i for the ith gene calculated from the training data. Letting g_{si} denote the expression level for gene i in sample s, the risk score is $r_s = \Sigma\beta_i(g_{si}-\mu_i)/\tau_i$, where μ_i and τ_i are the mean and standard deviation of the gene expression values for gene i across all training samples. This calculated risk score is then used to calculate the 5-year probability for any subject, taking into account information about tumor stage, with the model $Sr(t, stage, r) = Sr0(t, stage)*e^{(r*\gamma)}$. In computing the values, stage specific baseline rates (Sr0 = 0.716, 0.562, 0.314 and 0.152 respectively for stages Ia, Ib, II & III) and γ (0.118) were determined from the training set.

Class prediction or risk calculation for the test data:

For the samples in the test set the standardized gene expression values were calculated based on the test set data to account for systematic bias by site. All other values were used from calculations performed on the training data. The 5-year probabilities were calculated from the test set array data as described above. This classifier gives a continuous risk score.

Method C (clustering of samples combined with minimum gene selection).

Samples used in developing the classifier:

All UM samples in the training set

Initial data processing and filtering:

Set all values less than 1 equal to one and then log2 transform the data Exclusion of all control probesets (total of 68); Exclusion of 2 probesets: 207140_at and 207226_at due to a high proportion of samples having the value of 0 or missing values for these probesets; 1534 probesets differentially expressed between the 14 UM samples and the 4 HLM samples of Stratagene Universal RNA arrays identified using SAM analysis (100

permutations; FDR = 4.1%; Tusher et al. Proc Natl Acad Sci U S A. 2001 98(9):5116-5121).

Gene Selection or feature reduction:

A Maximizing Chi-square Analysis (MCA) approach was employed for identification of a minimum set of genes that had the highest independent ability of classifying patients into good/bad prognostic subgroups. The MCA approach includes two main steps: 1) identifying probesets whose expressions are significantly associated with prognosis, and 2) reducing the initial list to a minimum set of genes. The initial identification of significant probesets was done by univariate analysis using Cox proportional hazards regression. This resulted in the selection of 167 probesets using a p-value threshold of 0.001 for the selection. The second step involved stepwise exclusion of probesets, one probeset at a time. For the identified *n* probesets, MCA left one probeset out, used the *n*-1 probesets for k-means clustering (k=2, iteration=99), and calculated the Chi-square of cluster separation in terms of survival prediction using Cox proportional regression while adjusting for stage. The resulting *n* Chi-squares were compared and a probeset was excluded when its exclusion resulted in the largest Chi-square. When multiple probesets had the same largest Chi-square, the one with smallest univariate Chi-square was excluded. The procedure was repeated on the progressively smaller set of probesets until only one probeset was left. Finally, the minimum set of probesets necessary for maximum cluster separation was selected as the minimal set having the largest Chi-square. All statistical analyses, including MCA for probeset selection, were carried out using SAS v9.1 (SAS Institute Inc. NC, USA).

Classifier:

Binary Tree-Structured Vector Quantization (BTSVQ) program generates a binary cluster tree by using k=2 such that, at each level, the data are partitioned into two groups based on the degree of similarity of their expression profiles. When clustering samples, they are successively divided into two groups, until the resulting clusters at a given level of the cluster tree are homogeneous. Each partition is made to maximize the inter-cluster distance while continuing to minimize the intra-cluster distance. This hierarchical clustering method generates a binary tree that can be used to determine which sample types have the most similar gene expression profiles (4; Sultan et al). For classification of the samples in the training sets, the data from either of the training sets corresponding to the classifier probe sets identified by MCA were used to cluster the samples of each of the training sets separately in BTSVQ. For each of the training sets the first 2 sample clusters, which as per BTSVQ would contain the first separation of samples that differ in their classifier expression profile, were assigned “good or “bad” outcome according to the cumulative clinical outcome of the samples in the two clusters. For the classification of the test sets, we obtained a vector containing the median expression values for each of the probesets of the classifier corresponding to all the samples in either the “good” or “bad” clusters from the UM training set. These vectors of median expression values were used as outlined below.

Class prediction or risk calculation for the test data:

For the classification of the test sets samples, the data corresponding to the classifier probesets for each of the test sets was clustered in BTSVQ and clusters from the first level split in the cluster tree were obtained. Subsequently, Spearman's correlation was computed between individual test samples in each of the 2 clusters produced in BTSVQ and the 'good' and 'bad' median expression vectors determined from the UM training data, as mentioned above. The test set clusters were assigned 'good' or 'bad' based on the proportion of samples with higher correlation values to the median expression vector of either the 'good' or 'bad' clusters of UM training set. Finally, individual samples were assigned 'good' or 'bad' outcome according to assignment given to the entire cluster in which they were.

Method D. (clustering of samples combined with minimum gene selection).

Samples used in developing the classifier:

All UM Stage 1 samples in the training set.

Initial data processing and filtering:

Set all values less than 1 equal to one and then $\log_2$ transform the data Exclusion of all control probesets (total of 68); Exclusion of 2 probesets: 207140_at and 207226_at due to a high proportion of samples having the value of 0 or missing values for these probesets; 1534 probesets differentially expressed between the 14 UM samples and the 4 HLM samples of Stratagene Universal RNA arrays identified using SAM analysis (100 permutations; FDR = 4.1%; Tusher et al. Proc Natl Acad Sci U S A. 2001 98(9):5116-5121).

Gene Selection or feature reduction:
The same procedures were used as for method C. This procedure gave 42 probesets.

Classifier:
The same procedures were used as for method C.

Class prediction or risk calculation for test data:

The same procedures were used as for method C.

Method E. (Single gene classifier)

Samples used in developing the classifier:

A subset of extreme patients in the UM and HLM samples were used as the training set. Twenty-three samples from patients that survived less than 24 months and 25 samples from patients that survived longer than 60 months were utilized.

Initial data processing and filtering:

Exclude all probesets with median expression below 100.

Gene Selection or feature reduction:

Using only training data from HLM those probesets were identified for which a greater than 2-fold difference was observed between the medians of the short term survivor samples relative to the long term survivor samples. The same was done with the training data from UM. The intersection of these lists was the probeset 204351_at.

Classifier:

This is a single gene classifier with the classification of samples based on the relative expression of S100 calcium binding protein P as measured by the probeset 204351_at.

Class prediction or risk calculation for the test data:

Class is determined by post-hoc split of S100P expression determined from the distribution of the test data. After completing the study, we realized that this was inconsistent with our protocol (which only allowed for standardization of test set data or for calibration against percentile points of test set data). Since the method did not perform well, the impact of this on our overall findings is minimal.

Method F. (Principal component classifier on univariately selected genes)

Samples used in developing the classifier:

The HLM samples were used in the training set excluding 4 cases with death occurring less than 4 months after surgery. See Table 1 below for specific samples.

Initial data processing and filtering:

Remove all probesets where less than 20% of the samples had expression values 1.5-fold higher or lower than the median expression value.

Gene Selection or feature reduction:

Genes were selected by fitting Cox proportional hazards models to each gene in the training set and selecting only those with p <= 0.001. This yielded 42 probesets. The algorithm used was the Survival Analysis Prediction Tool from BRB ArrayTools (developed by Dr. Richard Simon and Amy Peng Lam - http://linus.nci.nih.gov/BRB-ArrayTools.html).

Classifier:

A Cox proportional hazards model was built using the first two principal components calculated from the 42 probe set list using BRB ArrayTools. The principal components are combinations of the individual probesets with coefficients determined from the training set. A Cox proportional hazards model was fit to the training set using the first two principal components.

Class prediction or risk calculation for the test data:

The first two principal component coefficients from the training set were applied to the gene expression levels from the test set. These scores were then further reduced using the coefficients from the Cox model linear predictor that was fit to the training set. Using the median risk score from the training data, the risks can be categorized as high or low.

Table 1: Samples used for method F.

524145	44.94	1	training
950017	34.76	1	training
323092	19.84	1	training
885163	8.61	1	training
781068	37.55	1	training
956943	60	0	training
247671	60	0	training
943565	56.8	1	training
933667	30.12	1	training
942053	13.33	1	training
936089	60	0	training
819465	44.42	1	training
366281	20.86	1	training
701581	27.63	1	training
759490	60	0	training
525810	60	0	training
265738	41.33	1	training
903414	5.65	1	training
792784	16.56	1	training

559504	60	0	training
894965	60	0	training
675746	60	0	training
148178	3.52	1	training
517608	17.54	1	training
845868	27.56	1	training
144587	16.39	1	training
836317	60	0	training
419309	60	0	training
659306	60	0	training
697490	45.24	1	training
316985	60	0	training
510932	39.06	1	training
193820	60	0	training
463273	2.27	1	training
932215	60	0	training
221190	42.18	1	training
897384	60	0	training
517541	60	0	training
249311	39.36	1	training
159395	60	0	training
376123	31.7	0	training
130882	42.02	1	training
684308	5.78	1	training
733168	7.69	1	training
696540	47.9	1	training
153791	58.84	0	training
816895	60	0	training
334341	24.51	1	training
189573	7.03	1	training
430943	9.63	1	training
391232	40.21	1	training
123673	60	0	training
678515	40.97	1	training
510311	60	0	training
631005	60	0	training
391371	11.99	1	training
112527	6.24	1	training
489693	21.88	1	training
824309	60	0	training
814773	60	0	training
216432	28.25	1	training
125786	53.39	0	training
568752	39.03	0	training
996308	52.2	1	training
448123	15.77	1	training
306267	27.66	1	training
238604	29.67	1	training
950812	12.02	1	training
993179	60	0	training

728260	15.47	1	training
281485	24.74	1	training
360466	16.49	1	training
925629	8.94	1	training
733074	26.38	1	training
701182	18.2	1	training
888397	60	0	training

Method G. (Principal component classifier on univariately selected genes)

Samples used in developing the classifier:

UM and HLM samples were used in the training set. All 100 samples with survival greater than 1 month and less than 60 months were used along with 42 samples selected at random from the patients that survived longer than 60 months. See Table 2 below for specific samples.

Table 2: Samples used for method G, including a mix of HLM and UM samples.

Experiment	Names	time	censor	traintest
463273	HLM	2.27	1	training
148178	HLM	3.52	1	training
903414	HLM	5.65	1	training
684308	HLM	5.78	1	training
112527	HLM	6.24	1	training
189573	HLM	7.03	1	training
733168	HLM	7.69	1	training
885163	HLM	8.61	1	training
925629	HLM	8.94	1	training
430943	HLM	9.63	1	training
391371	HLM	11.99	1	training
950812	HLM	12.02	1	training
942053	HLM	13.33	1	training
728260	HLM	15.47	1	training
448123	HLM	15.77	1	training
144587	HLM	16.39	1	training
360466	HLM	16.49	1	training
792784	HLM	16.56	1	training
517608	HLM	17.54	1	training
701182	HLM	18.2	1	training
323092	HLM	19.84	1	training
366281	HLM	20.86	1	training
489693	HLM	21.88	1	training
334341	HLM	24.51	1	training
281485	HLM	24.74	1	training
733074	HLM	26.38	1	training
845868	HLM	27.56	1	training
701581	HLM	27.63	1	training
306267	HLM	27.66	1	training

216432	HLM	28.25	1	training
238604	HLM	29.67	1	training
933667	HLM	30.12	1	training
950017	HLM	34.76	1	training
781068	HLM	37.55	1	training
510932	HLM	39.06	1	training
249311	HLM	39.36	1	training
391232	HLM	40.21	1	training
678515	HLM	40.97	1	training
265738	HLM	41.33	1	training
130882	HLM	42.02	1	training
221190	HLM	42.18	1	training
819465	HLM	44.42	1	training
524145	HLM	44.94	1	training
697490	HLM	45.24	1	training
696540	HLM	47.9	1	training
996308	HLM	52.2	1	training
943565	HLM	56.8	1	training
631005	HLM	60	0	training
894965	HLM	60	0	training
675746	HLM	60	0	training
800868	UM	2.24	1	training
663986	UM	2.4	1	training
252020	UM	2.7	1	training
659223	UM	3.3	1	training
567176	UM	4	1	training
432758	UM	5.8	1	training
271063	UM	5.9	1	training
404793	UM	6.38	1	training
321219	UM	6.6	1	training
455846	UM	7.2	1	training
333004	UM	7.3	1	training
772404	UM	8.4	1	training
474137	UM	8.7	1	training
209144	UM	8.85	1	training
569229	UM	9	1	training
599309	UM	9.2	1	training
826326	UM	9.6	1	training
913626	UM	9.74	1	training
150128	UM	10.1	1	training
518640	UM	11.68	1	training
265064	UM	11.7	1	training
850333	UM	12.2	1	training
233644	UM	12.5	1	training
618265	UM	12.63	1	training
145764	UM	14.5	1	training
173905	UM	14.9	1	training
496607	UM	16.3	1	training
683907	UM	16.8	1	training
398109	UM	16.9	1	training

952867	UM	17.6	1	training
466286	UM	17.6	1	training
403432	UM	18.6	1	training
702722	UM	19	1	training
880814	UM	19	1	training
411207	UM	19.5	1	training
661131	UM	19.6	1	training
798369	UM	19.9	1	training
602007	UM	20.6	1	training
712593	UM	20.9	1	training
452459	UM	21	1	training
949065	UM	21.2	1	training
859039	UM	21.48	1	training
341749	UM	22.3	1	training
202710	UM	22.7	1	training
714441	UM	24.97	1	training
769668	UM	25.95	1	training
555993	UM	26.9	1	training
720794	UM	27.6	1	training
139107	UM	28.2	1	training
493037	UM	29.6	1	training
161979	UM	30.2	1	training
402191	UM	30.7	1	training
977057	UM	31.28	1	training
524193	UM	31.94	1	training
110567	UM	32.24	1	training
284987	UM	32.6	1	training
443769	UM	33.2	1	training
744956	UM	33.9	1	training
185234	UM	34.8	1	training
812142	UM	36.25	1	training
476148	UM	40.2	1	training
496854	UM	42.04	1	training
445613	UM	42.5	1	training
940717	UM	43.8	1	training
172712	UM	45.3	1	training
183391	UM	45.33	1	training
424579	UM	45.8	1	training
255153	UM	47.8	1	training
980608	UM	48	1	training
951138	UM	48.8	1	training
370811	UM	50.2	1	training
222148	UM	54.2	1	training
352567	UM	54.93	1	training
651553	UM	59	1	training
999364	UM	59.11	1	training
913728	UM	60	0	training
450566	UM	60	0	training
579761	UM	60	0	training
763623	UM	60	0	training

668131	UM	60	0	training
914179	UM	60	0	training
826771	UM	60	0	training
943098	UM	60	0	training
687055	UM	60	0	training
165242	UM	60	0	training
878703	UM	60	0	training
425795	UM	60	0	training
509008	UM	60	0	training
542450	UM	60	0	training
418227	UM	60	0	training
747041	UM	60	0	training
440283	UM	60	0	training

Initial data processing and filtering:

Same procedure as for method F.

Gene Selection or feature reduction:

Same procedure as for method F, which yielded 38 probesets in this case.

Classifier:

Same procedure as for method F.

Class prediction or risk calculation for the test data:

Same procedure as for method F.

Method H. (Majority vote classifier on mitosis genes).

Samples used in developing the classifier:

All HLM samples in the training set, for both all stage prediction and for stage 1 prediction.

Initial data processing and filtering:

An initial list of 614 probesets suspected of having predictive capacity for survival because they were related to mitosis was initially developed with independent data. This set of probesets served as the initial gene set G_1 described below.

Gene Selection or feature reduction:

The probesets and thresholds for classification were selected by an iterative application of the classification process described below. The process was initiated with 614 pre-selected probesets (G_1). This set of probesets was used to score the HLM samples, as described below. Based on these scores, the highest scoring third of the samples and the lowest scoring third of the samples were selected. A t-test was performed to identify genes that best discriminated between these two subsets of the HLM training data ($p_i < 0.05/n$ for Bonferroni correction). This list of identified genes formed the new gene set G_2, which was used to start the process over again. After 3 iterations a static group of 313 probesets was identified that were used for further classification of all other data sets, following the "majority vote" procedure below.

Scoring:

Terminology
C = majority vote of individual classifiers (sum of vote by all probesets in classifier)
c_k = individual classifier k (a single probeset and thresholds for that probeset)
x_j = sample j (microarray dataset for an array from a single tumor sample)
G = set of genes used for individual classifiers (set of all probesets in classifier)
g_{jk} = gene expression value for sample j and probeset k
S = sign (+/-) indicating trend relative to outcome, + = high expression in Poor class, - = high expression in Good class

Individual Classifier

For each element g_k of G define LO and HI as the 33rd percentile and 66th percentile of expression values in the dataset under investigation. Then define

$$c_k = \begin{cases} -1 * S_k & g_{jk} < LO \\ +1 * S_k & g_{jk} > HI \\ 0 & \text{otherwise} \end{cases}$$

Majority Vote Classifier

The final risk score is the sum of all the individual classifiers, $C_j = \Sigma (c_{kj})$.
For training purposes thresholds were set to define the extreme classes used for gene selection.

$$\text{Pred}(x_j) = \begin{cases} \text{Poor} & C_j > 0.15\,|G| \\ \text{Good} & C_j < -0.15\,|G| \\ \text{Grey} & \text{otherwise} \end{cases}$$

0.15 was heuristically determined. A Pred value of |G| (or –|G|) indicates complete agreement with one class whereas values near zero (e.g. 0.15 * |G|) indicate uncertainty in classification.

Class prediction or risk calculation for the test data:

The thresholds for individual classifiers were reset based on the distribution of the gene expression values in each test set. The majority vote classifier was calculated to generate a score. This score exists within a continuum of values between 313 and –313, which can serve as a risk score. For the 2 class discrimination in this study, the threshold was set at C_j >or< 0.15 |G|.

Method I. (Clinical variables only)

Samples used in developing the classifier:

All UM and HLM samples in the training set.

Initial data processing and filtering:

None.

Gene Selection or feature reduction:

N/A

Classifier:

AJCC stage (coded as 1, 2, and 3/4), gender, and age were included in a multivariate Cox proportional hazards model.

Class prediction or risk calculation for the test data:

The fitted Cox model coefficients calculated from the training data are used to construct a continuous risk score.

Method J and K. (Based on Chen et al. NEJM 2007 set of genes)

Samples used in developing the classifier:

All UM and HLM samples in the training set.

Initial data processing and filtering:

All probe sets for which fewer than five samples showed greater than 50 raw expression units were removed. For the remaining probesets all values less than 0 were set to zero, all values were increased by 1, and then all the values were $\log_2$ transformed. An additional 274 probe sets were removed that had a greater than 2 fold difference in expression between the UM samples and the HLM samples. Expression levels within each of the four sample sets (UM, HLM, MSK, and CAN/DF) were mean centered at the probe-set level.

Gene Selection or feature reduction:

All probe sets on the Affymetrix U133A array that detected the same gene names reported in Chen et al., were identified. For method J the five gene signature was translated into 9 probe sets. For method K the 16 gene signature was translated into 33 probe sets covering 15 of the 16 genes used in Chen et al. No U133A probe set detected transcripts for the CPEB4 gene.

The probe sets for the 5 genes are:

DUSP6	208891_at, 208892_s_at, 208893_s_at
ERBB3	202454_s_at
LCK	204890_s_at, 204891_s_at
MMD	203414_at
STAT1	200887_s_at, 209969_s_at

The probe sets for the 16 genes are:

HMMR	207165_at, 209709_s_at
ERBB3	202454_s_at

DLG2	206253_at
ANXA5	200782_at
STAT2	205170_at, 217199_s_at
STAT1	200887_s_at, 209969_s_at
IRF4	216987_at, 216986_s_at, 204562_at
FRAP1	202288_at, 215381_at
ZNF264	205917_at
NF1	211914_x_at, 210631_at, 204323_x_at, 204325_s_at, 211095_at, 211094_s_at
HGF	209961_s_at, 210755_at, 209960_at,210998_s_at, 210997_at
MMD	203414_at
RNF4	212696_s_at
DUSP6	208891_at, 208892_s_at, 208893_s_at
LCK	204891_s_at, 204890_s_at

Classifier:

The gene expression values for the selected probe sets were used for fitting ridged Cox proportional hazards models to our training set data, as describe in method A.

Class prediction or risk calculation for the test data:

The fitted Cox model coefficients calculated from the training data and the gene expression values from each sample in the test data are used to construct a continuous risk score.

Method L, M, and N. (Modeled after Potti et al, NEJM 2006)

Samples used in developing the classifier:

All UM and HLM samples in the training set.

Initial data processing and filtering:

All probe sets for which fewer than five samples showed greater than 50 raw expression units were removed. For the remaining probe sets all values less than 0 were set to zero, all values were increased by 1, and then all the values were $\log_2$ transformed. Standard deviations were calculated for each probe set and one quarter of the probe sets with the smallest standard deviations were removed. An additional 274 probe sets were removed that had a greater than 2 fold difference in expression between the UM samples and the HLM samples. Expression levels within each of the four sample sets (UM, HLM, MSK, and CAN/DF) were mean centered at the probe-set level.

Gene Selection or feature reduction:

In the Supplementary Material of Potti et al., U133 probe sets belonging to 9 metagenes are listed (note that this is a subset of all metagenes used by their method). We reduced the original set of 132 probesets to 80 probe sets following the filtering procedure described above. We then considered three distinct approaches for constructing risk scores based on these 80 probesets. For method L, the first principal component was used to summarize the expression data in each metagene. The 9 metagene summaries were used as covariates in a ridged Cox proportional hazards model, as described for method A above. The model was developed with a ridging parameter of 1, but other values were found to give similar or inferior results on the testing data. Method M fit a 45-dimensional ridged Cox model in which the metagene principal component scores were main effects, and in addition all two-way interactions were included. Method N fit an 80-dimensional ridged Cox model using the gene expression levels for all genes in the 9 metagenes, ignoring their assignment to metagenes. We also built models using gene expression and clinical covariates, in which case the three clinical covariates were incorporated as additional additive terms in the Cox model linear predictor. This gave 12, 48, and 83 dimensional models, respectively.

Classifier:

Following data reduction ridged Cox proportional hazards models were constructed as described above.

Class prediction or risk calculation for the test data:

Linear predictors from the ridged Cox proportional hazards models developed in methods L-N were used as continuous risk scores.

Metagene	Probesets
35	202539_s_at, 203766_s_at, 206912_at, 211165_x_at, 213575_at
69	203969_at,203971_at,206691_s_at,206924_at,208902_s_at,209024_s_at,209252_at,212016_s_at

40	200881_s_at,201116_s_at,201120_s_at,201138_s_at,201182_s_at,201455_s_at,202666_s_at,203018_s_at,206302_s_at,211442_x_at,211778_s_at,212360_at,213244_at,213693_s_at,214421_x_at
41	204168_at,205844_at,206066_s_at,206370_at,209971_x_at,210431_at,210609_s_at,211062_s_at,213188_s_at,213517_at
74	203850_s_at,205887_x_at,210785_s_at,213650_at
79	200983_x_at,201481_s_at,201627_s_at,202812_at,203796_s_at,203856_at,206022_at,209479_at,210220_at,211207_s_at,213457_at
19	200908_s_at,203914_x_at,204928_s_at,205767_at,207108_s_at,207814_at
86	203741_s_at,204122_at,204381_at,205339_at,207355_at,207606_s_at,209590_at,209751_s_at,210794_s_at,212169_at,213143_at,213433_at
31	201037_at,201637_s_at,201850_at,204633_s_at,204727_at,209343_at,210381_s_at,212034_s_at,212410_at

Challenges in developing prognostic models for early stage lung cancer

Some potentially significant additional challenges will arise if gene expression profiling is to be used for lung cancer prognosis in practice. Most likely, a method would be trained on a large reference set, and the fitted model used to make predictions at distinct treatment sites. To account for this, we constructed our testing sets around institutional samples, rather than randomly assigning the entire set of 400+ samples to training and testing subsets. The presence of lower signal in one test set and moderate stage and gender biases in both test sets relative to the training set are therefore realistic elements of our study. However the fact that we produced gene expression summaries by running DChip on the entire sample (training and testing sets together) is somewhat unrealistic, as the adaptive aspect of DChip's approach to expression quantitation may have removed some of the inter-site differences. A related point is that it was permitted for a method to standardize gene expression data within test sets, or to refer to percentile points of summary features within test sets. We view this as being realistic, since in practice some form of standardization and calibration is sure to be used.

There are a number of challenges that should be considered in further efforts to develop accurate prognostic models for lung cancer. The well recognized biologic heterogeneity

of lung adenocarcinomas may contribute to difficulties in prediction. These tumors show considerable histological variation and in most instances contain mixed subtypes with several distinct regions present in a tumor. Thus the region selected for gene expression analysis may not always correspond to the most informative component of the lesion in regards to accurately describing tumor behavior. It is also likely that some of this heterogeneity in lung adenocarcinomas results from genetic alterations related to smoking tobacco (Muira et al, 2002; Powell et al, 2003), which was highly prevalent in the subject cohorts examined. However specific information about smoking in our study was too limited to incorporate into the analysis.

There are also difficulties in using overall survival as an endpoint in prognostic modeling in cancer. Although genes associated with aggressive disease, such as those involved in cell proliferation or increased glycolysis (Beer et al, 2002, Chen et al 2003), may be clearly present in a subject's tumor, this may not always correlate with reduced survival if, for example, a complete resection of this lesion was successful. Similarly, a subject's tumor might exhibit a favorable gene expression profile, but the subject might develop and succumb to a second primary or to some other clinical condition shortly after diagnosis. In both instances, there would be a lack of correlation between gene expression and subject survival. Time to recurrence may be more directly related to tumor biology, but is difficult to accurately define in a retrospective setting. In the UM/HLM, MSK and CAN/DF sets, respectively 99, 32, and 31 subjects died during the study with an explicit indication of disease recurrence. Only 27, 7 and 2 subjects, respectively, died during the study with no indication of lung cancer recurrence in their clinical files. While we did not view our recurrence data as being sufficiently accurate to use as an analysis endpoint, this suggests that a large fraction of the deaths in our study were due to lung cancer. For future profiling projects, methods must be developed to more accurately track this information.

References for Supplementary Materials 2:

Tusher et al. Proc Natl Acad Sci U S A. 2001 98(9):5116-5121).

Chen HY, Yu SL, Chen CH, Chang GC, Chen CY, Yuan A, Cheng CL, Wang CH, Terng HJ, Kao SF, Chan WK, Li HN, Liu CC, Singh S, Chen WJ, Chen JJ, Yang PC A five-gene signature and clinical outcome in non-small-cell lung cancer. *N Engl J Med*. 2007 Jan 4;356(1):11-20.

Potti A, Mukherjee S, Petersen R, Dressman HK, Bild A, Koontz J, Kratzke R, Watson MA, Kelley M, Ginsburg GS, West M, Harpole DH, Nevins JR.A genomic strategy to refine prognosis in early-stage non–small-cell lung cancer. *N Engl J Med* 2006;355:570–80.

Li C, Wong WH. Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. Proc Natl Acad Sci U S A 2001;98(1):31-6.

Sultan M, Wigle DA, Cumbaa CA, Maziarz M, Glasgow J, Tsao MS, Jurisica I. Binary tree-structured vector quantization approach to clustering and visualizing microarray data. Bioinformatics 2002;18 Suppl 1:S111-9. PMID: 12169538.

Miura K, Bowman ED, Simon R, Peng AC, Robles AI, Jones RT, Katagiri T, He P, Mizukami H, Charboneau L, Kikuchi T, Liotta LA, Nakamura Y, Harris CC. Laser capture microdissection and microarray expression analysis of lung adenocarcinoma reveals tobacco smoking- and prognosis-related molecular profiles. *Cancer Res*. 2002 Jun 1;62(11):3244-50.

Powell CA, Spira A, Derti A, DeLisi C, Liu G, Borczuk A, Busch S, Sahasrabudhe S, Chen Y, Sugarbaker D, Bueno R, Richards WG, Brody JS. Gene expression in lung adenocarcinomas of smokers and nonsmokers. Am J Respir Cell Mol Biol. 2003 Aug;29(2):157-62.

Beer DG, Kardia SL, Huang CC, Giordano TJ, Levin AM, Misek DE, Lin L, Chen G, Gharib TG, Thomas DG, Lizyness ML, Kuick R, Hayasaka S, Taylor JM, Iannettoni MD, Orringer MB, Hanash S Gene-expression profiles predict survival of subjects with lung adenocarcinoma. 2002; *Nat Med.,* 8(8):816-24.

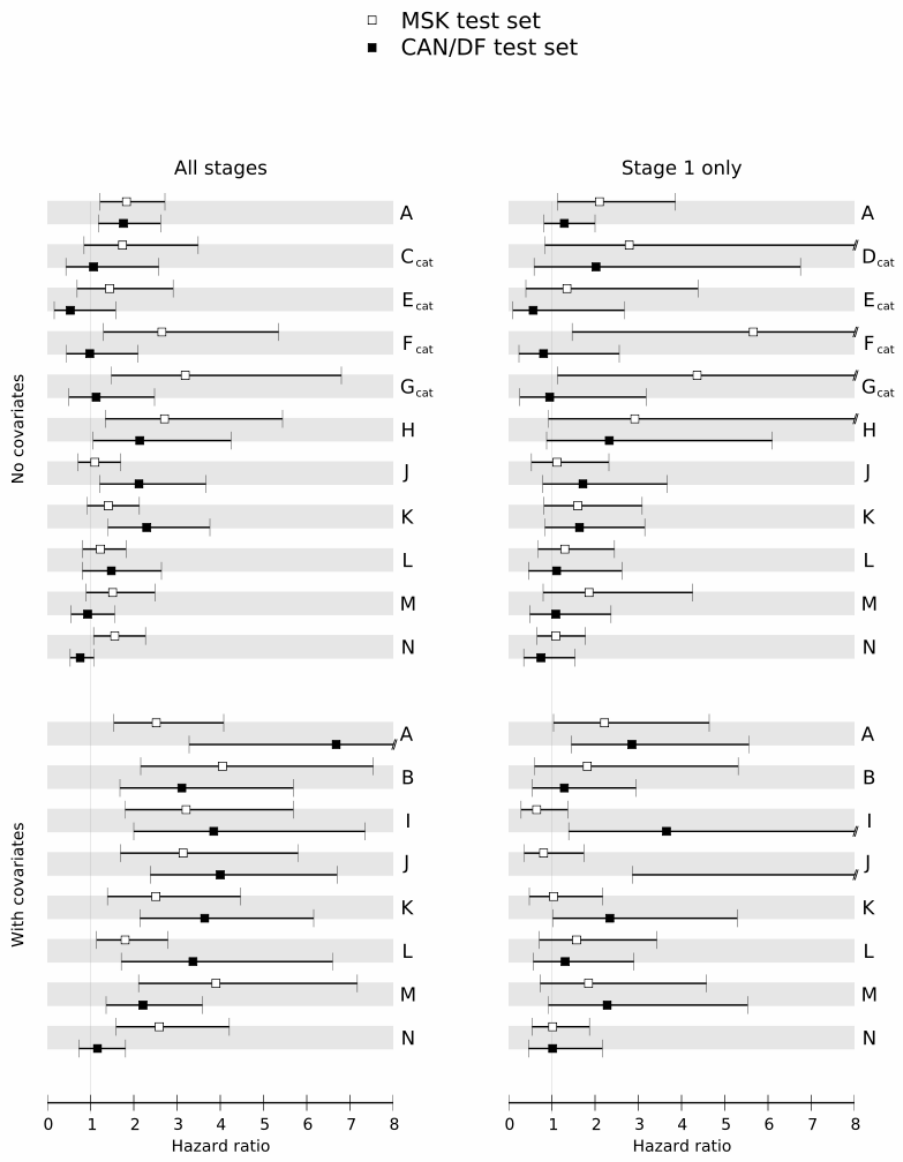
Chen G, Gharib TG, Wang H, Huang CC, Kuick R, Thomas DG, Shedden KA, Misek DE, Taylor JM, Giordano TJ, Kardia SL, Iannettoni MD, Yee J, Hogg PJ, Orringer MB, Hanash SM, Beer DG. Protein profiles associated with survival in lung adenocarcinoma. *Proc Natl Acad Sci* U S A. 2003; Nov 11;100(23):13537-42.

Supplementary Materials 3a: detailed hazard ratio results

Figure. Results on two independent testing data sets. Figure showing hazard ratios for all the methods A-N.

Codes:

A-H: Methods developed by the consortia using gene expression with and without clinical data.
I: Clinical covariates only (stage, gender and age)
J: derived from Chen et al 5 gene signature
K: derived from Chen et al 16 gene signature
L: derived from Potti et al metagenes (Approach 1, main effects of metagenes)
M: derived from Potti et al metagenes (Approach 2, main effects and interactions of metagenes)
N: derived from Potti et al metagenes (Approach 3, all genes in 9 metagenes)


MSK test set
CAN/DF test set
All stages
Stage 1 only
No covariates
With covariates
A
C cat
D cat
E cat
F cat
G cat
H
I
J
K
L
M
N
B
0 1 2 3 4 5 6 7 8
Hazard ratio

Expanded Table of performance evaluations.

Hazard ratios for hypothesis 1. Gene expression only, all stages, MSK test set.

Classifier	Hazard ratio	95% CI	p-value	CPE
A	1.83	(1.24,2.70)	0.002	0.627
C	1.74	(0.87,3.47)	0.111	0.561
E	1.44	(0.71,2.89)	0.301	0.542
F	2.65	(1.32,5.33)	0.005	0.611
G	3.19	(1.50,6.78)	0.002	0.632
H	2.71	(1.36,5.42)	0.004	0.634
J	1.10	(0.73,1.68)	0.640	0.520
K	1.41	(0.94,2.11)	0.090	0.580
L	1.23	(0.84,1.80)	0.280	0.550
M	1.51	(0.92,2.47)	0.090	0.590
N	1.56	(1.09,2.25)	0.010	0.610

Hazard ratios for hypothesis 1. Gene expression only, all stages, CAN/DF test set.

Classifier	Hazard ratio	95% CI	p-value	CPE
A	1.76	(1.20,2.60)	0.003	0.623
C	1.07	(0.45,2.56)	0.878	0.506
E	0.53	(0.18,1.56)	0.239	0.553
F	0.98	(0.46,2.08)	0.947	0.503
G	1.13	(0.52,2.46)	0.751	0.515
H	2.14	(1.08,4.23)	0.025	0.621
J	2.12	(1.23,3.65)	0.010	0.640
K	2.30	(1.42,3.74)	0.000	0.660
L	1.48	(0.83,2.62)	0.180	0.570
M	0.93	(0.57,1.54)	0.780	0.520
N	0.76	(0.55,1.06)	0.100	0.580

Hazard ratios for hypothesis 2. Gene expression and clinical variables, all stages, MSK test set.

Classifier	Hazard ratio	95% CI	p-value	CPE
A	2.52	(1.56,4.06)	0.000	0.671
B	4.05	(2.18,7.52)	0.000	0.706
I	3.21	(1.82,5.67)	0.000	0.670
J	3.14	(1.71,5.78)	0.000	0.670
K	2.51	(1.41,4.45)	0.000	0.660
L	1.80	(1.16,2.77)	0.010	0.610

M	3.90	(2.13,7.15)	0.000	0.710
N	2.59	(1.61,4.19)	0.000	0.680

Hazard ratios for hypothesis 2. Gene expression and clinical variables, all stages, CAN/DF test set.

Classifier	**Hazard ratio**	**95% CI**	**p-value**	**CPE**
A	6.68	(3.30,13.52)	0.000	0.762
B	3.11	(1.70,5.67)	0.000	0.690
I	3.85	(2.02,7.533)	0.000	0.700
J	4.00	(2.40,6.68)	0.000	0.740
K	3.64	(2.16,6.14)	0.000	0.720
L	3.37	(1.73,6.58)	0.000	0.710
M	2.21	(1.38,3.57)	0.000	0.660
N	1.16	(0.75,1.79)	0.500	0.540

Hazard ratios for hypothesis 3. Gene expression only, stage 1 only, MSK test set.

Classifier	**Hazard ratio**	**95% CI**	**p-value**	**CPE**
A	2.10	(1.15,3.84)	0.014	0.656
D	2.79	(0.86,9.00)	0.080	0.579
E	1.35	(0.42,4.36)	0.610	0.535
F	5.66	(1.49,21.48)	0.009	0.668
G	4.36	(1.15,16.57)	0.027	0.656
H	2.92	(0.94,9.10)	0.060	0.640
J	1.12	(0.55,2.30)	0.740	0.520
K	1.60	(0.84,3.07)	0.150	0.610
L	1.30	(0.70,2.43)	0.390	0.570
M	1.86	(0.82,4.24)	0.130	0.630
N	1.09	(0.68,1.76)	0.700	0.530

Hazard ratios for hypothesis 3. Gene expression only, stage 1 only, CAN/DF test set.

Classifier	**Hazard ratio**	**95% CI**	**p-value**	**CPE**
A	1.29	(0.84,1.98)	0.243	0.574
D	2.02	(0.61,6.73)	0.242	0.558
E	0.57	(0.12,2.66)	0.465	0.547
F	0.81	(0.26,2.54)	0.715	0.526
G	0.95	(0.28,3.16)	0.928	0.506
H	2.33	(0.90,6.08)	0.077	0.638
J	1.72	(0.81,3.65)	0.150	0.610
K	1.64	(0.86,3.13)	0.130	0.610
L	1.11	(0.48,2.60)	0.800	0.520

M	1.09	(0.51,2.35)	0.820	0.520
N	0.75	(0.37,1.52)	0.410	0.560

Hazard ratios for hypothesis 4. Gene expression and clinical variables, stage 1 only, MSK test set.

Classifier	**Hazard ratio**	**95% CI**	**p-value**	**CPE**
A	2.22	(1.06,4.62)	0.030	0.659
B	1.82	(0.62,5.29)	0.264	0.596
I	0.65	(0.31,1.35)	0.230	0.590
J	0.81	(0.38,1.72)	0.580	0.550
K	1.04	(0.50,2.16)	0.920	0.510
L	1.57	(0.73,3.41)	0.240	0.590
M	1.85	(0.75,4.56)	0.170	0.620
N	1.02	(0.56,1.85)	0.950	0.510

Hazard ratios for hypothesis 4. Gene expression and clinical variables, stage 1 only, CAN/DF test set.

Classifier	**Hazard ratio**	**95% CI**	**p-value**	**CPE**
A	2.85	(1.47,5.54)	0.002	0.699
B	1.29	(0.57,2.93)	0.537	0.551
I	3.65	(1.41,9.46)	0.010	0.720
J	9.18	(2.88,29.32)	0.000	0.800
K	2.34	(1.04,5.27)	0.040	0.660
L	1.31	(0.59,2.88)	0.500	0.550
M	2.28	(0.94,5.51)	0.060	0.660
N	1.02	(0.48,2.15)	0.960	0.500

Supplementary Materials 3b: survivor function methodology

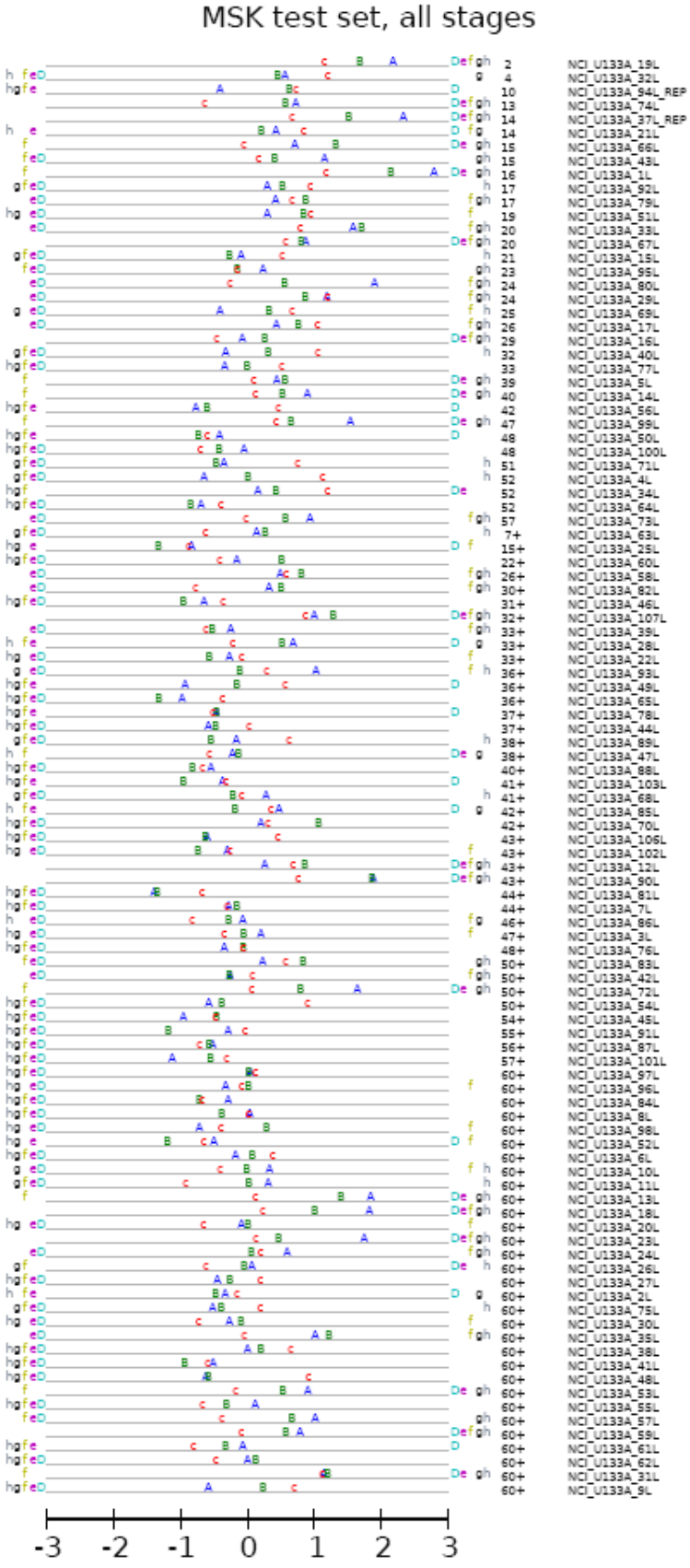
Method A survivor function plots for the training set (Figure 3 of the main article) were generated on the HLM+UM training data using a cross validation procedure. In each of 100 cross-validation replications, a random sample of 200 cases from the combined HLM+UM training set was drawn and used to completely retrain method A. The fitted method was then applied to the held-out samples to produce risk scores. Each risk score was paired with the actual event time and censoring data. The total set of all such score/outcome pairs was pooled over all cross validation replications, then divide into three groups of equal size based on the scores. The survivor function for each of the three subgroups was then estimated using the Kaplan-Meier method. Survivor functions for the test sets (Figure 2 in the main paper) were estimated directly using the Kaplan-Meier method based on the fitted classifier from the training sets. Additional survivor function plots for the other methods not shown in the main article are in Supplementary Figure S1.

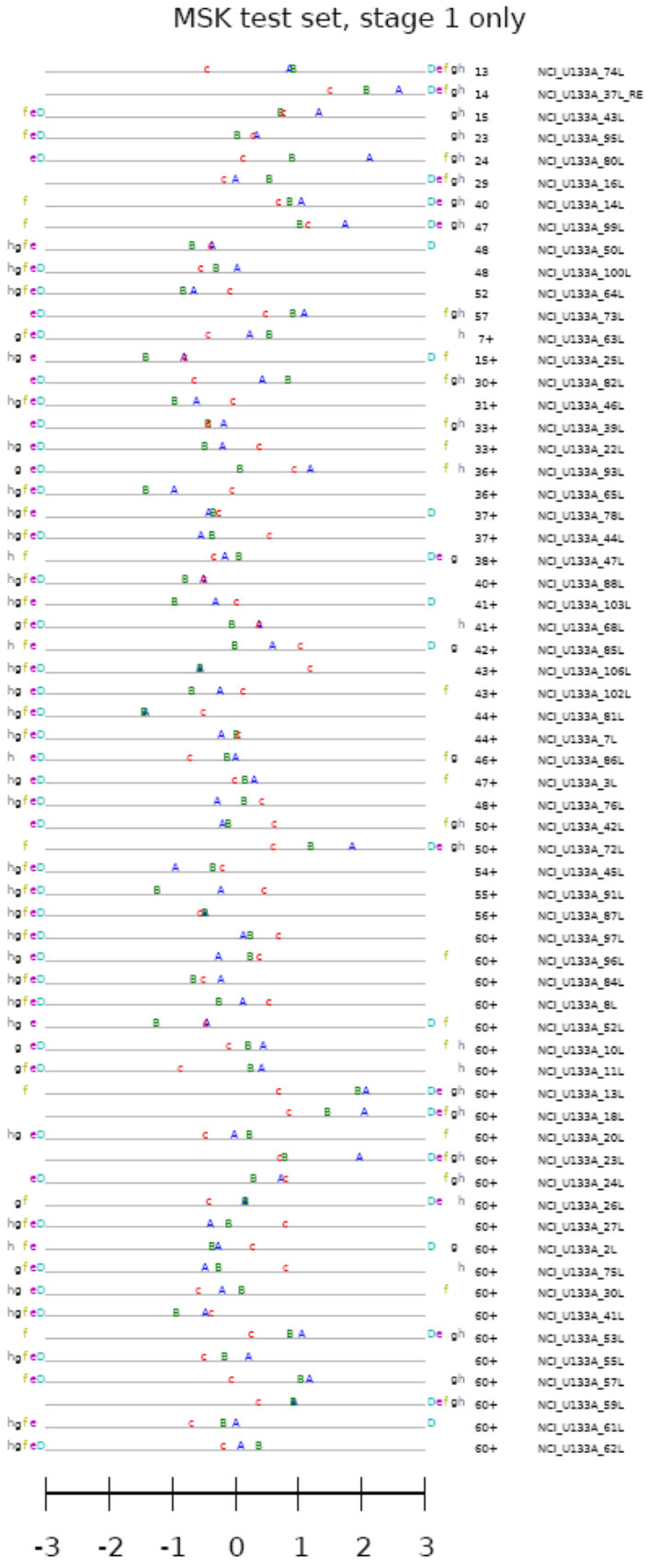
Supplementary Materials 4: classifier concordance

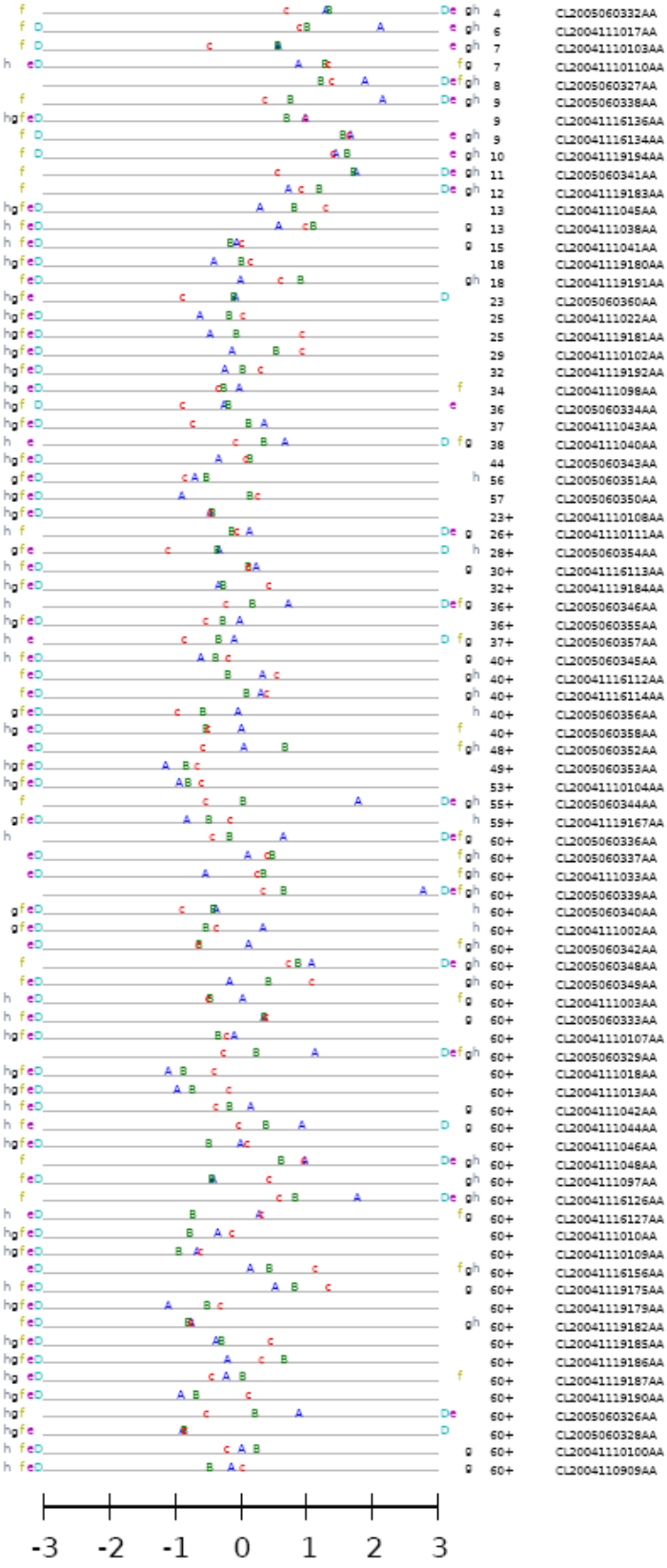
Concordance between classifiers.

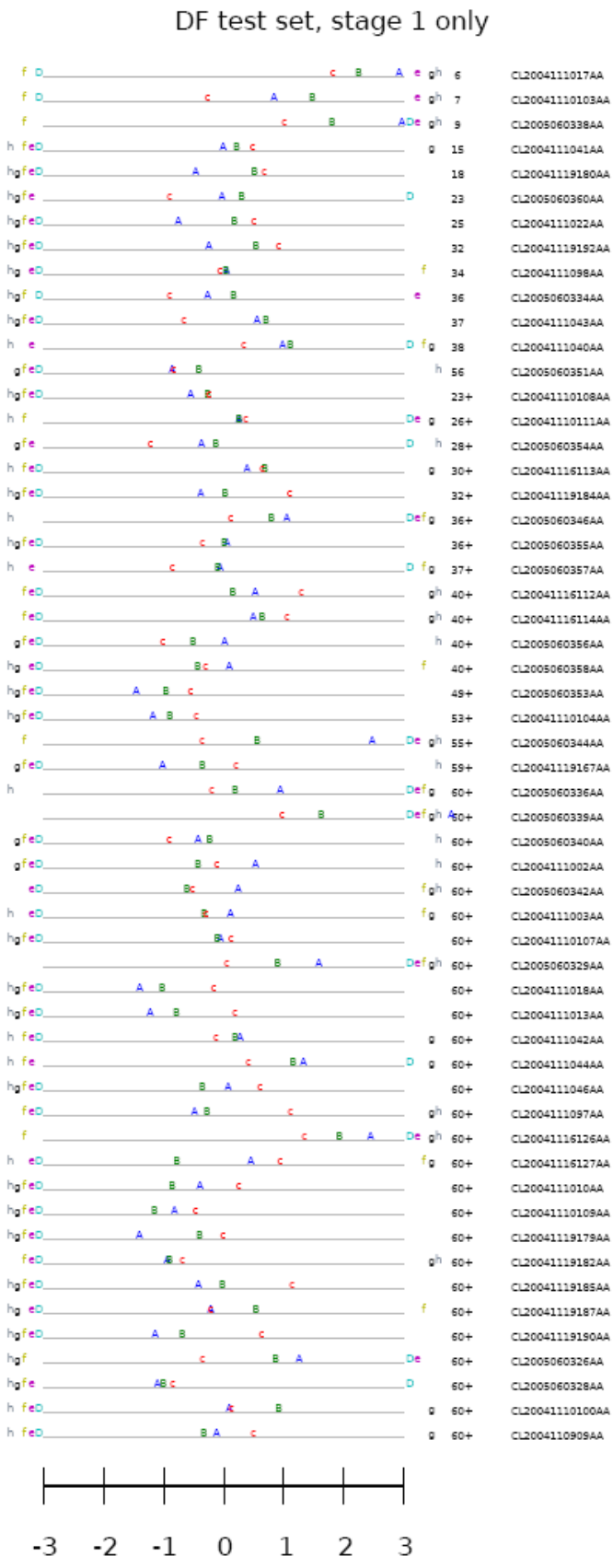
Each of the classifiers assigns a risk to every patient in the test set. For some classifiers (A,B and C) this is a continuous score, for other classifiers (D,E,F,G and H) it is a binary indication of low or high risk. For analysis of the concordance see Supplementary Table T1. Additionally, the concordance between the predictions from the classifiers is represented graphically in the following figures. For the continuous classifiers the risk scores are standardized to have mean zero and variance 1, and this is shown on the horizontal scale. For the binary classifiers the high risk samples are depicted at the right hand end of the horizontal scale and the low risk at the left hand end. Classifiers that were overall successful for that particular hypothesis are shown in upper case and those that were not successful are shown in lower case. For each of the test sets the patients are ordered based on time to death for those who died and time to last follow up for those who are censored.

The figures reveal an overall heterogeneity between the classifiers. Thus it is not the case that all the classifiers give similar predictions on a majority of the patients and differ on a small subset. Never-the-less there is a subset of patients who appear to be correctly predicted by nearly all the classifiers. For example in the MSK all stages prediction, patients 19L, 37L_REP and 67L are correctly predicted to die early by all the classifiers. Similarly 41L is predicted to be low risk by all the classifiers and does live for more than 5 years. However, there are also patients in which every classifier gives the same prediction, but the predictions are incorrect.


MSK test set, all stages
-3
-2
-1
0
1
2
3


MSK test set, stage 1 only
13 NCI_U133A_74L
14 NCI_U133A_37L_RE
15 NCI_U133A_43L
23 NCI_U133A_95L
24 NCI_U133A_80L
29 NCI_U133A_16L
40 NCI_U133A_14L
47 NCI_U133A_99L
48 NCI_U133A_50L
48 NCI_U133A_100L
52 NCI_U133A_64L
57 NCI_U133A_73L
7+ NCI_U133A_63L
15+ NCI_U133A_25L
30+ NCI_U133A_82L
31+ NCI_U133A_46L
33+ NCI_U133A_39L
33+ NCI_U133A_22L
36+ NCI_U133A_93L
36+ NCI_U133A_65L
37+ NCI_U133A_78L
37+ NCI_U133A_44L
38+ NCI_U133A_47L
40+ NCI_U133A_88L
41+ NCI_U133A_103L
41+ NCI_U133A_68L
42+ NCI_U133A_85L
43+ NCI_U133A_106L
43+ NCI_U133A_102L
44+ NCI_U133A_81L
44+ NCI_U133A_7L
46+ NCI_U133A_86L
47+ NCI_U133A_3L
48+ NCI_U133A_76L
50+ NCI_U133A_42L
50+ NCI_U133A_72L
54+ NCI_U133A_45L
55+ NCI_U133A_91L
56+ NCI_U133A_87L
60+ NCI_U133A_97L
60+ NCI_U133A_96L
60+ NCI_U133A_84L
60+ NCI_U133A_8L
60+ NCI_U133A_52L
60+ NCI_U133A_10L
60+ NCI_U133A_11L
60+ NCI_U133A_13L
60+ NCI_U133A_18L
60+ NCI_U133A_20L
60+ NCI_U133A_23L
60+ NCI_U133A_24L
60+ NCI_U133A_26L
60+ NCI_U133A_27L
60+ NCI_U133A_2L
60+ NCI_U133A_75L
60+ NCI_U133A_30L
60+ NCI_U133A_41L
60+ NCI_U133A_53L
60+ NCI_U133A_55L
60+ NCI_U133A_57L
60+ NCI_U133A_59L
60+ NCI_U133A_61L
60+ NCI_U133A_62L
-3
-2
-1
0
1
2
3


DF test set, all stages
-3
-2
-1
0
1
2
3


DF test set, stage 1 only
6 CL2004111017AA
7 CL2004111010 3AA
9 CL2005060338AA
15 CL2004111041AA
18 CL2004111918 0AA
23 CL2005060360AA
25 CL2004111022AA
32 CL2004111919 2AA
34 CL2004111098AA
36 CL2005060334AA
37 CL2004111043AA
38 CL2004111040AA
56 CL2005060351AA
23+ CL2004111010 8AA
26+ CL2004111011 1AA
28+ CL2005060354AA
30+ CL2004111611 3AA
32+ CL2004111918 4AA
36+ CL2005060346AA
36+ CL2005060355AA
37+ CL2005060357AA
40+ CL2004111611 2AA
40+ CL2004111611 4AA
40+ CL2005060356AA
40+ CL2005060358AA
49+ CL2005060353AA
53+ CL2004111010 4AA
55+ CL2005060344AA
59+ CL2004111916 7AA
60+ CL2005060336AA
60+ CL2005060339AA
60+ CL2005060340AA
60+ CL2004111002AA
60+ CL2005060342AA
60+ CL2004111003AA
60+ CL2004111010 7AA
60+ CL2005060329AA
60+ CL2004111018AA
60+ CL2004111013AA
60+ CL2004111042AA
60+ CL2004111044AA
60+ CL2004111046AA
60+ CL2004111097AA
60+ CL2004111612 6AA
60+ CL2004111612 7AA
60+ CL2004111010AA
60+ CL2004111010 9AA
60+ CL2004111917 9AA
60+ CL2004111918 2AA
60+ CL2004111918 5AA
60+ CL2004111918 7AA
60+ CL2004111919 0AA
60+ CL2005060326AA
60+ CL2005060328AA
60+ CL2004111010 0AA
60+ CL2004110909AA
-3
-2
-1
0
1
2
3

Supplementary Materials 5: lists of genes and gene clusters

The genes constituting the 100 clusters used for Method A are given at URL (http://www.stat.lsa.umich.edu/~kshedden/Beer-Test/cluster_profile). Additional probeset information for Methods B, C, D and H is provided below:

Method B:

Probe	row	median(exp)	mad(exp)	beta-hat
200066_at	136	595.98	146.61	-.26
200621_at	691	1066.44	360.4	-.31
205308_at	5376	229.25	94.6	-.27
205396_at	5464	84.34	19.83	.2
205409_at	5477	89.4	41.02	.23
205592_at	5660	23.17	8.59	.23
206791_s_at	6858	67.37	12.18	.29
206799_at	6866	13.96	4.85	.17
206986_at	7053	22.08	7.08	-.28
207228_at	7293	6.45	3.87	-.27
207313_x_at	7377	100.02	23.35	.17
208291_s_at	8335	32.09	7.6	.11
208559_at	8597	.54	.57	-.01
208644_at	8681	990.54	238.71	-.27
209237_s_at	9271	73.46	23.83	.19
209281_s_at	9315	121.85	49.73	.18
209418_s_at	9451	224.37	62.99	-.22
209890_at	9916	44.79	27.88	.19
210220_at	10241	84.09	61.24	-.21
211099_s_at	11083	15.05	7.11	-.28
211123_at	11107	94.12	16.66	.24
211327_x_at	11286	121.03	26.26	.26
211610_at	11555	119.29	46.04	.18
211829_s_at	11765	24.98	10.07	-.28
213628_at	13548	200.44	50.77	-.26
213855_s_at	13775	52.19	11.86	.28

214375_at	14294	115.54	33.46	.2
214443_at	14362	84.01	18.32	.18
214538_x_at	14457	94.05	33.23	-.29
214717_at	14633	12.73	9.86	-.25
215100_at	15014	151.37	29.99	.18
215809_at	15722	59.49	15.54	.25
216018_at	15930	37.11	13.35	.3
216227_at	16139	33.73	7.98	.28
216251_s_at	16163	760.29	187.04	.24
216398_at	16309	113.08	18.19	.28
217245_at	17150	45.91	11.49	.25
217538_at	17443	69.44	24.04	.22
218342_s_at	18247	287.76	100.94	.19
218498_s_at	18403	393.06	193.27	.16
218881_s_at	18785	85.56	34.52	.27
219043_s_at	18947	398.9	128.3	-.22
219069_at	18973	265.05	69.08	-.24
219184_x_at	19088	460.21	108.01	.23
219394_at	19298	494.34	168.57	-.29
219930_at	19834	19.2	19.43	-.25
220020_at	19924	46.79	14.64	.18
220495_s_at	20399	655.61	224.63	-.27
220550_at	20454	26.49	12.58	.27
220620_at	20524	31.94	6.78	.22
220987_s_at	20891	225.52	91.3	-.21
221852_at	21752	8.5	6.18	-.18

Method C

Probe Set ID	Gene Symbol	Gene Title
202260_s_at	STXBP1	syntaxin binding protein 1
202917_s_at	S100A8	S100 calcium binding protein A8
203881_s_at	DMD	dystrophin (muscular dystrophy, Duchenne and Becker types)
203967_at	CDC6	cell division cycle 6 homolog (S. cerevisiae)

204497_at	ADCY9	adenylate cyclase 9
204753_s_at	HLF	hepatic leukemia factor
205394_at	CHEK1	CHK1 checkpoint homolog (S. pombe)
205542_at	STEAP1	six transmembrane epithelial antigen of the prostate 1
206896_s_at	GNG7	guanine nucleotide binding protein (G protein), gamma 7
208025_s_at	HMGA2	high mobility group AT-hook 2
208454_s_at	PGCP	plasma glutamate carboxypeptidase
208502_s_at	PITX1	paired-like homeodomain 1
208772_at	ANKHD1 /// MASK-BP3	ankyrin repeat and KH domain containing 1 /// MASK-4E-BP3 alternate reading frame gene
209368_at	EPHX2	epoxide hydrolase 2, cytoplasmic
209460_at	ABAT	4-aminobutyrate aminotransferase
209487_at	RBPMS	RNA binding protein with multiple splicing
209488_s_at	RBPMS	RNA binding protein with multiple splicing
210762_s_at	DLC1	deleted in liver cancer 1
212741_at	MAOA	monoamine oxidase A
213643_s_at	INPP5B	inositol polyphosphate-5-phosphatase, 75kDa
213804_at	INPP5B	inositol polyphosphate-5-phosphatase, 75kDa
214370_at	S100A8	S100 calcium binding protein A8
214764_at	RRP15	ribosomal RNA processing 15 homolog (S. cerevisiae)
215054_at	EPOR	erythropoietin receptor
222351_at	PPP2R1B	protein phosphatase 2 (formerly 2A), regulatory subunit A, beta isoform
40093_at	BCAM	basal cell adhesion molecule (Lutheran blood group)

Method D

Probe Set ID	Gene Symbol	Gene Title
201250_s_at	SLC2A1	solute carrier family 2 (facilitated glucose transporter), member 1
201998_at	ST6GAL1	ST6 beta-galactosamide alpha-2,6-sialyltranferase 1
202917_s_at	S100A8	S100 calcium binding protein A8
203439_s_at	STC2	stanniocalcin 2
203881_s_at	DMD	dystrophin (muscular dystrophy, Duchenne and Becker types)
204388_s_at	MAOA	monoamine oxidase A
204420_at	FOSL1	FOS-like antigen 1
204497_at	ADCY9	adenylate cyclase 9
204687_at	DKFZP564O0823	DKFZP564O0823 protein
204740_at	CNKSR1	connector enhancer of kinase suppressor of Ras 1
204754_at	HLF	hepatic leukemia factor
205518_s_at	CMAH	cytidine monophosphate-N-acetylneuraminic acid hydroxylase (CMP-N-acetylneuraminate monooxygenase)

206754_s_at	CYP2B7P1	cytochrome P450, family 2, subfamily B, polypeptide 7 pseudogene 1
207855_s_at	CLCC1	chloride channel CLIC-like 1
208025_s_at	HMGA2	high mobility group AT-hook 2
208454_s_at	PGCP	plasma glutamate carboxypeptidase
209030_s_at	CADM1	cell adhesion molecule 1
209031_at	CADM1	cell adhesion molecule 1
209441_at	RHOBTB2	Rho-related BTB domain containing 2
209459_s_at	ABAT	4-aminobutyrate aminotransferase
209460_at	ABAT	4-aminobutyrate aminotransferase
209488_s_at	RBPMS	RNA binding protein with multiple splicing
211596_s_at	LRIG1	leucine-rich repeats and immunoglobulin-like domains 1
212389_at	SBF1	SET binding factor 1
213355_at	ST3GAL6	ST3 beta-galactoside alpha-2,3-sialyltransferase 6
213582_at	ATP11A	ATPase, Class VI, type 11A
213643_s_at	INPP5B	inositol polyphosphate-5-phosphatase, 75kDa
213804_at	INPP5B	inositol polyphosphate-5-phosphatase, 75kDa
214370_at	S100A8	S100 calcium binding protein A8
214385_s_at	MUC5AC	mucin 5AC, oligomeric mucus/gel-forming
214428_x_at	C4A /// C4B	complement component 4A (Rodgers blood group) /// complement component 4B (Childo blood group)
215054_at	EPOR	erythropoietin receptor
215059_at	C4orf31	Chromosome 4 open reading frame 31
215268_at	KIAA0754	hypothetical LOC643314
215555_at	---	CDNA FLJ13712 fis, clone PLACE2000394
219922_s_at	LTBP3	latent transforming growth factor beta binding protein 3
220431_at	TMPRSS11E /// TMPRSS11E2	transmembrane protease, serine 11E /// transmembrane protease, serine 11E2
220512_at	DLC1	deleted in liver cancer 1
220621_at	FOXE3	forkhead box E3
222073_at	COL4A3	collagen, type IV, alpha 3 (Goodpasture antigen)
222351_at	PPP2R1B	protein phosphatase 2 (formerly 2A), regulatory subunit A, beta isoform
37986_at	EPOR	erythropoietin receptor

Method H gene set

Probe Set ID	Gene Symbol	Gene Title
1053_at	RFC2	replication factor C (activator 1) 2, 40kDa
200783_s_at	STMN1	stathmin 1/oncoprotein 18
200799_at	HSPA1A	heat shock 70kDa protein 1A
200800_s_at	HSPA1A /// HSPA1B	heat shock 70kDa protein 1A /// heat shock 70kDa protein 1B
200853_at	H2AFZ	H2A histone family, member Z
200913_at	PPM1G	protein phosphatase 1G (formerly 2C), magnesium-dependent, gamma isoform
200934_at	DEK	DEK oncogene (DNA binding)
201088_at	KPNA2 /// LOC643995	karyopherin alpha 2 (RAG cohort 1, importin alpha 1) /// similar to Importin alpha-2 subunit (Karyopherin alpha-2 subunit) (SRP1-alpha) (RAG cohort protein 1)
201090_x_at	K-ALPHA-1	alpha tubulin
201091_s_at	CBX3 /// LOC653972	chromobox homolog 3 (HP1 gamma homolog, Drosophila) /// similar to chromobox homolog 3
201111_at	CSE1L	CSE1 chromosome segregation 1-like (yeast)
201112_s_at	CSE1L	CSE1 chromosome segregation 1-like (yeast)
201202_at	PCNA	proliferating cell nuclear antigen
201291_s_at	TOP2A	topoisomerase (DNA) II alpha 170kDa
201292_at	TOP2A	topoisomerase (DNA) II alpha 170kDa
201475_x_at	MARS	methionine-tRNA synthetase
201477_s_at	RRM1	ribonucleotide reductase M1 polypeptide
201478_s_at	DKC1	dyskeratosis congenita 1, dyskerin
201479_at	DKC1	dyskeratosis congenita 1, dyskerin
201535_at	UBL3	ubiquitin-like 3
201555_at	MCM3	MCM3 minichromosome maintenance deficient 3 (S. cerevisiae)
201584_s_at	DDX39	DEAD (Asp-Glu-Ala-Asp) box polypeptide 39
201697_s_at	DNMT1	DNA (cytosine-5-)-methyltransferase 1
201710_at	MYBL2	v-myb myeloblastosis viral oncogene homolog (avian)-like 2
201761_at	MTHFD2	methylenetetrahydrofolate dehydrogenase (NADP+ dependent) 2, methenyltetrahydrofolate cyclohydrolase
201762_s_at	PSME2	proteasome (prosome, macropain) activator subunit 2 (PA28 beta)
201770_at	SNRPA	small nuclear ribonucleoprotein polypeptide A
201774_s_at	CNAP1	chromosome condensation-related SMC-associated protein 1

201833_at	HDAC2	histone deacetylase 2
201890_at	RRM2	ribonucleotide reductase M2 polypeptide
201896_s_at	PSRC1	proline/serine-rich coiled-coil 1
201897_s_at	CKS1B	CDC28 protein kinase regulatory subunit 1B
201930_at	MCM6	MCM6 minichromosome maintenance deficient 6 (MIS5 homolog, S. pombe) (S. cerevisiae)
202094_at	BIRC5	baculoviral IAP repeat-containing 5 (survivin)
202095_s_at	BIRC5	baculoviral IAP repeat-containing 5 (survivin)
202105_at	IGBP1	immunoglobulin (CD79A) binding protein 1
202107_s_at	MCM2	MCM2 minichromosome maintenance deficient 2, mitotin (S. cerevisiae)
202153_s_at	NUP62	nucleoporin 62kDa
202338_at	TK1	thymidine kinase 1, soluble
202483_s_at	RANBP1	RAN binding protein 1
202503_s_at	KIAA0101	KIAA0101
202580_x_at	FOXM1	forkhead box M1
202589_at	TYMS	thymidylate synthetase
202633_at	TOPBP1	topoisomerase (DNA) II binding protein 1
202666_s_at	ACTL6A	actin-like 6A
202715_at	CAD /// ARHGEF5 /// LOC653691	carbamoyl-phosphate synthetase 2, aspartate transcarbamylase, and dihydroorotase /// Rho guanine nucleotide exchange factor (GEF) 5 /// FLJ40722-like
202726_at	LIG1	ligase I, DNA, ATP-dependent
202738_s_at	PHKB	phosphorylase kinase, beta
202754_at	R3HDM1	R3H domain containing 1
202779_s_at	UBE2S /// LOC651816	ubiquitin-conjugating enzyme E2S /// similar to Ubiquitin-conjugating enzyme E2S (Ubiquitin-conjugating enzyme E2-24 kDa) (Ubiquitin-protein ligase) (Ubiquitin carrier protein) (E2-EPF5)
202854_at	HPRT1	hypoxanthine phosphoribosyltransferase 1 (Lesch-Nyhan syndrome)
202870_s_at	CDC20	CDC20 cell division cycle 20 homolog (S. cerevisiae)
202904_s_at	LSM5	LSM5 homolog, U6 small nuclear RNA associated (S. cerevisiae)
202911_at	MSH6	mutS homolog 6 (E. coli)
202954_at	PAK3 /// UBE2C	p21 (CDKN1A)-activated kinase 3 /// ubiquitin-conjugating enzyme E2C
202983_at	SMARCA3	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 3
203022_at	RNASE	ribonuclease H2, subunit A

	H2A	
203046_s_at	TIMELESS	timeless homolog (Drosophila)
203087_s_at	KIF2	kinesin heavy chain member 2
203145_at	SPAG5	sperm associated antigen 5
203209_at	RFC5	replication factor C (activator 1) 5, 36.5kDa
203210_s_at	RFC5	replication factor C (activator 1) 5, 36.5kDa
203213_at	CDC2	cell division cycle 2, G1 to S and G2 to M
203214_x_at	CDC2	cell division cycle 2, G1 to S and G2 to M
203270_at	DTYMK /// LOC653208	deoxythymidylate kinase (thymidylate kinase) /// similar to deoxythymidylate kinase (thymidylate kinase)
203276_at	LMNB1	lamin B1
203302_at	DCK	deoxycytidine kinase
203358_s_at	EZH2	enhancer of zeste homolog 2 (Drosophila)
203362_s_at	MAD2L1	MAD2 mitotic arrest deficient-like 1 (yeast)
203414_at	MMD	monocyte to macrophage differentiation-associated
203418_at	CCNA2	cyclin A2
203420_at	FAM8A1	family with sequence similarity 8, member A1
203432_at	TMPO	thymopoietin
203554_x_at	PTTG1	pituitary tumor-transforming 1
203693_s_at	E2F3	E2F transcription factor 3
203696_s_at	RFC2	replication factor C (activator 1) 2, 40kDa
203755_at	BUB1B	BUB1 budding uninhibited by benzimidazoles 1 homolog beta (yeast)
203764_at	DLG7	discs, large homolog 7 (Drosophila)
203832_at	SNRPF	small nuclear ribonucleoprotein polypeptide F
203856_at	VRK1	vaccinia related kinase 1
203967_at	CDC6	CDC6 cell division cycle 6 homolog (S. cerevisiae)
203968_s_at	CDC6	CDC6 cell division cycle 6 homolog (S. cerevisiae)
203976_s_at	CHAF1A	chromatin assembly factor 1, subunit A (p150)
204023_at	RFC4	replication factor C (activator 1) 4, 37kDa
204026_s_at	ZWINT	ZW10 interactor
204033_at	TRIP13	thyroid hormone receptor interactor 13
204092_s_at	AURKA	aurora kinase A
204126_s_at	CDC45L	CDC45 cell division cycle 45-like (S. cerevisiae)
204127_at	RFC3	replication factor C (activator 1) 3, 38kDa
204128_s_at	RFC3	replication factor C (activator 1) 3, 38kDa
204146_at	RAD51AP1	RAD51 associated protein 1

204162_at	KNTC2	kinetochore associated 2
204165_at	WASF1	WAS protein family, member 1
204170_s_at	CKS2	CDC28 protein kinase regulatory subunit 2
204240_s_at	SMC2	structural maintenance of chromosomes 2
204244_s_at	DBF4	DBF4 homolog (S. cerevisiae)
204315_s_at	GTSE1	G-2 and S-phase expressed 1
204317_at	GTSE1	G-2 and S-phase expressed 1
204318_s_at	GTSE1	G-2 and S-phase expressed 1
204407_at	TTF2	transcription termination factor, RNA polymerase II
204444_at	KIF11	kinesin family member 11
204492_at	ARHGAP11A	Rho GTPase activating protein 11A
204510_at	CDC7	CDC7 cell division cycle 7 (S. cerevisiae)
204558_at	RAD54L	RAD54-like (S. cerevisiae)
204641_at	NEK2	NIMA (never in mitosis gene a)-related kinase 2
204649_at	TROAP	trophinin associated protein (tastin)
204695_at	CDC25A	cell division cycle 25A
204709_s_at	KIF23	kinesin family member 23
204727_at	WDHD1	WD repeat and HMG-box DNA binding protein 1
204728_s_at	WDHD1	WD repeat and HMG-box DNA binding protein 1
204752_x_at	PARP2	poly (ADP-ribose) polymerase family, member 2
204766_s_at	NUDT1	nudix (nucleoside diphosphate linked moiety X)-type motif 1
204767_s_at	FEN1	flap structure-specific endonuclease 1
204768_s_at	FEN1	flap structure-specific endonuclease 1
204822_at	TTK	TTK protein kinase
204825_at	MELK	maternal embryonic leucine zipper kinase
204886_at	PLK4	polo-like kinase 4 (Drosophila)
204887_s_at	PLK4	polo-like kinase 4 (Drosophila)
204947_at	E2F1	E2F transcription factor 1
204962_s_at	CENPA	centromere protein A
205034_at	CCNE2	cyclin E2
205046_at	CENPE	centromere protein E, 312kDa
205053_at	PRIM1	primase, polypeptide 1, 49kDa
205063_at	SIP1	survival of motor neuron protein interacting protein 1
205085_at	ORC1L	origin recognition complex, subunit 1-like (yeast)
205167_s_at	CDC25C	cell division cycle 25C
205234_at	SLC16A4	solute carrier family 16, member 4 (monocarboxylic acid transporter 5)
205296_at	---	---
205393_s_at	CHEK1	CHK1 checkpoint homolog (S. pombe)

205394_at	CHEK1	CHK1 checkpoint homolog (S. pombe)
205395_s_at	MRE11A	MRE11 meiotic recombination 11 homolog A (S. cerevisiae)
205436_s_at	H2AFX	H2A histone family, member X
205644_s_at	SNRPG	small nuclear ribonucleoprotein polypeptide G
205733_at	BLM	Bloom syndrome
206055_s_at	SNRPA1	small nuclear ribonucleoprotein polypeptide A'
206074_s_at	HMGA1	high mobility group AT-hook 1
206102_at	GINS1	GINS complex subunit 1 (Psf1 homolog)
206272_at	SPHAR	S-phase response (cyclin-related)
206316_s_at	KNTC1	kinetochore associated 1
206364_at	KIF14	kinesin family member 14
206445_s_at	PRMT1	protein arginine methyltransferase 1
206550_s_at	NUP155	nucleoporin 155kDa
207165_at	HMMR	hyaluronan-mediated motility receptor (RHAMM)
207183_at	GPR19	G protein-coupled receptor 19
207268_x_at	ABI2	abl interactor 2
207416_s_at	NFATC3	nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 3
207740_s_at	NUP62	nucleoporin 62kDa
207828_s_at	CENPF	centromere protein F, 350/400ka (mitosin)
208079_s_at	AURKA	aurora kinase A
208149_x_at	DDX11 /// LOC652053	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 11 (CHL1-like helicase homolog, S. cerevisiae) /// DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 11 (CHL1-like helicase homolog, S. cerevisiae) /// similar to DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 11 isoform 1 /// similar to DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 11 isoform 1
208159_x_at	DDX11	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 11 (CHL1-like helicase homolog, S. cerevisiae)
208696_at	CCT5	chaperonin containing TCP1, subunit 5 (epsilon)
208741_at	SAP18	Sin3A-associated protein, 18kDa
208760_at	UBE2I	Ubiquitin-conjugating enzyme E2I (UBC9 homolog, yeast)
208765_s_at	HNRPR	heterogeneous nuclear ribonucleoprotein R
208766_s_at	HNRPR	heterogeneous nuclear ribonucleoprotein R
208795_s_at	MCM7	MCM7 minichromosome maintenance deficient 7 (S. cerevisiae)
208808_s_at	HMGB2	high-mobility group box 2
208821_at	SNRPB	small nuclear ribonucleoprotein polypeptides B and B1
208931_s_at	ILF3	interleukin enhancer binding factor 3, 90kDa
209026_x_at	TUBB	tubulin, beta
209046_s_at	GABARAPL2	GABA(A) receptor-associated protein-like 2
209052_s_at	WHSC1	Wolf-Hirschhorn syndrome candidate 1
209053_s_at	WHSC1	Wolf-Hirschhorn syndrome candidate 1

209068_at	HNRPDL	heterogeneous nuclear ribonucleoprotein D-like
209172_s_at	CENPF	centromere protein F, 350/400ka (mitosin) /// centromere protein F, 350/400ka (mitosin)
209251_x_at	TUBA6	tubulin, alpha 6
209375_at	XPC	xeroderma pigmentosum, complementation group C
209408_at	KIF2C	kinesin family member 2C
209464_at	AURKB	aurora kinase B
209642_at	BUB1	BUB1 budding uninhibited by benzimidazoles 1 homolog (yeast)
209680_s_at	KIFC1	kinesin family member C1
209714_s_at	CDKN3	cyclin-dependent kinase inhibitor 3 (CDK2-associated dual specificity phosphatase)
209773_s_at	RRM2	ribonucleotide reductase M2 polypeptide
209832_s_at	CDT1	chromatin licensing and DNA replication factor 1
209856_x_at	ABI2	abl interactor 2
209891_at	SPBC25	spindle pole body component 25 homolog (S. cerevisiae)
210052_s_at	TPX2	TPX2, microtubule-associated, homolog (Xenopus laevis)
210115_at	RPL39L	ribosomal protein L39-like
210175_at	C2orf3	chromosome 2 open reading frame 3
210334_x_at	BIRC5	baculoviral IAP repeat-containing 5 (survivin)
210527_x_at	TUBA2	tubulin, alpha 2
210559_s_at	CDC2	cell division cycle 2, G1 to S and G2 to M
210568_s_at	RECQL	RecQ protein-like (DNA helicase Q1-like)
210766_s_at	CSE1L	CSE1 chromosome segregation 1-like (yeast)
210983_s_at	MCM7	MCM7 minichromosome maintenance deficient 7 (S. cerevisiae)
211058_x_at	K-ALPHA-1	alpha tubulin /// alpha tubulin
211072_x_at	K-ALPHA-1	alpha tubulin /// alpha tubulin
211080_s_at	NEK2	NIMA (never in mitosis gene a)-related kinase 2 /// NIMA (never in mitosis gene a)-related kinase 2
211375_s_at	ILF3	interleukin enhancer binding factor 3, 90kDa
211519_s_at	KIF2C	kinesin family member 2C
211714_x_at	TUBB	tubulin, beta /// tubulin, beta
211747_s_at	LSM5	LSM5 homolog, U6 small nuclear RNA associated (S. cerevisiae) /// LSM5 homolog, U6 small nuclear RNA associated (S. cerevisiae)
211750_x_at	TUBA6	tubulin, alpha 6 /// tubulin, alpha 6
211762_s_at	KPNA2 /// LOC643995	karyopherin alpha 2 (RAG cohort 1, importin alpha 1) /// karyopherin alpha 2 (RAG cohort 1, importin alpha 1) /// similar to Importin alpha-2 subunit (Karyopherin alpha-2 subunit) (SRP1-alpha) (RAG cohort protein 1) /// similar to Importin alpha-2 subunit (Karyopherin alpha-2 subunit) (SRP1-alpha) (RAG cohort protein 1)
211814_s_at	CCNE2	cyclin E2

211931_s_at	HNRPA3P1 /// HNRPA3 /// LOC643689 /// LOC647474	heterogeneous nuclear ribonucleoprotein A3 pseudogene 1 /// heterogeneous nuclear ribonucleoprotein A3 /// heterogeneous nuclear ribonucleoprotein A3 pseudogene /// similar to heterogeneous nuclear ribonucleoprotein A3
212020_s_at	MKI67	antigen identified by monoclonal antibody Ki-67
212022_s_at	MKI67	antigen identified by monoclonal antibody Ki-67
212023_s_at	MKI67	antigen identified by monoclonal antibody Ki-67
212202_s_at	TMEM87A	transmembrane protein 87A
212219_at	PSME4	proteasome (prosome, macropain) activator subunit 4
212247_at	NUP205	nucleoporin 205kDa
212297_at	ATP13A3	ATPase type 13A3
212315_s_at	NUP210	nucleoporin 210kDa
212316_at	NUP210	nucleoporin 210kDa
212331_at	RBL2	retinoblastoma-like 2 (p130)
212343_at	YIPF6	Yip1 domain family, member 6
212621_at	KIAA0286	KIAA0286 protein
212639_x_at	K-ALPHA-1	alpha tubulin
212832_s_at	CKAP5	cytoskeleton associated protein 5
212949_at	BRRN1	barren homolog 1 (Drosophila)
213007_at	KIAA1794	KIAA1794
213008_at	KIAA1794	KIAA1794
213088_s_at	DNAJC9	DnaJ (Hsp40) homolog, subfamily C, member 9
213175_s_at	SNRPB	small nuclear ribonucleoprotein polypeptides B and B1
213215_at	---	CDNA clone IMAGE:4157286
213226_at	CCNA2	cyclin A2
213253_at	SMC2L1	SMC2 structural maintenance of chromosomes 2-like 1 (yeast)
213346_at	C13orf27	chromosome 13 open reading frame 27
213453_x_at	GAPDH	glyceraldehyde-3-phosphate dehydrogenase
213520_at	RECQL4	RecQ protein-like 4
213646_x_at	K-	alpha tubulin

	ALPHA-1	
213911_s_at	H2AFZ	H2A histone family, member Z
213947_s_at	NUP210	nucleoporin 210kDa
213951_s_at	PSMC3IP	PSMC3 interacting protein
214028_x_at	TDRD3	tudor domain containing 3
214086_s_at	PARP2	poly (ADP-ribose) polymerase family, member 2
214202_at	---	MRNA from chromosome 5q21-22, clone:357Ex
214426_x_at	CHAF1A	chromatin assembly factor 1, subunit A (p150)
214431_at	GMPS	guanine monophosphate synthetase
214700_x_at	RIF1	RAP1 interacting factor homolog (yeast)
214710_s_at	CCNB1	cyclin B1
214727_at	BRCA2	breast cancer 2, early onset
214804_at	---	---
215006_at	EZH2	Enhancer of zeste homolog 2 (Drosophila)
215075_s_at	GRB2	growth factor receptor-bound protein 2
215286_s_at	PHTF2	putative homeodomain transcription factor 2
215509_s_at	BUB1	BUB1 budding uninhibited by benzimidazoles 1 homolog (yeast)
215773_x_at	PARP2	poly (ADP-ribose) polymerase family, member 2
216228_s_at	WDHD1	WD repeat and HMG-box DNA binding protein 1
216237_s_at	MCM5	MCM5 minichromosome maintenance deficient 5, cell division cycle 46 (S. cerevisiae)
216952_s_at	LMNB2	lamin B2
217094_s_at	ITCH	itchy homolog E3 ubiquitin protein ligase (mouse)
217640_x_at	C18orf24	chromosome 18 open reading frame 24
217714_x_at	STMN1	stathmin 1/oncoprotein 18
217805_at	ILF3	interleukin enhancer binding factor 3, 90kDa
218009_s_at	PRC1	protein regulator of cytokinesis 1
218039_at	NUSAP1	nucleolar and spindle associated protein 1
218073_s_at	TMEM48	transmembrane protein 48
218115_at	ASF1B	ASF1 anti-silencing function 1 homolog B (S. cerevisiae)
218142_s_at	CRBN	cereblon
218204_s_at	FYCO1	FYVE and coiled-coil domain containing 1
218252_at	CKAP2	cytoskeleton associated protein 2
218308_at	TACC3	transforming, acidic coiled-coil containing protein 3
218355_at	KIF4A	kinesin family member 4A
218381_s_at	U2AF2	U2 small nuclear RNA auxiliary factor 2
218432_at	FBXO3	F-box protein 3
218497_s_at	RNASE	ribonuclease H1

	H1	
218542_at	CEP55	centrosomal protein 55kDa
218585_s_at	DTL	denticleless homolog (Drosophila)
218602_s_at	FAM29A	family with sequence similarity 29, member A
218662_s_at	HCAP-G	chromosome condensation protein G
218663_at	HCAP-G	chromosome condensation protein G
218726_at	DKFZp762E1312	hypothetical protein DKFZp762E1312
218755_at	KIF20A	kinesin family member 20A
218782_s_at	ATAD2	ATPase family, AAA domain containing 2
218869_at	MLYCD	malonyl-CoA decarboxylase
218875_s_at	FBXO5	F-box protein 5
218883_s_at	MLF1IP	MLF1 interacting protein
218966_at	MYO5C	myosin VC
219000_s_at	DCC1	defective in sister chromatid cohesion homolog 1 (S. cerevisiae)
219135_s_at	LA16c-360B4.1	hypothetical protein FLJ12681
219148_at	PBK	PDZ binding kinase
219162_s_at	MRPL11	mitochondrial ribosomal protein L11
219306_at	KIF15 /// C7orf9	kinesin family member 15 /// chromosome 7 open reading frame 9
219493_at	SHCBP1	SHC SH2-domain binding protein 1
219506_at	C1orf54	chromosome 1 open reading frame 54
219512_at	C20orf172	chromosome 20 open reading frame 172
219556_at	C16orf59	chromosome 16 open reading frame 59
219588_s_at	LUZP5	leucine zipper protein 5
219650_at	FLJ20105	FLJ20105 protein
219787_s_at	ECT2	epithelial cell transforming sequence 2 oncogene
219918_s_at	ASPM	asp (abnormal spindle)-like, microcephaly associated (Drosophila)
219978_s_at	NUSAP1	nucleolar and spindle associated protein 1
219990_at	E2F8	E2F transcription factor 8
220060_s_at	C12orf48	chromosome 12 open reading frame 48

220085_at	HELLS	helicase, lymphoid-specific
220238_s_at	KLHL7	kelch-like 7 (Drosophila)
220239_at	KLHL7	kelch-like 7 (Drosophila)
220295_x_at	DEPDC1	DEP domain containing 1
220651_s_at	MCM10	MCM10 minichromosome maintenance deficient 10 (S. cerevisiae)
220753_s_at	CRYL1	crystallin, lambda 1
220788_s_at	RNF31	ring finger protein 31
220840_s_at	C1orf112	chromosome 1 open reading frame 112
221012_s_at	TRIM8	tripartite motif-containing 8 /// tripartite motif-containing 8
221156_x_at	CCPG1	cell cycle progression 1
221203_s_at	YEATS2	YEATS domain containing 2
221258_s_at	KIF18A	kinesin family member 18A /// kinesin family member 18A
221436_s_at	CDCA3	cell division cycle associated 3 /// cell division cycle associated 3
221476_s_at	RPL15	ribosomal protein L15
221511_x_at	CCPG1	cell cycle progression 1
221520_s_at	CDCA8	cell division cycle associated 8
221564_at	PRMT2	protein arginine methyltransferase 2
221677_s_at	DONSON	downstream neighbor of SON
221703_at	BRIP1	BRCA1 interacting protein C-terminal helicase 1 /// BRCA1 interacting protein C-terminal helicase 1
221825_at	ANGEL2	angel homolog 2 (Drosophila)
222036_s_at	MCM4	MCM4 minichromosome maintenance deficient 4 (S. cerevisiae)
222039_at	SLC35E1 /// LOC146909	solute carrier family 35, member E1 /// hypothetical protein LOC146909
222077_s_at	RACGAP1	Rac GTPase activating protein 1
38158_at	ESPL1	extra spindle poles like 1 (S. cerevisiae)
49452_at	ACACB	acetyl-Coenzyme A carboxylase beta
57703_at	SENP5	SUMO1/sentrin specific peptidase 5

Supplementary Materials 6: ROC curve analysis

Sensitivity/Specificity Analysis

To construct ROC curves based on censored outcome data, we used Bayes' theorem

together with Kaplan-Meier estimates of the hazard function. Specifically, to calculate the sensitivity for a rule based on a given risk-score cutoff k, we expressed P(R>k | T<t) = P(T<t | R>k)*P(R>k)/P(T<t), where T is the survival time, t=36 months is the follow-up time we used for this analysis, and R is a risk score. The terms P(T<t) and P(T<t | R>k) can be estimated using the Kaplan-Meier procedure, and P(R>k) is estimated directly with its empirical probability. When the conditioning sets {R>k} (for sensitivity) and {R<k} (for specificity) become small, the Kaplan-Meier estimates are very unstable, therefore we exclude split points where either of these sets contains fewer than five values. Similarly, the specificity P(R<k | T>t) = P(T>t | R<k)*P(R<k)/P(T>t) can be estimated. Since the resulting ROC curve may not be monotonic, we used isotonic regression to transform it into a monotonic curve.

For ROC plots, see Supplementary Figure S2.

Method C								
	All stages				Stage 1 only			
	MSK		CAN/DF		MSK		CAN/DF	
	Sens	Spec	Sens	Spec	Sens	Spec	Sens	Spec
>=0	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
>=1	0.39	0.67	0.26	0.75	0.50	0.68	0.19	0.72
>1	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00

Method D								
	All stages				Stage 1 only			
	MSK		CAN/DF		MSK		CAN/DF	
	Sens	Spec	Sens	Spec	Sens	Spec	Sens	Spec
>=0	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
>=1	0.30	0.80	0.43	0.83	0.49	0.82	0.40	0.83
>1	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00

Method E								
	All stages				Stage 1 only			
	MSK		CAN/DF		MSK		CAN/DF	
	Sens	Spec	Sens	Spec	Sens	Spec	Sens	Spec
>=0	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00

>=1	0.56	0.69	0.13	0.74	0.67	0.67	0.10	0.76
>1	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00

Method F								
	All stages				Stage 1 only			
	MSK		CAN/DF		MSK		CAN/DF	
	Sens	Spec	Sens	Spec	Sens	Spec	Sens	Spec
>=0	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
>=1	0.68	0.67	0.56	0.47	0.98	0.69	0.39	0.54
>1	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00

Method G								
	All stages				Stage 1 only			
	MSK		CAN/DF		MSK		CAN/DF	
	Sens	Spec	Sens	Spec	Sens	Spec	Sens	Spec
>=0	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
>=1	0.78	0.61	0.43	0.62	1.00	0.63	0.29	0.67
>1	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00

Method H								
	All stages				Stage 1 only			
	MSK		CAN/DF		MSK		CAN/DF	
	Sens	Spec	Sens	Spec	Sens	Spec	Sens	Spec
>=0	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
>=1	0.81	0.56	0.74	0.44	0.98	0.65	0.61	0.50
>=2	0.08	0.88	0.18	0.78	0.24	0.93	0.11	0.81
>2	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00