Multistate models for colon cancer recurrence and death with a cured fraction

Jeremy M. G. Taylor

Department of Biostatistics, University of Michigan

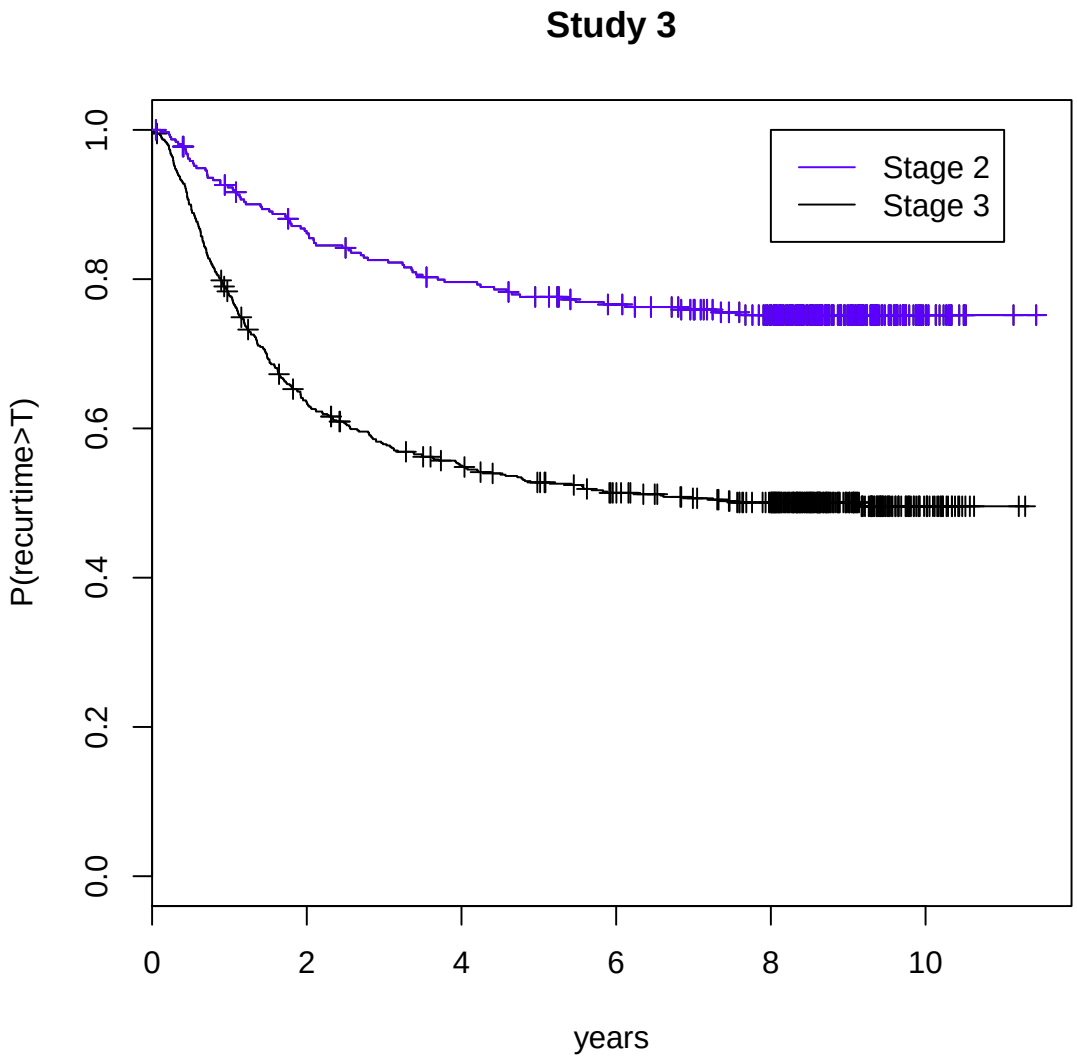
10th October 2013

- Introduction
  - Cure models
  - Multi-state models
  - Colon cancer data
- Model description
  - Model components
  - Likelihood and estimation
  - Results and parameter interpretation
- Utilizing the model
  - Prediction of marginal survival distributions
  - Prediction of conditional survival distributions
  - Using the predictions for efficiency gains

Cure models

- A generalization of standard survival analysis models
- Allows more detailed interpretation
  - Whether an event happens - logistic model
  - When an event happens - survival model
- Useful when
  - Scientific rationale for a cured group
  - Empirical evidence for a cured fraction
- Applicable in many cancer settings

Cured model illustration: Recurrence in cancer clinical trial



Multi-state model examples

Fig. 1 Progressive three-state model

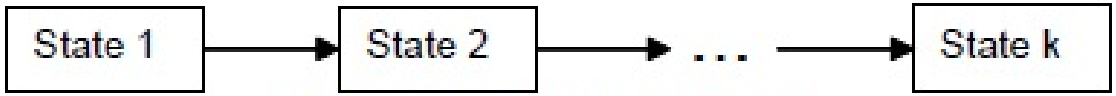


Fig. 2 Progressive k-state model

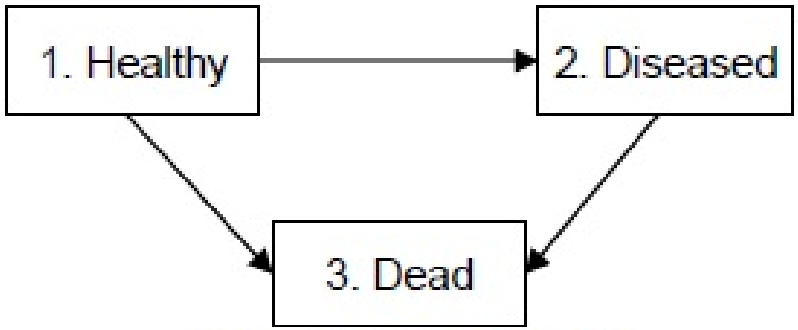


Fig. 3 Illness-death model

The setting

- Multiple randomized trials of colon cancer
- Two censored event times
  - Recurrence of the cancer
  - Death
- Primary endpoint of trials: treatment effect on time to death
- Cause of death not known
- Other covariates: stage of cancer, age of patient

Goal

- Build joint models for recurrence and death incorporating a cured fraction
  - Modeling the process that gave rise to the data
  - Incorporate the context into the model structure
  - Scientifically interpretable models
  - No emphasis given to statistically convenient or parsimonious models
- Use model for prediction and efficiency gain

Description of the data

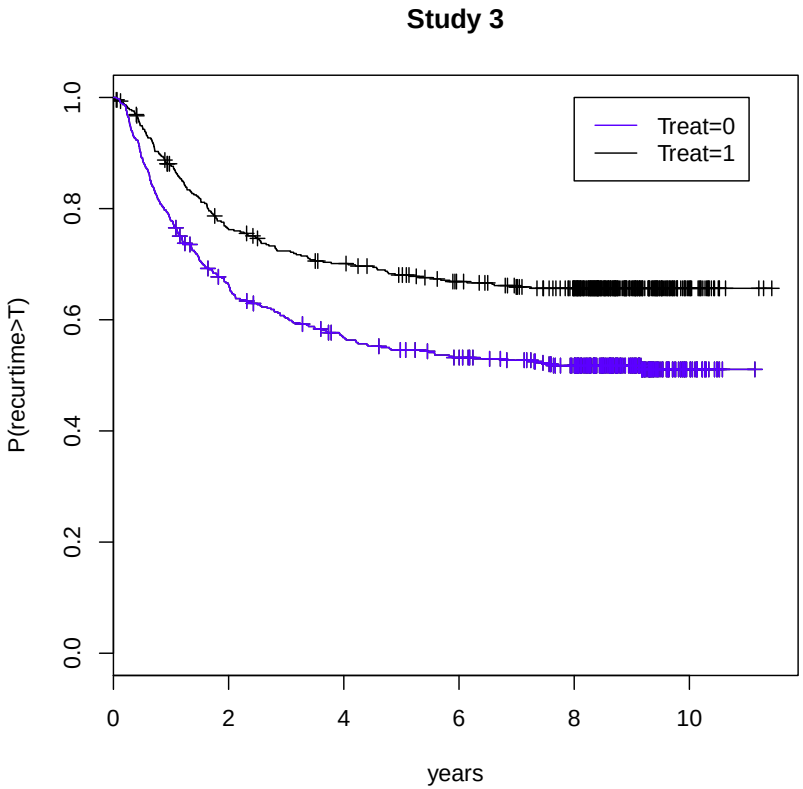
- Stage 2 and 3 colon cancer
  - Locally advanced cancer
- Twelve randomized trials
  - Surgery versus surgery + drug (5 trials)
  - Surgery + drug A versus surgery + drug B (7 trials)
- 13,983 patients in total
  - 4346 recurrences, 4990 deaths
- Most, but not all, censoring was administrative censoring
- Conlon et al, 2011, Clinical Trials

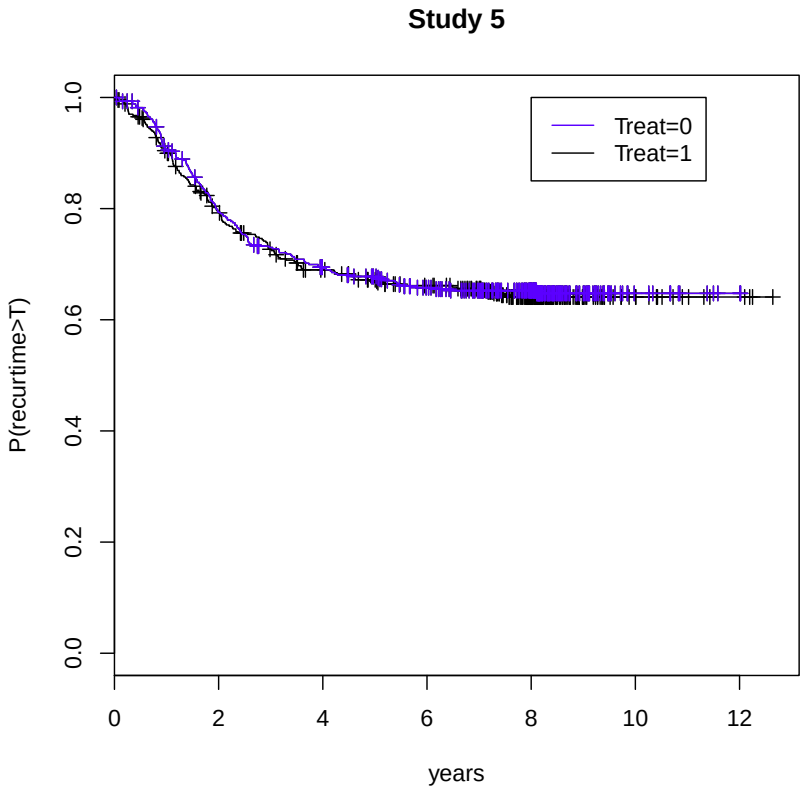
Data Summary

Study	N	Recur	Recur Without Death	Death Without Recur	Total Deaths	Longest Follow-Up (years)	% with Stage 3 Cancer	% in Trtmnt Group	Age (mean)
1	247	116	14	13	115	9.9	66	49	60
2	408	139	11	44	172	9.1	82	63	61
3	926	377	31	76	422	11.4	66	49	60
4	914	380	36	106	450	9.9	83	75	63
5	878	297	33	74	338	12.6	74	50	61
6	724	275	10	132	397	13.2	57	48	60
7	683	206	32	129	303	12.9	43	50	63
8	1040	356	36	67	387	9.7	72	50	56
9	2077	605	57	176	724	9.4	59	67	57
10	2128	574	66	192	700	10.3	56	50	58
11	1549	394	71	115	438	8.0	54	50	61
12	2409	627	189	106	544	6.0	71	50	58
Total	13983	4346	586	1230	4990	13.2	64	54	59

Plots of data

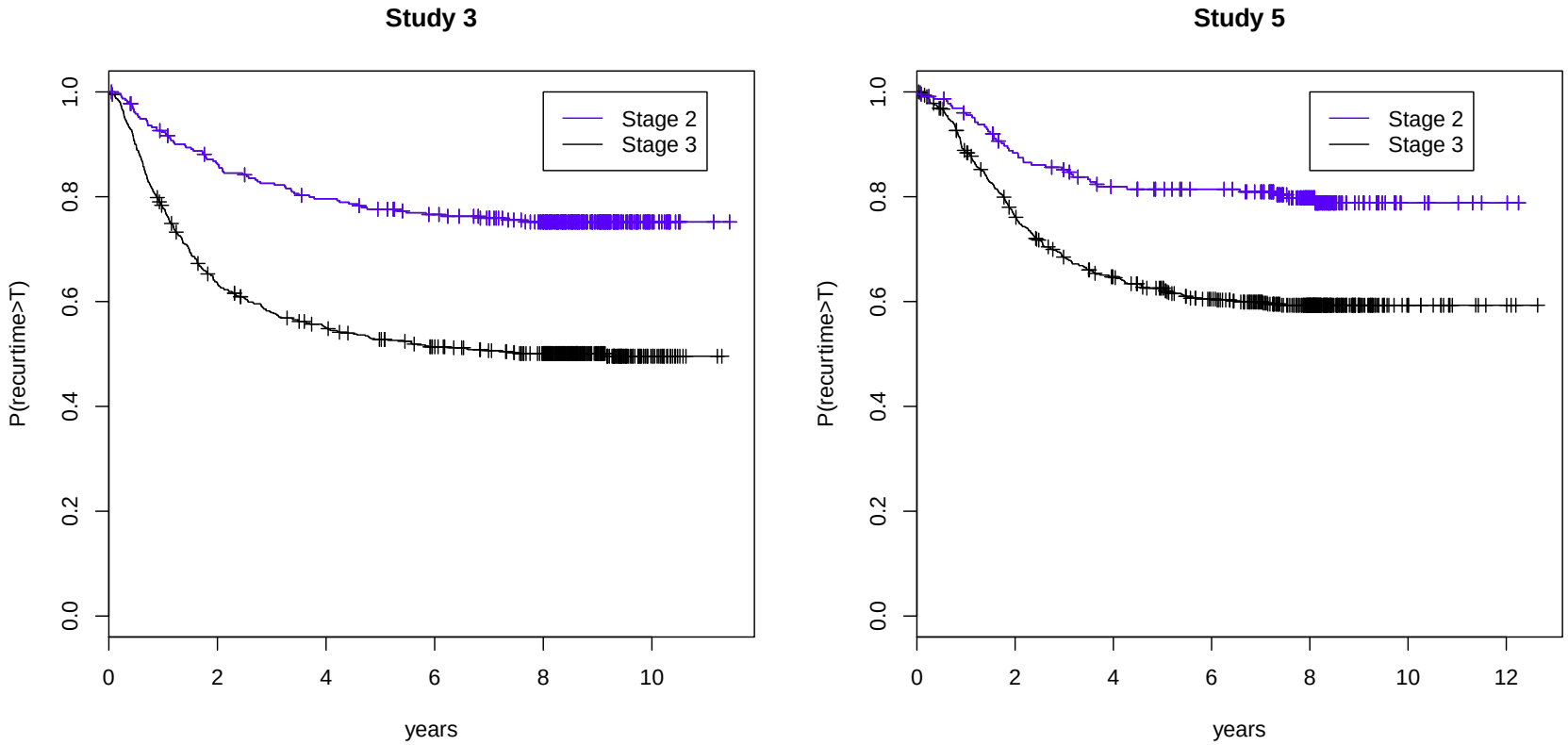
Time to recurrence by treatment





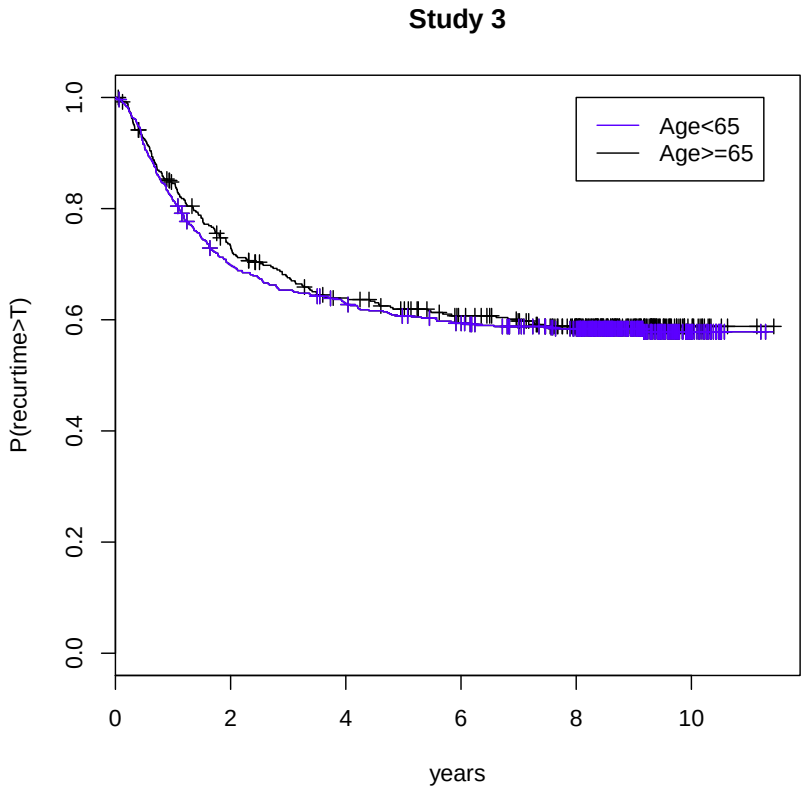
Plots of data

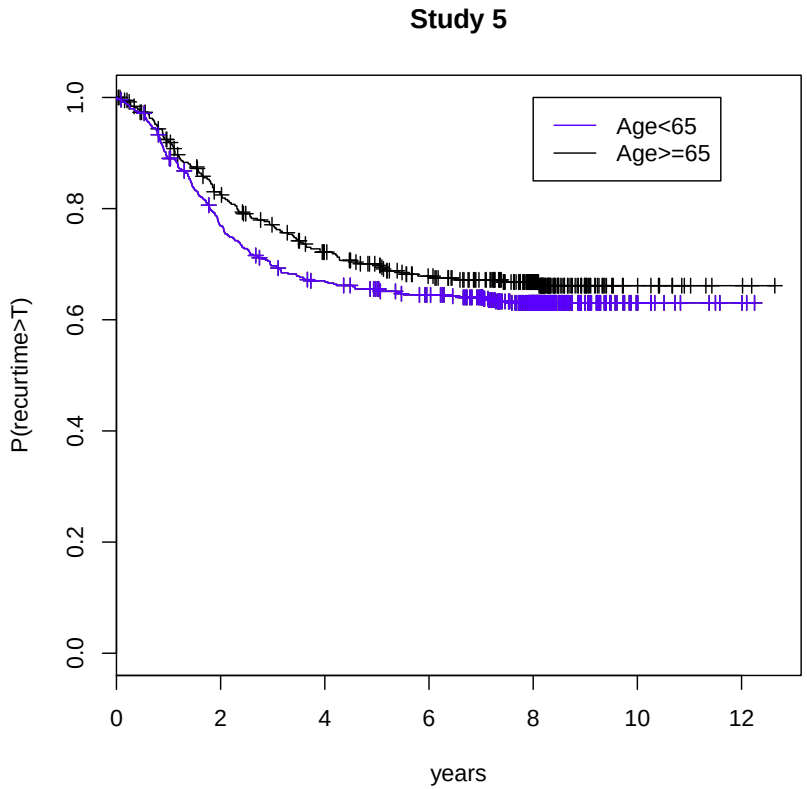
Time to recurrence by stage



Plots of data

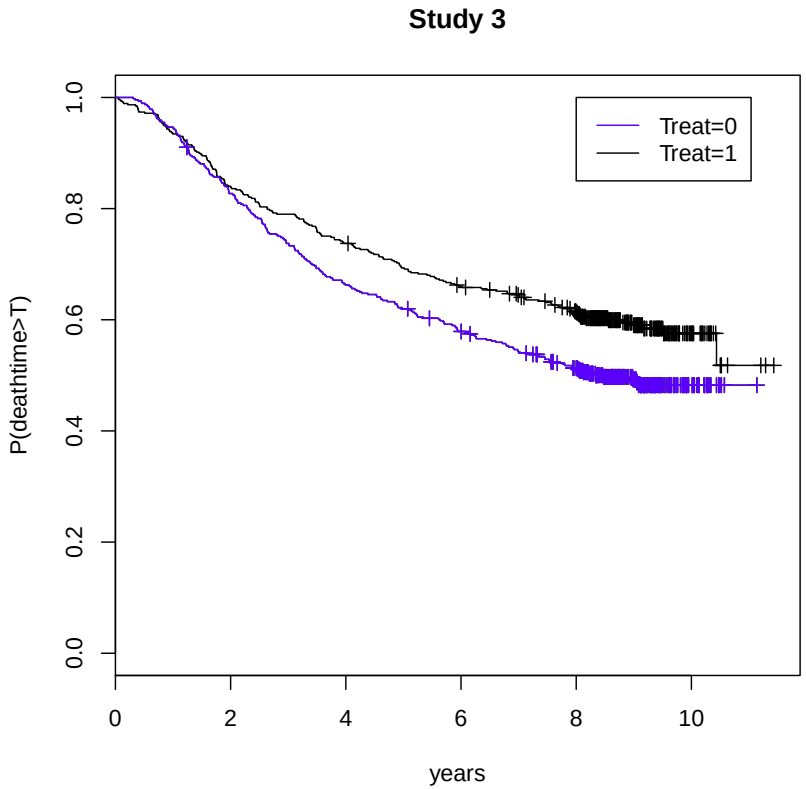
Time to recurrence by age

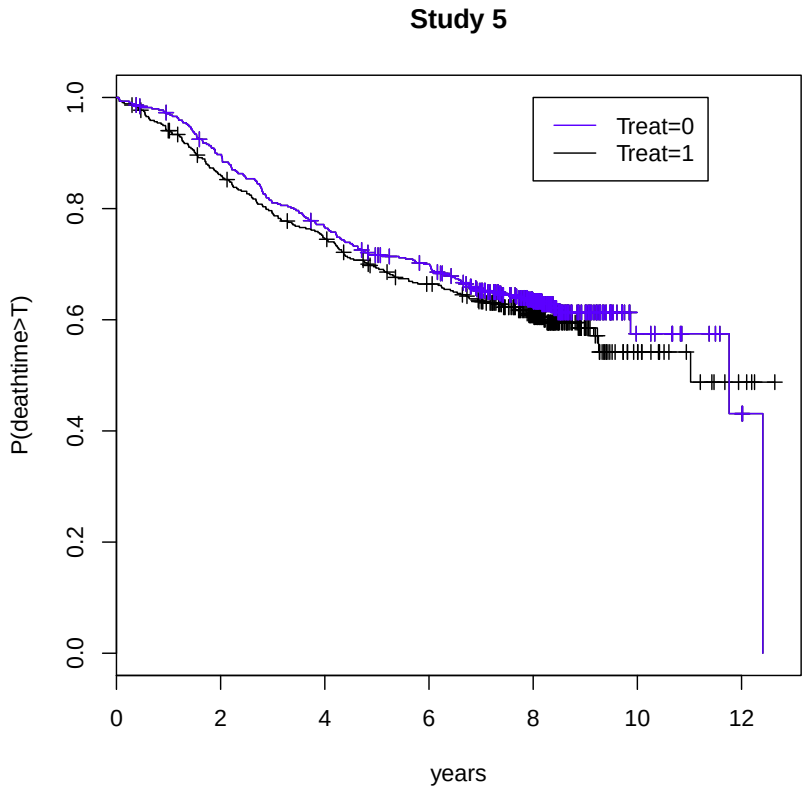




Plots of data

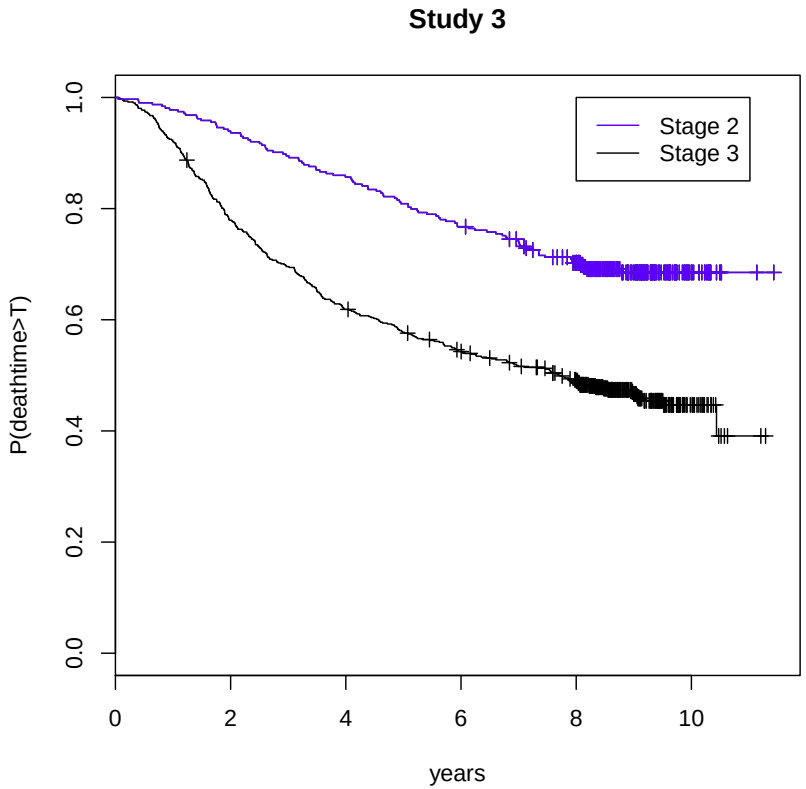
Time to death by treatment

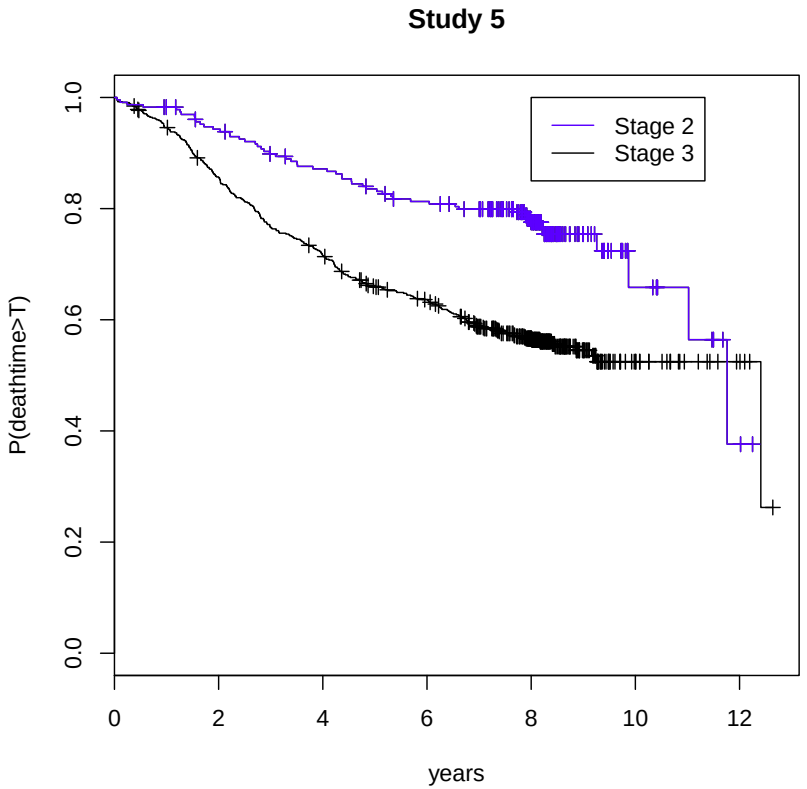




Plots of data

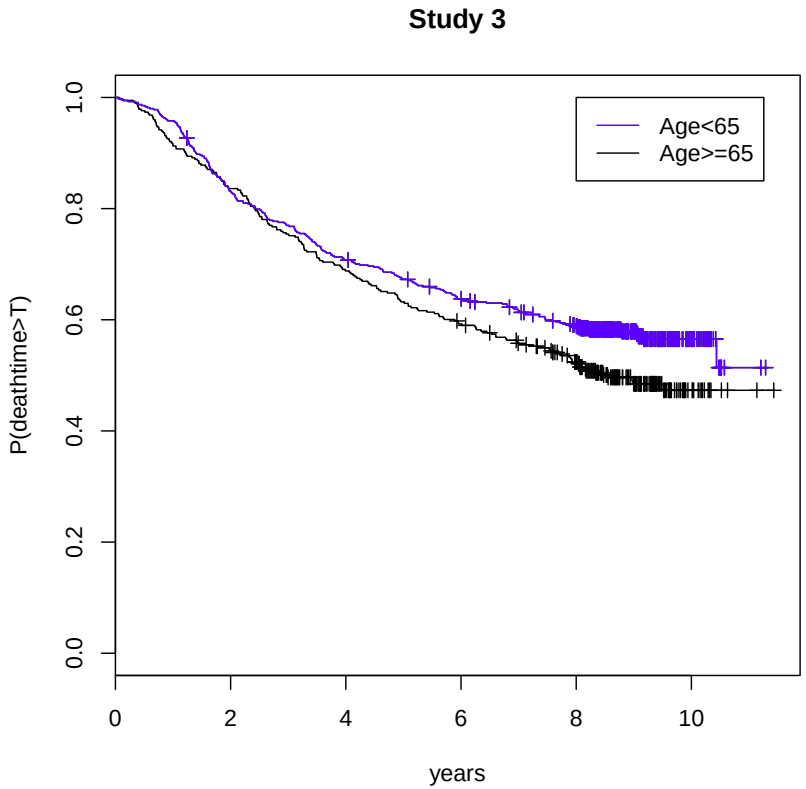
Time to death by stage

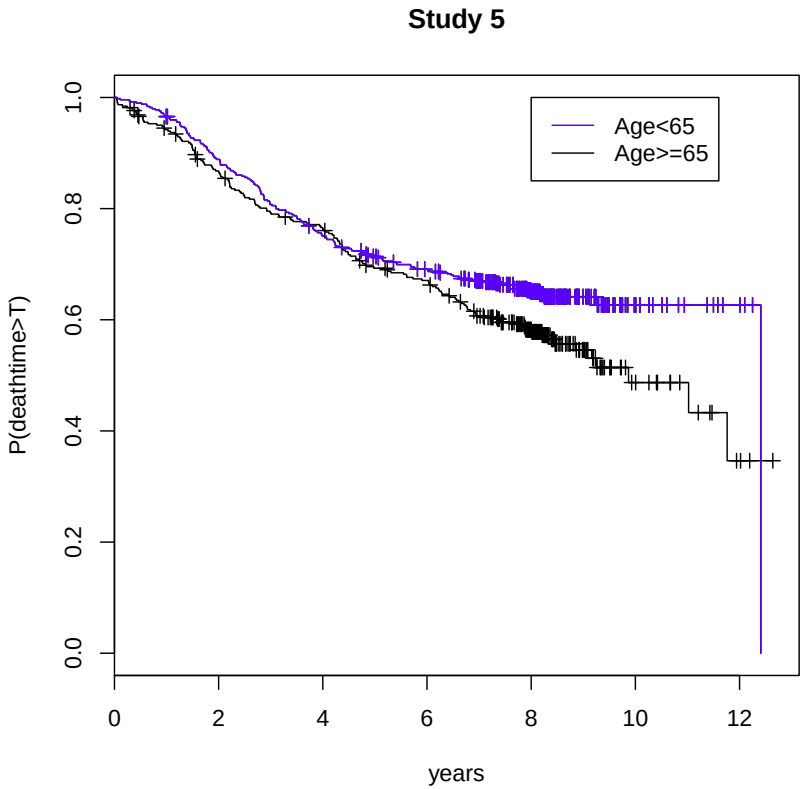




Plots of data

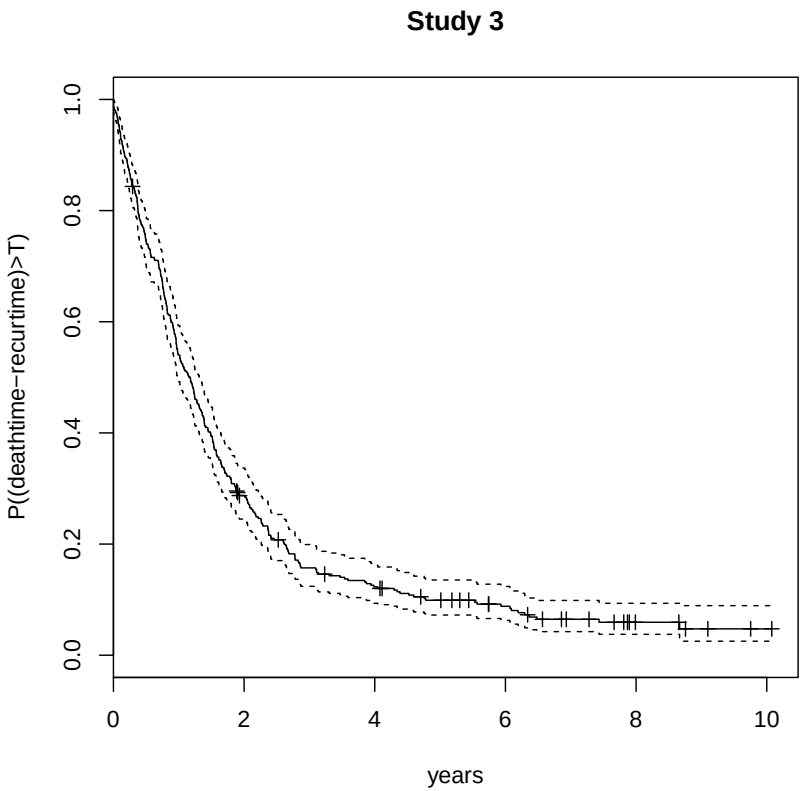
Time to death by age

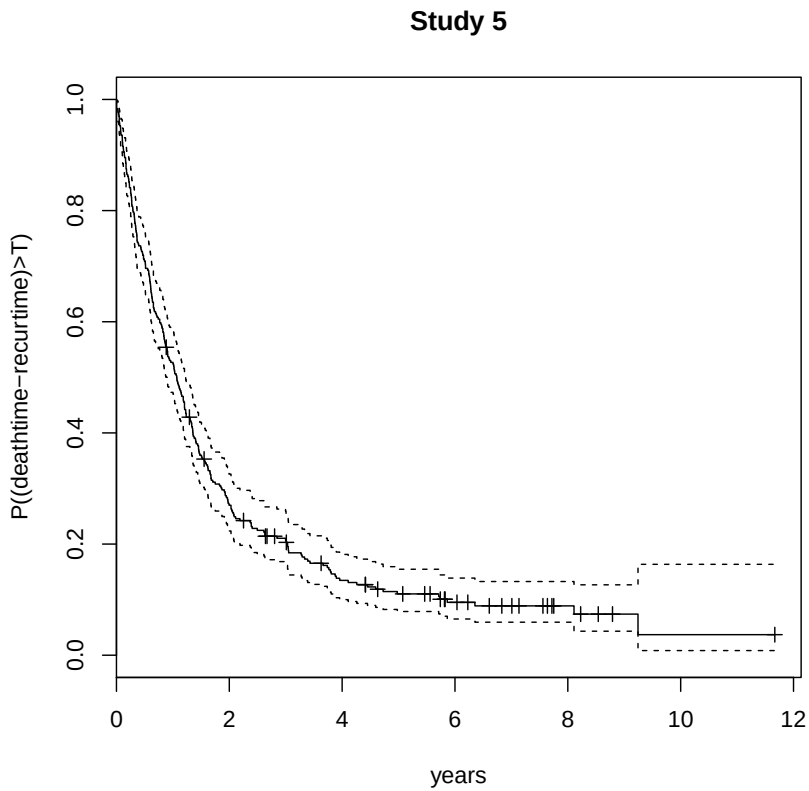




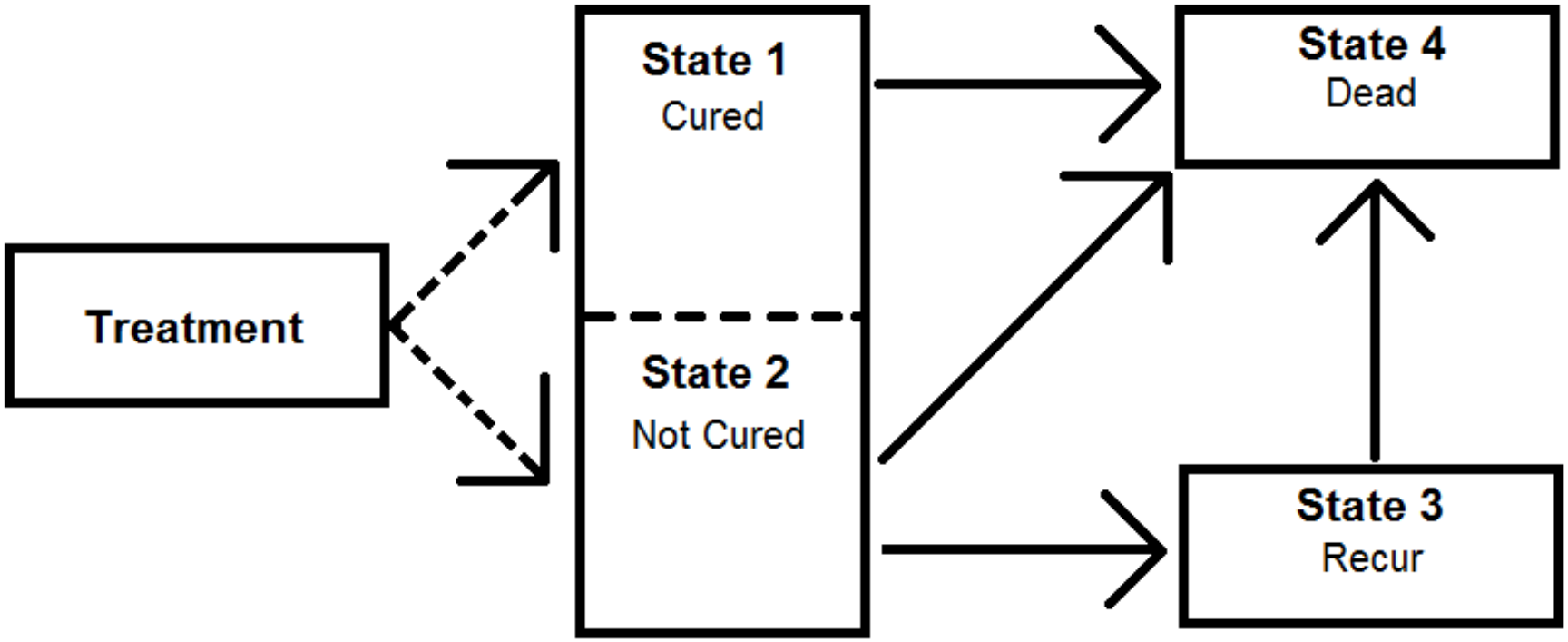
Plots of data

Time to death from recurrence





Graph of the multistate model



True States

- State 1: alive and cured of disease
- State 2: alive and not cured
- State 3: alive with recurrence
- State 4: death
- Don't fully observe progression through states
  - Did not recur and alive: 1 or 2
  - Recurred and alive: $2 \rightarrow 3$
  - Did not recur and dead: $1 \rightarrow 4$ or $2 \rightarrow 4$
  - Recurred and dead: $2 \rightarrow 3 \rightarrow 4$

Component models

- 5 components to the overall model
- Model for initial placement into state 1 or state 2
- Model four transition times:
  - $1 \rightarrow 4$
  - $2 \rightarrow 3$
  - $2 \rightarrow 4$
  - $3 \rightarrow 4$

Motivation for the model structure

- The treatment can eliminate the cancer, cured group
- Cure happens at the time of treatment, but not immediately observable
- If the cancer is not eliminated recurrence happens when the tumor has regrown to a detectable size
- Most recurrences within 4 years
- Almost no recurrences after a fixed time window, 7 years

Interpretation of the model components

- Pr(cure), about the tumor, cell killing effect of the treatment
- Hazard_{23}(recur|uncured), about tumor regrowth
- Hazard_{14}(death|cured), about the person, not about the cancer
- Hazard_{24}(death|non-cured, not yet recurred), mainly about the person, not much about the cancer
- Hazard_{34}(death|recurred), about the person and the tumor regrowth
- Covariates can be expected to be associated with these separate components in differing ways

Details of the models

- logistic for P(Cure),
  - $p_i = \frac{exp(\gamma X_i)}{1+exp(\gamma X_i)}$
- Weibull hazard models, state k to state j, semi-Markov

$$\lambda_{kj}(T_i; X_i) = \left(\frac{\rho_{kj}}{\lambda_{kj}}\right)\left(\frac{T_i}{\lambda_{kj}}\right)^{\rho_{kj}-1} exp\left(\beta_{kj} X_i\right)$$

- T_i is either a recurrence time or death time, measured from entry into current state
- Model for recur-to-death contains time-to-recurrence as a covariate
- 25 parameters

Details of the models: Data and notation

- $(Y_{ir}, \delta_{ir}, Y_{id}, \delta_{id}, X_i)$, $i = 1, ..., n$
- $Y_{ir} = min(T_{ir}, C_{ir}), Y_{id} = min(T_{id}, C_{id})$
- $G_i =$ latent variable for cured status (partially observed)
- $p_i = P(G_i = 1|X_i)$
- Censoring time for recurrence may differ from censoring time for death

Details of the models: Types of observations

- Different types of observations
  - recurred and alive
  - recurred and died
  - not recurred and alive
  - not recurred and died

Contribution to the likelihood

Recurred and Died:

$$(1-p_i)\lambda_{23}(Y_{ir})exp\left(-\int_0^{Y_{ir}}\lambda_{23}(u)du-\int_0^{Y_{ir}}\lambda_{24}(u)du\right)$$
$$\times\lambda_{34}(Y_{id}-Y_{ir})exp\left(-\int_0^{Y_{id}-Y_{ir}}\lambda_{34}(u)du\right)$$

where: $\lambda(t)=\left(\frac{\rho}{\lambda}\right)\left(\frac{t}{\lambda}\right)^{\rho-1}\exp(\beta X_i)$

Contribution to the likelihood

Not Recurred and Died, censored for recurrence at death time :

$$(1-p_i)\lambda_{24}(Y_{id})exp\left(-\int_0^{Y_{id}}\lambda_{24}(u)du-\int_0^{Y_{id}}\lambda_{23}(u)du\right)$$

$$+\,p_i\lambda_{14}(Y_{id})exp\left(-\int_0^{Y_{id}}\lambda_{14}(u)du\right)$$

Estimation: MCMC

- Metropolis-Hastings algorithm to draw parameters
  - Acceptance rate around 0.4
- Draw latent variable, G_i
- Mildly informative prior distributions
  - $log(\lambda) \sim N(0, 2^2)$
  - $\rho \sim Gamma(mean = 1, sd = 0.6)$
  - $\beta_{14\,Treat}$ and $\beta_{14Stage} \sim N(0, 0.25^2)$
  - remaining β's $\sim N(0, 1)$
- 50,000 iterations
- 5,000 draws from posterior distribution of each parameter

Parameter Estimates

Table: Parameter estimates (SD) for Trial 3

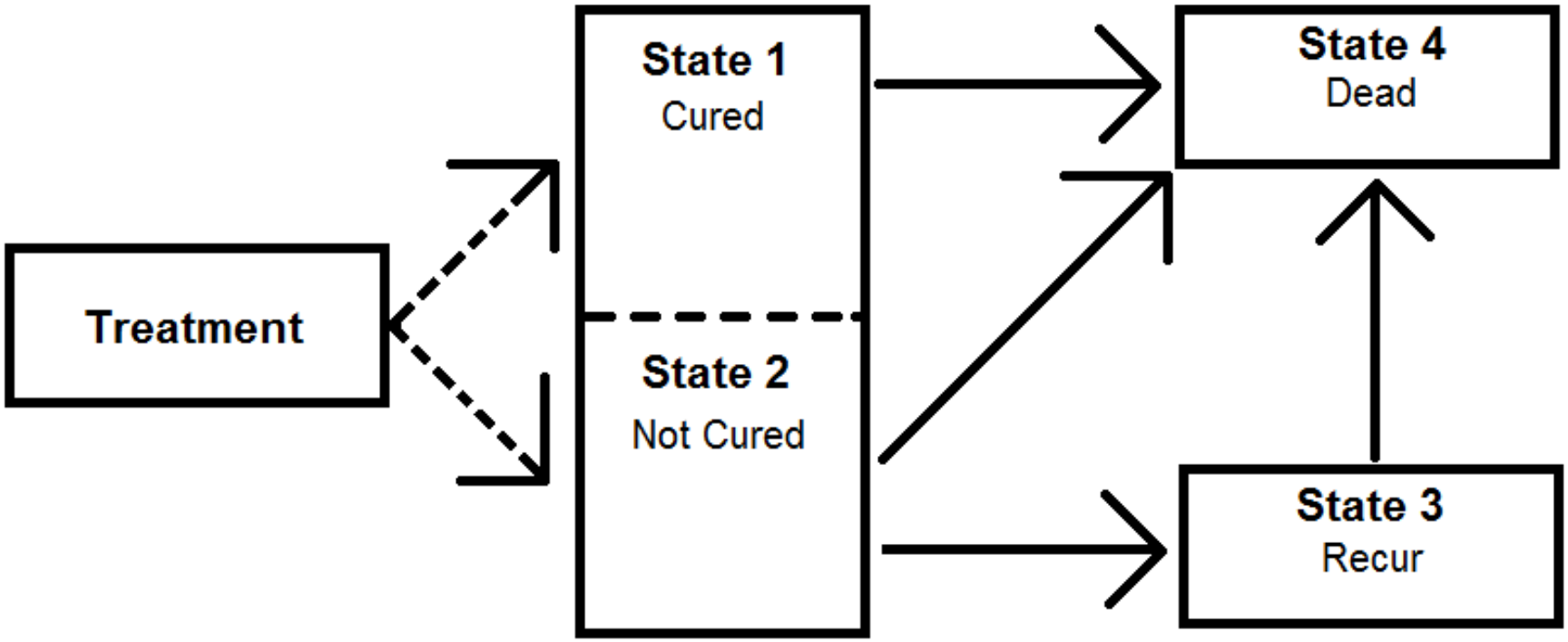
	Treatment	Stage	Age (10 yrs)	T_r
P(Cure)	0.72 (0.20)	-1.15 (0.19)	-0.08 (0.05)	-
Time to recur	-0.25 (0.12)	0.54 (0.14)	-0.11 (0.04)	-
Cure to death	0.41 (0.44)	-0.37 (0.42)	0.55 (0.20)	-
Noncure to death	-0.09 (0.33)	0.12 (0.34)	0.62 (0.15)	-
Recur to death	0.27 (0.11)	0.39 (0.14)	0.04 (0.04)	-0.12 (0.04)

Parameter Estimates

Table: Parameter estimates (SD) for Trial 3

	Treatment	Stage	Age (10 yrs)	T_r
P(Cure)	0.55 (0.14)	-1.04 (0.16)	0.01 (0.06)	-
Time to recur	-0.11 (0.12)	0.39 (0.16)	-0.10 (0.05)	-
Cure to death	0.05 (0.20)	-0.08 (0.19)	0.75 (0.16)	-
Noncure to death	0.71 (0.54)	-0.33 (0.53)	0.65 (0.30)	-
Recur to death	0.27 (0.11)	0.39 (0.14)	0.07 (0.04)	-0.11 (0.04)

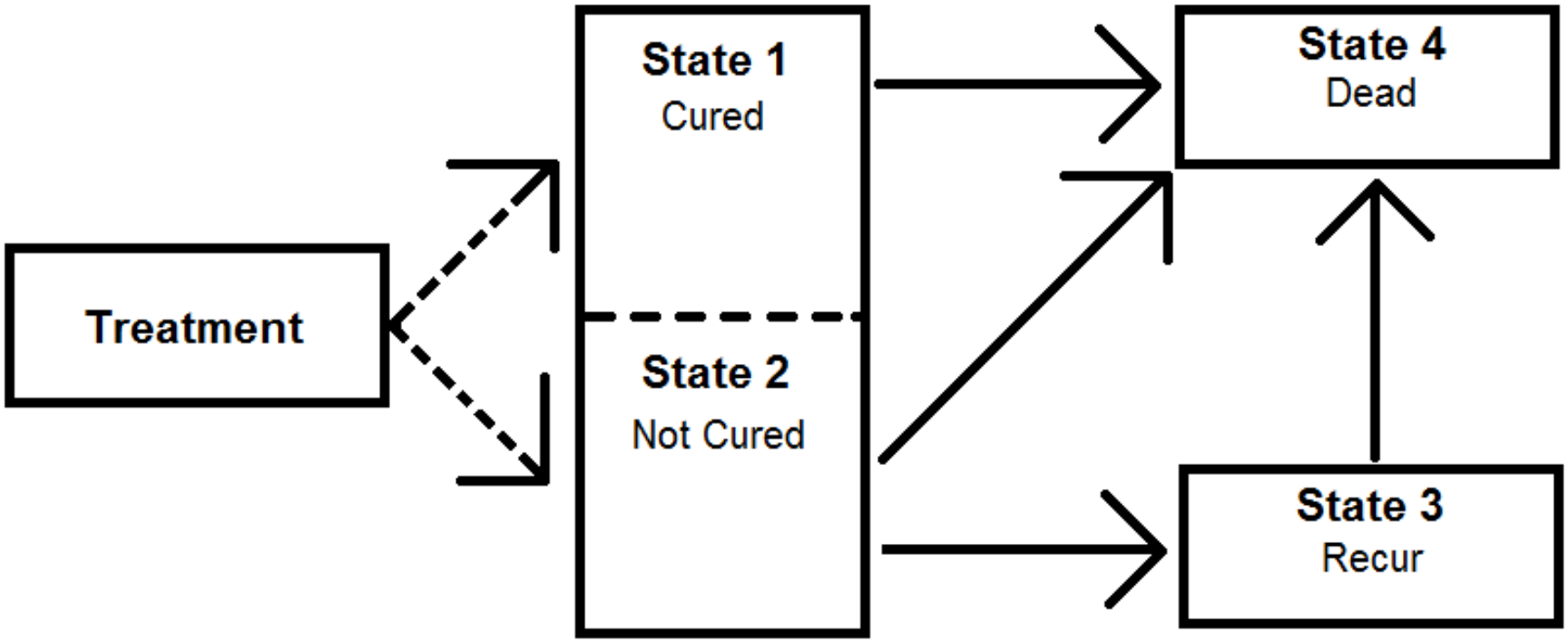
Graph of the multistate model



Parameter estimates follow some expected patterns

- Treatment effects depend on trial
- Stage
  - P(Cure), Very strong effect
  - Recur to death, Strong effect
  - Time to recur, Modest effect
  - Cure-to-death and Noncure-to-death, Smaller and mixed effects
- Age
  - Cure to death, Strong effect
  - Noncure to death, Modest effect
  - Recur to death, Mild effect
  - Time to recur, Mild negative effect
- Quick recurrence associated with short time from recur to death

Graph of the multistate model



Model without cured fraction

- Could we have fit a simpler model?
- Multistate with cured fraction, 25 parameters
- Multistate without cured fraction, 16 parameters

Model comparison by DIC values

# of parameters	Complete Model 25	No Cured Fraction 16
	DIC	
Trial 1	**1093.7**	1130.8
Trial 2	**1789.4**	2106.3
Trial 3	**4206.5**	4375.7
Trial 4	**4392.7**	4507.4
Trial 5	**3572.5**	3694.5
Trial 6	**3607.8**	3740.2
Trial 7	**3176.4**	3301.0
Trial 8	**4084.6**	4303.0
Trial 9	**7677.2**	8083.2
Trial 10	**7706.4**	8061.3
Trial 11	**5225.7**	5407.6
Trial 12	**7172.9**	7349.0

Treatment effects: Overall survival results, log rank tests

- Trials 3,8,9 showed strong treatment effects on overall survival
- Trials 2,7,12 showed marginal treatment effects on overall survival

Application of the model: Treatment effect on overall survival at 5 years

- Have expression for $S(5|X_i, \theta) = P(T_{id} > 5|X_i, \theta)$
- Average over stage and age values
- Calculate $S(5|trt) - S(5|control)$
- Compare point estimates and SE's to 5 year Kaplan Meier estimates

Formula for overall survival distribution

$$
\begin{aligned}
& S(5|X_i, \theta) \\
& = p_i P(\text{Don't die when cured}) \\
& + (1 - p_i) P(\text{Don't recur or die when not cured}) \\
& + (1 - p_i) P(\text{Recur but don't die})
\end{aligned}
$$

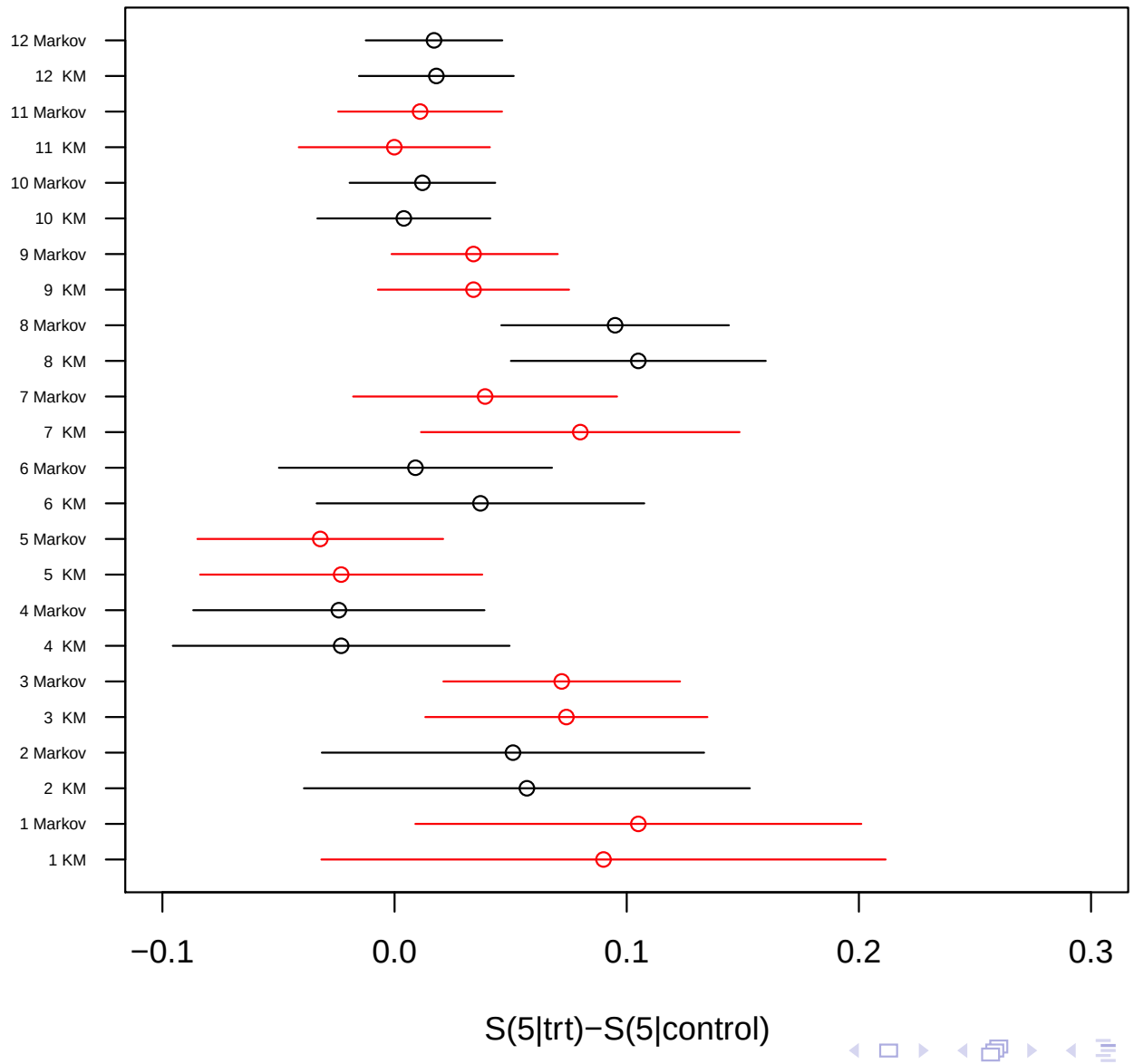
$$
\begin{aligned}
& S(5|X_i, \theta) = \\
& p_i exp\left(-\int_0^5 \lambda_{14}(u)du\right) + \\
& (1 - p_i) exp\left(-\int_0^5 \lambda_{23}(u)du - \int_0^5 \lambda_{24}(u)du\right) + \\
& (1 - p_i) \int_0^5 exp\left(-\int_0^u \lambda_{23}(v)dv - \int_0^u \lambda_{24}(v)dv\right) \lambda_{23}(u) exp\left(-\int_0^{5-u} \lambda_{34}(v)dv\right) du
\end{aligned}
$$

5 Year Treatment Effect Estimates

Table: Results for two trials

	$S(5\|Trt) - S(5\|Cntl)$
Trial 3	
Kaplan-Meier	0.074 (0.031)
Multistate-Cure model	0.072 (0.026)
Trial 5	
Kaplan-Meier	-0.023 (0.031)
Multistate-Cure model	-0.032 (0.027)

Treatment effect on 5 year survival



Auxiliary variable goals

- Utilize the information in recurrence times
- Gain efficiency in the treatment effect estimate on the primary endpoint of interest, overall survival
- Could be used to shorten the length of a clinical trial

Auxiliary variable idea

- Someone who has recurred prior to censoring is likely to die sooner than someone who has not recurred
- Use the model to impute death times for censored people
- Analyse the combination of observed and imputed death times
- Do multiple imputation

Auxiliary variable implementation

- Impute death time from residual time to death distribution
- Impute from $P(T_{id} > Y_{id} + a | T_{id} > Y_{id}, \delta_{id} = 0, Y_{ir}, \delta_{ir}, X_i)$
- Analyze time to death in the combined observed and imputed data
- log-rank tests, Cox models, estimate S(5|Trt)-S(5|Cntl)
- Imputation approach uses the multistate model in a mild way

Auxiliary variable implementation

- Impute from $P(T_{id} > Y_{id} + a | T_{id} > Y_{id}, \delta_{id} = 0, Y_{ir}, \delta_{ir}, X_i)$
- Have expression for $P(T_{id} > Y_{id} + a | T_{id} > Y_{id}, \delta_{id} = 0, Y_{ir}, \delta_{ir}, X_i, \theta)$
- Example when $\delta_{ir} = 1$

$$P(T_{id} > Y_{id} + a | T_{id} > Y_{id}, \delta_{id} = 0, Y_{ir}, \delta_{ir} = 1, X_i, \theta)$$
$$= \exp\left(-\int_{Y_{id}-Y_{ir}}^{Y_{id}+a-Y_{ir}} \lambda_{34}(u)du\right)$$

Auxiliary variable implementation

- Multiple imputation
  - Draw θ from the posterior distribution
  - Impute values of T_{id}
  - Censor at the maximum follow up
  - Analyse the M imputed datasets separately
  - Use the multiple imputation combining rules to get final result from M separate analyses

Could the trial have been shorter?

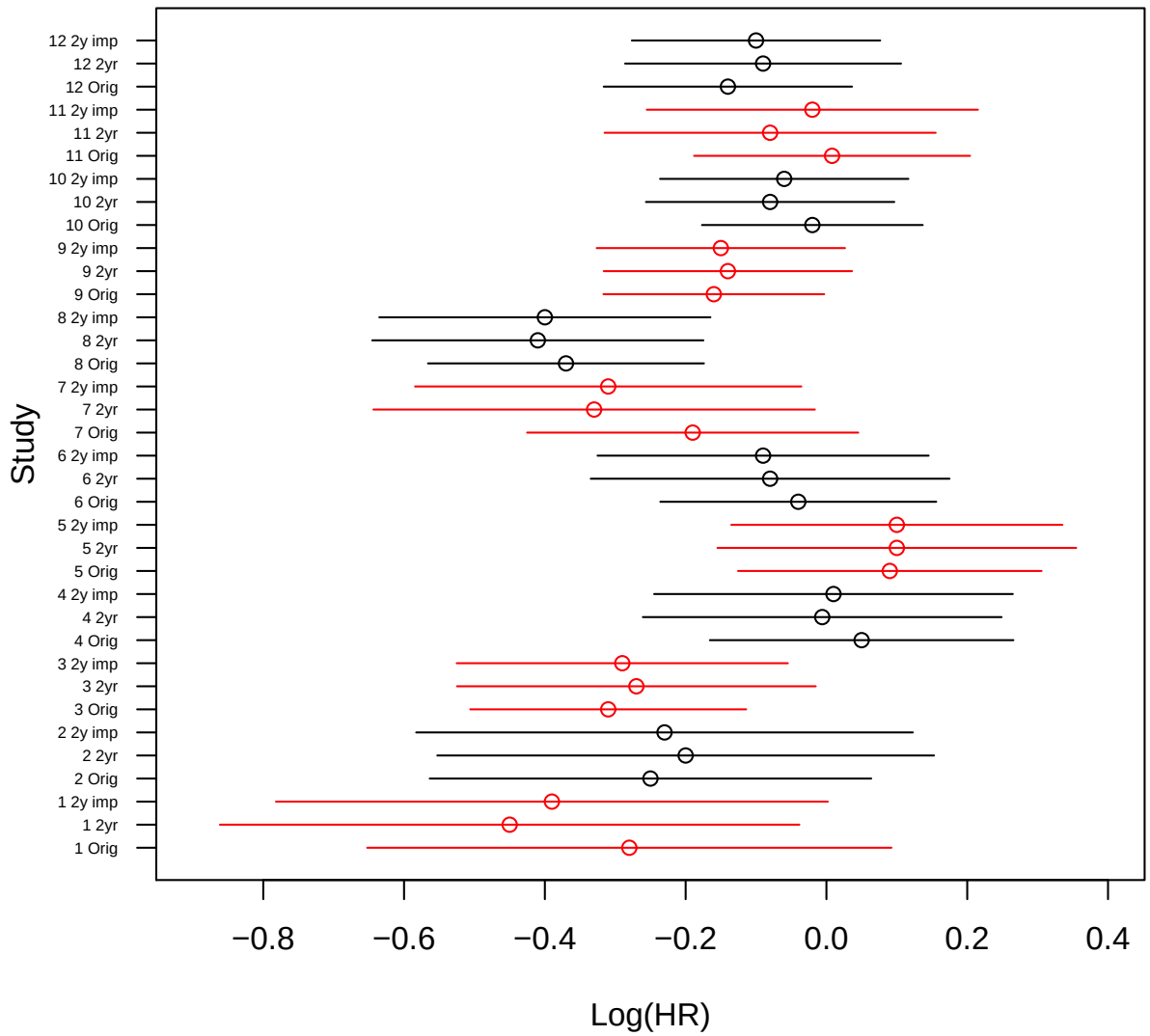
- Reduce the follow-up in the original data by censoring at earlier date
  - Stop trial 2 years after last accrual or 5.5 years follow-up
- Fit multi-state model and apply imputation approach to the reduced follow-up data
- Compare results from original data, reduced follow-up data and imputed data
- Is lost information recovered?

Treatment Effect Results

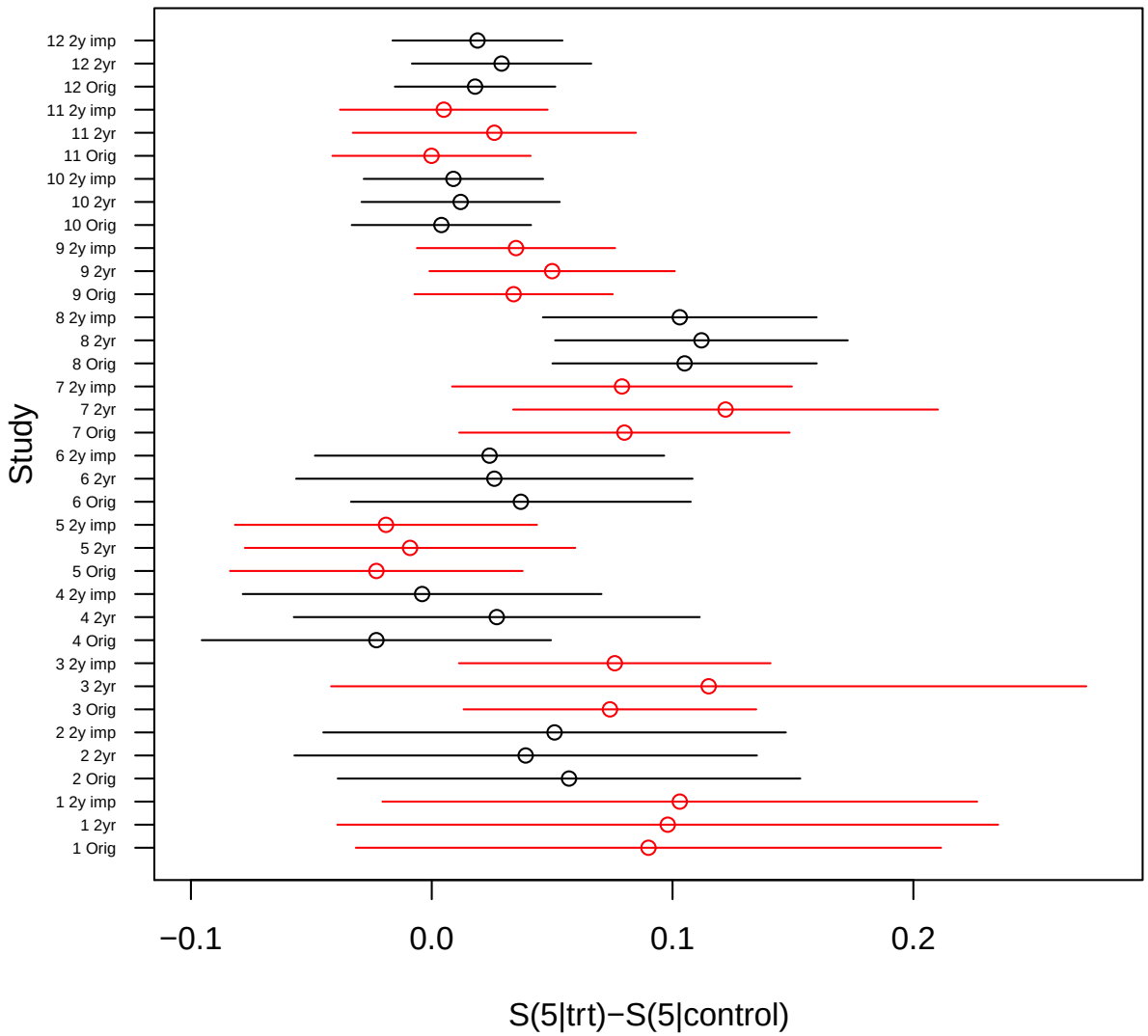
Table: Results for two trials

	log-rank (p-value)	log(HR) (adjusted)	$S(5\|Trt) - S(5\|Cntl)$
Trial 3			
Original	0.002	-0.31 (0.10)	0.074 (0.031)
Reduced FU	0.05	-0.27 (0.13)	0.115 (0.080)
Reduced FU, Imputed	0.04	-0.26 (0.12)	0.076 (0.033)
Trial 5			
Original	0.35	0.09 (0.11)	-0.023 (0.031)
Reduced FU	0.46	0.10 (0.13)	-0.009 (0.035)
Reduced FU, Imputed	0.37	0.11 (0.13)	-0.020 (0.032)

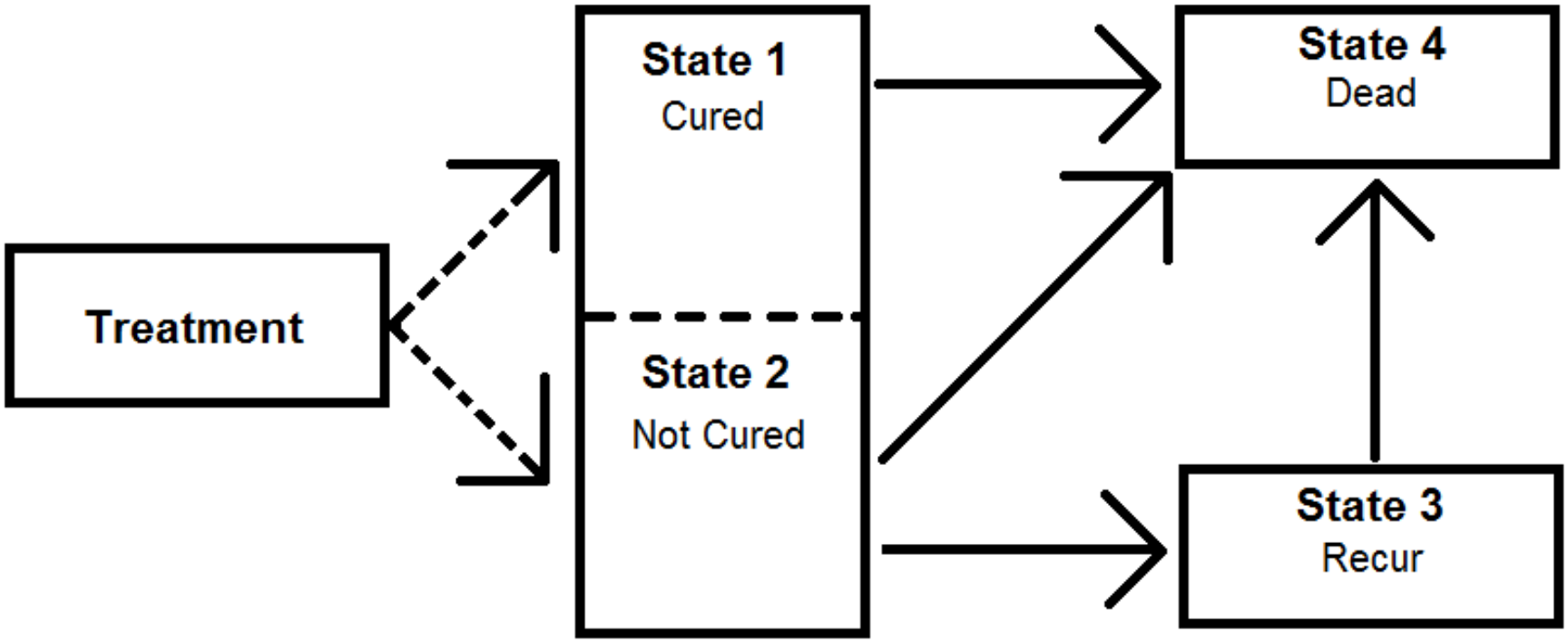
Efficiency comparisons: Hazard ratio



Efficiency comparisons: 5 year survival



Graph of the multistate model



Other ways to gain efficiency

- Impose structure or constraints
- Gains in efficiency may be greater if
  - Shrink treatment coefficient for Recur to Death towards zero
  - Shrink treatment coefficient for Cure to Death towards zero
  - Shrink treatment coefficient for Noncure to Death towards zero
- Li et al, (2011, Biometrics): Cook and Lawless, (2001, J Stat. Plan. Inference); Broglio and Berry, (2009, JNCI)

Efficiency gain from restricting parameters

Table: Treatment effect on survival, original data, reduced FU data and reduced FU data with restrictions and imputation

Study	Data	Log-Rank P-Value	Cox model Log Hazard Ratio (SE)	5 year KM Estimate (SE)
3	Original	0.002	-0.31 (0.10)	0.074 (0.031)
	Reduced FU	0.05	-0.27 (0.13)	0.115 (0.080)
	Reduced FU, Imputed	0.04	-0.26 (0.12)	0.076 (0.033)
	Reduced FU, Imputed, $\beta_{14} = \beta_{24} = \beta_{34} = 0$	0.005	-0.33 (0.12)	0.092 (0.033)
	Reduced FU, Imputed, $\beta_{14}, \beta_{24}, \beta_{34}$ shrunk	0.01	-0.31 (0.12)	0.082 (0.033)
5	Original	0.35	0.09 (0.11)	-0.023 (0.031)
	Reduced FU,	0.46	0.10 (0.13)	-0.009 (0.035)
	Reduced FU, Imputed	0.37	0.11 (0.13)	-0.020 (0.032)
	Reduced FU, Imputed, $\beta_{14} = \beta_{24} = \beta_{34} = 0$	0.46	0.09 (0.12)	-0.017 (0.032)
	Reduced FU, Imputed, $\beta_{14}, \beta_{24}, \beta_{34}$ shrunk	0.41	0.10 (0.12)	-0.019 (0.032)

Discussion

- Sources of gain in efficiency of treatment effect on time-to-death
  - The recurrence time information
  - The assumptions in the model
  - Data from other trials

Discussion

- The comprehensive model allows calculation of any conditional distribution of interest
- Is it worth building a comprehensive model if prediction is the goal?
- Landmark approach: for each landmark time t directly model residual time distribution for time to death amongst those at risk
- $S(a;\theta_t) = P(T_{id} > t + a | T_{id} > t, \delta_{id} = 0, Y_{ir}, \delta_{ir}, X_i, \theta_t)$
- Standard survival analysis with time-independent covariates for each t
- Assume θ_t is a smooth function of t

Discussion

- Landmark approach
- Less assumptions
- Less ability to incorporate scientific context
  - Cured group
  - Age effects
- Can it handle awkward data issues?
  - recurrence time censored before t
  - interval censoring

Acknowledgements

- Anna Conlon, PhD student, University of Michigan
- Dan Sargent, Mayo Clinic
- National Cancer Institute grant funding

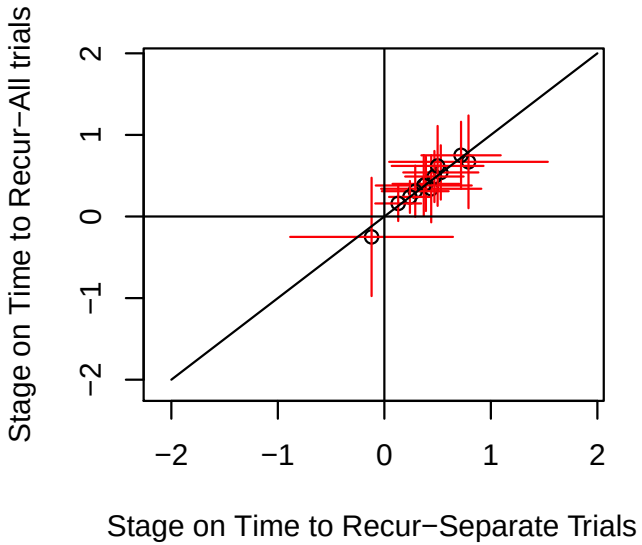
Extension to multiple trials

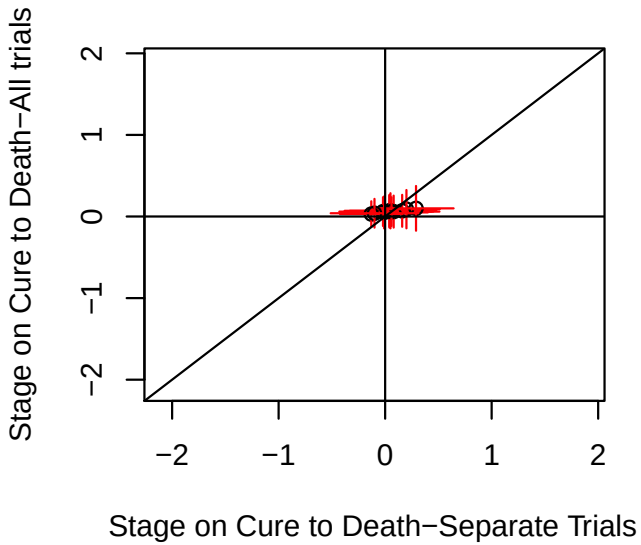
- Borrow information from other trials
- Hierarchical models, Bayesian estimation
- Shrink selected parameters towards common values, age, stage
- Shrink Weibull shape parameter towards common value

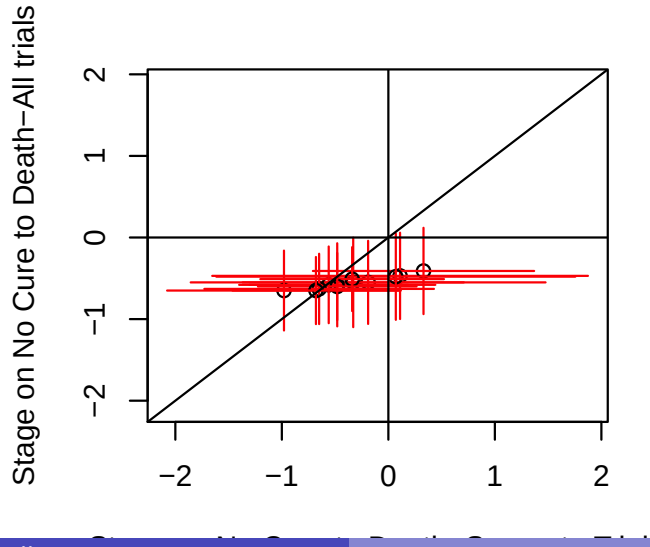
Details of the hierarchical models

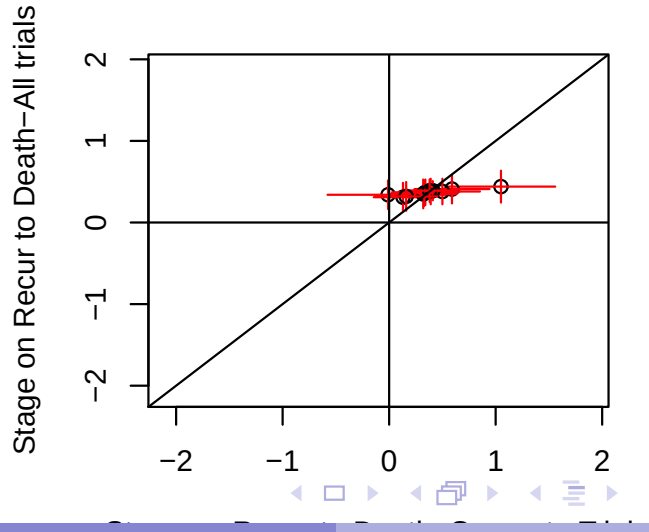
- subject i in study s
- Random effects logistic model for $p_{is} = \mathrm{P}(\mathrm{Cure}|X_{is})$, $logit(p_{is}) = \gamma_s X_{is}$
  - $\gamma_s^{(age)} \sim N(\gamma^{(age)}, \sigma^{(\gamma age)2})$
  - $\gamma_s^{(stage)} \sim N(\gamma^{(stage)}, \sigma^{(\gamma stage)2})$
- Random effects Weibull hazard models
  $\lambda_{kjs}(T_{is}; X_{is}) = \left(\frac{\rho_{kjs}}{\lambda_{kj}}\right)\left(\frac{T_{is}}{\lambda_{kj}}\right)^{\rho_{kjs}-1} exp\left(\beta_{kjs} X_{is}\right)$
  - $\beta_{kjs}^{(age)} \sim N(\beta_{kj}^{(age)}, \sigma_{kj}^{(\beta age)2})$
  - $\beta_{kjs}^{(stage)} \sim N(\beta_{kj}^{(stage)}, \sigma_{kj}^{(\beta stage)2})$
  - $\beta_{34s}^{(recur)} \sim N(\beta_{34}^{(recur)}, \sigma_{34}^{(\beta recur)2})$
  - $\rho_{kjs} \sim N(\rho_{kj}, \sigma_{kj}^{(\rho)2})$
- 150 population parameters $(= 12 \times (2 + 4 \times 2) + 2 \times (3 + 4 \times 3))$

Stage effect

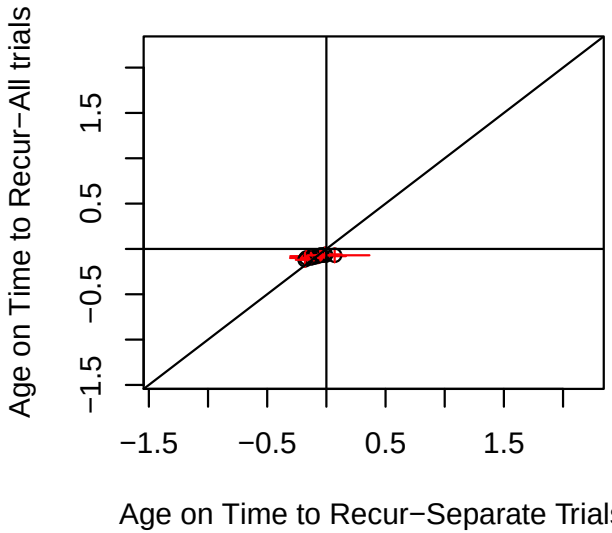


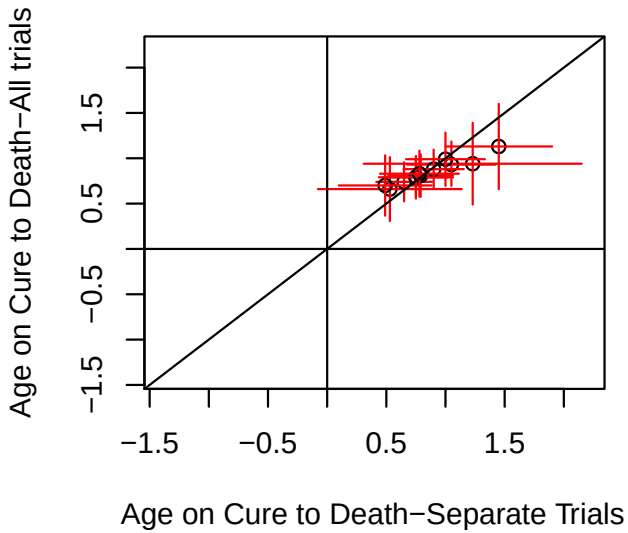


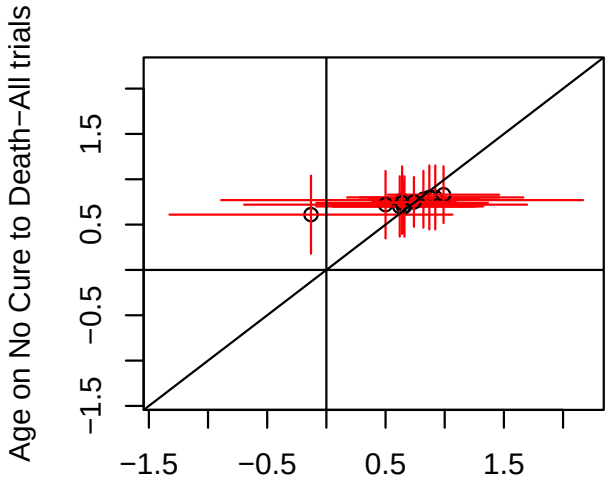


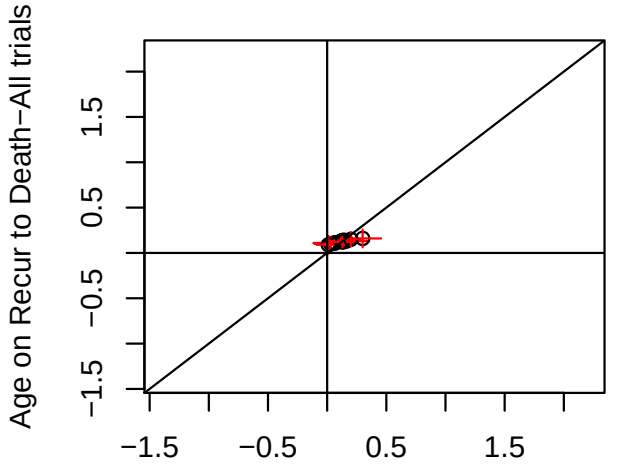


Age effect


Age on Time to Recur−All trials
Age on Time to Recur−Separate Trials
−1.5
−0.5
0.5
1.5


Age on Cure to Death−All trials
Age on Cure to Death−Separate Trials
−1.5
−0.5
0.5
1.5


Age on No Cure to Death−All trials
−1.5
−0.5
0.5
1.5


Age on Recur to Death−All trials
−1.5
−0.5
0.5
1.5

Treatment effect