Introduction to Biostatistics

Mithat Gonen

Brenden Bready

Today's Plan

- Review
  - 2x2 table
  - Censoring, Kaplan-Meier estimator and the log-rank test
  - Hazard Function and Hazard Ratio
- New Material
  - RMST
  - Competing Risks
  - Immortal Time Bias
  - Left Truncation

2x2 Table

	Response	No Response	TOTAL
Treatment A	R_1	$N_1 - R_1$	N_1
Treatment B	R_2	$N_2 - R_2$	N_2
TOTAL	$R_1 + R_2$	$N_1 - R_1 + N_2 - R_2$	$N_1 + N_2$

- Think of a 2x2 table anytime you have a binary outcome and 2 groups
- Think of a KxP table any time you have an outcome with K categories and P groups
- All the information in the data we need is here
- Very useful and flexible display

Difference, Relative Risk, Odds Ratio

- Proportions: R1/N1 = p1 and R2/N2 = p2
- Three metrics for summarizing the difference
  - Difference: p1 – p2
  - Relative Risk: p1/p2
  - Odds Ratio: [p1/(1-p1)] / [p2/(1-p2)]
- Important note: Odds Ratio is often interpreted as a relative risk (e.g., 2-fold increase in risk) but it is not

Summary of the 2x2 Table

- Three metrics: Difference, Relative Risk, Odds Ratio
- Two test statistics
  - Chi-Square test
  - Fisher’s exact test

Survival Analysis

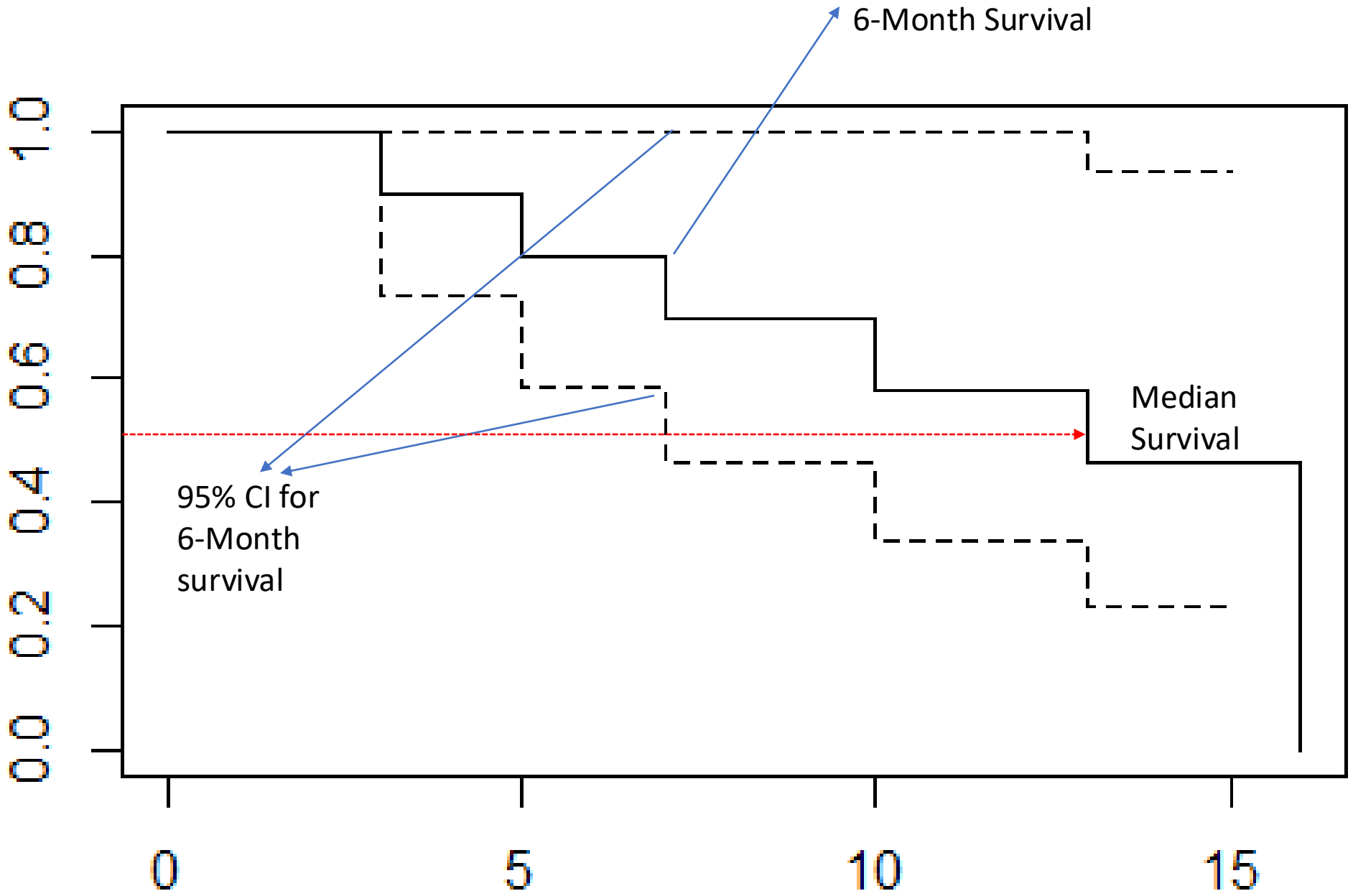
- Most important cancer outcomes are time-to-event
  - Time to death (overall survival)
  - Time to cancer death (disease specific survival)
  - Time to progression (progression free survival)
- Time interval between a “baseline” and an event
- Continuous variable?
- What is the big deal?

Censoring

- You are VERY VERY unlikely to observe all patients have the event of interest
- Some people will be alive or free of progression when the data analysis time comes
- This is called censoring, a data point that is not (yet) fully observed
  - Unfortunate terminology
- Primary reason why survival analysis requires special treatment

What Can We Do?

- Ignore censoring, analyze as continuous
  - Treats people under follow-up as if they died
  - Underestimates survival probabilities
- Ignore time, analyze binary
  - Someone who died at 10 years vs someone who is alive and under follow-up at 6 months
- Both wrong, should NEVER EVER be used


6-Month Survival
1.0
0.8
0.6
0.4
0.2
0.0
Median
Survival
95% CI for
6-Month
survival
0
5
10
15

Characteristics of KM Curves

- Full information about survival
  - All probabilities
  - All quantiles, quartiles etc
- Jumps down at event times
- Stays the same at censoring times
  - This is why censoring times are indicated by tick marks
- Step sizes get larger and larger
  - Because the denominator gets smaller and smaller

Key Concept

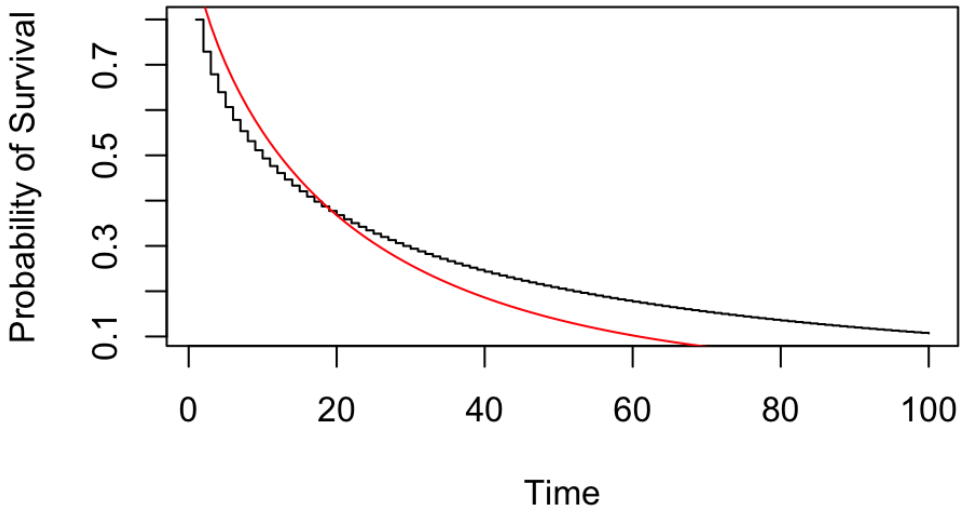
- Risk set, or the number at risk changes at each time point
- This is true even when there is no censoring; those who die move out of the risk set
- But when censoring is present, those who are censored also move out of the risk set at the time of censoring
- The idea of risk set underlies all survival analysis (Cox regression, competing risks, truncation etc)

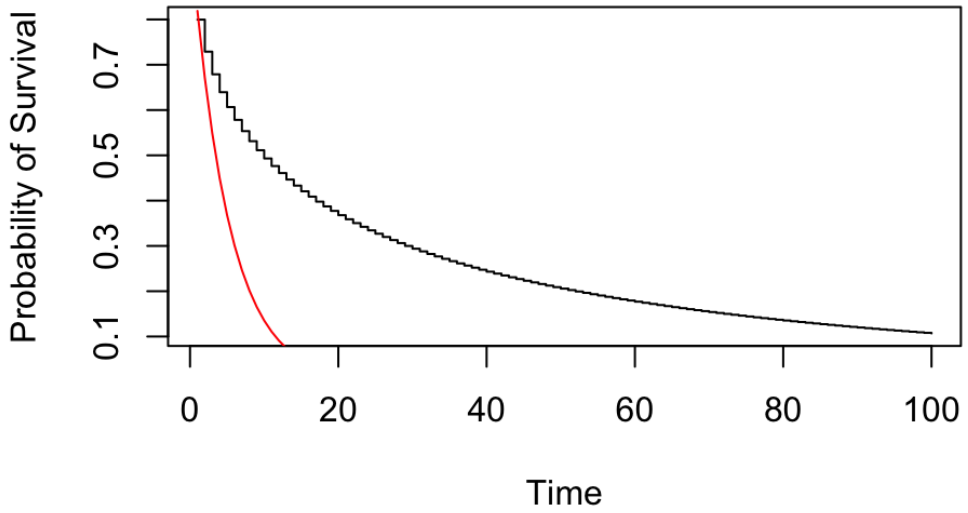
Log-Rank Test

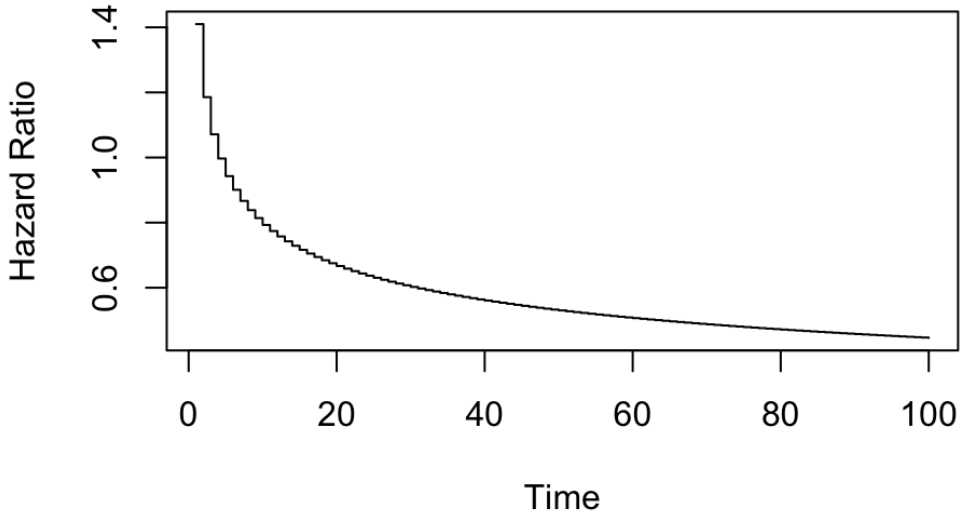
- Standard method of comparing survival curves
- At each time point one of the curves (or both) jumps, calculate the expected number of events at that time point if there is no difference between the curves
  - Why do we do this calculation assuming there is no difference?
- Compare the expected number with the observed (similar to the chi-square test)
- Add up the difference between observed and expected over all time points

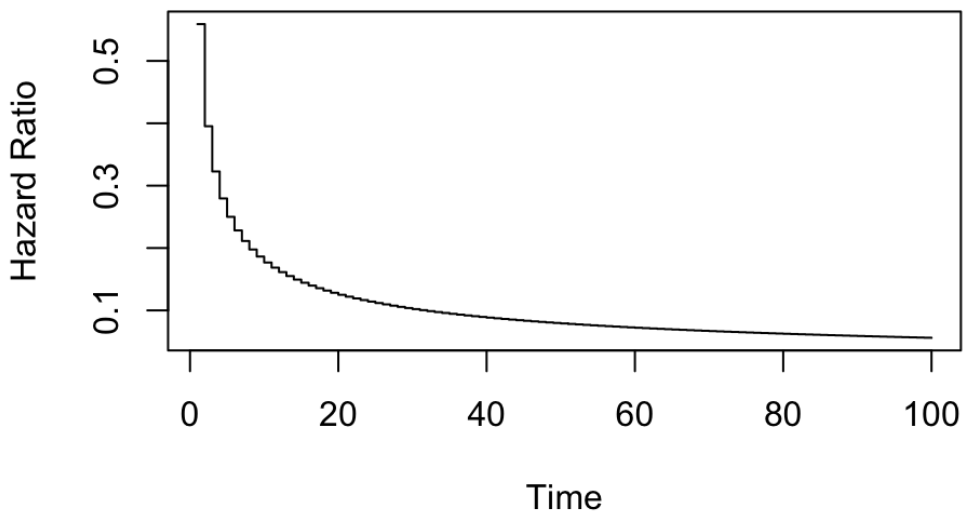
Hazard Function: Definition & Interpretation

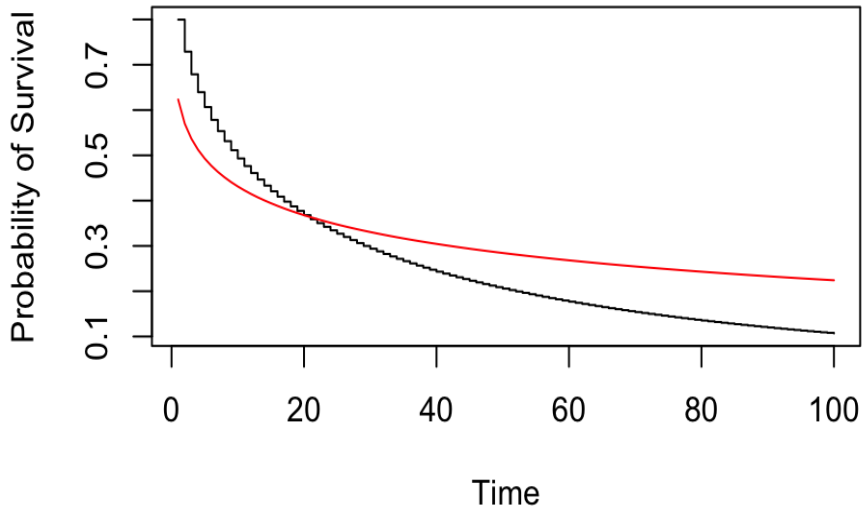
- $h(t) = (d/dt) \log(S(t))$
- Probability of dying between t and t+d, given one has survived up to t, divided by d, where d is a very, very small time interval
- Hazard is NOT probability
- It is probability scaled by a small length of time
- Numerically, not useful at all by itself
- But the ratio of two hazard functions is quite informative

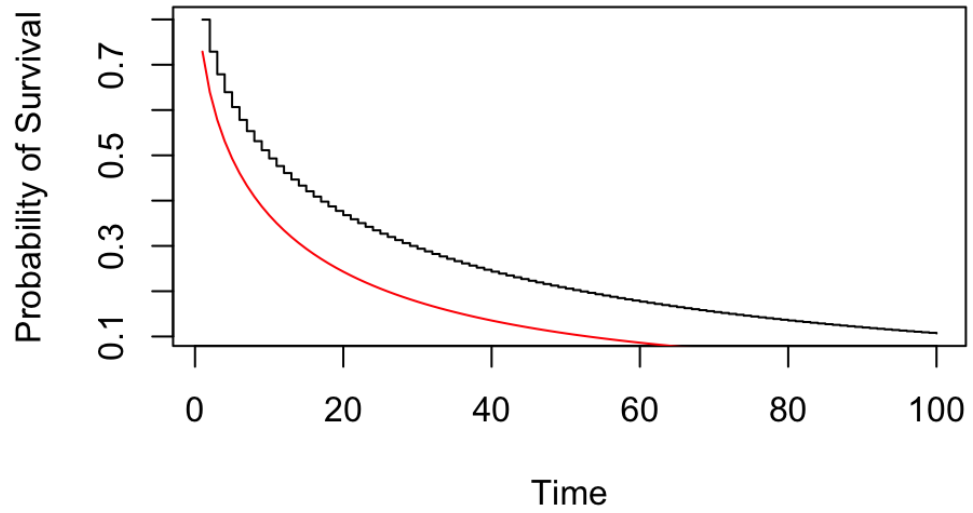

Probability of Survival
0.1
0.3
0.5
0.7
0
20
40
60
80
100
Time

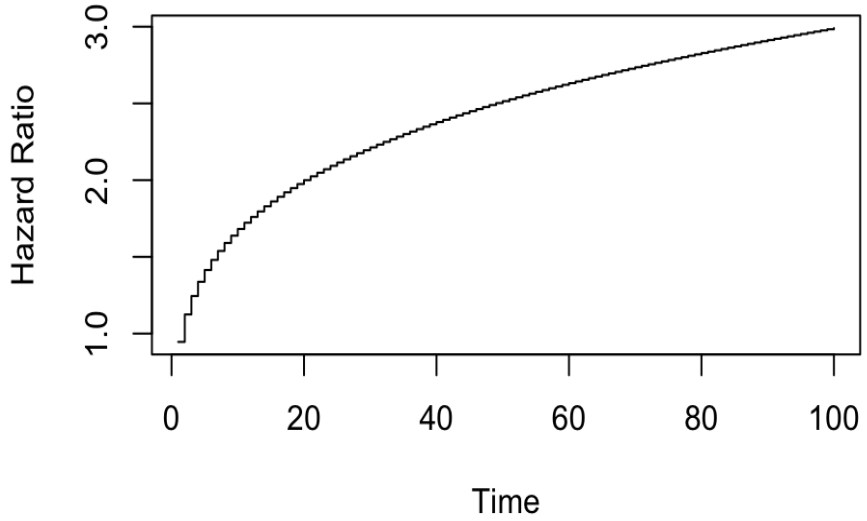

Probability of Survival
0.1
0.3
0.5
0.7
0
20
40
60
80
100
Time

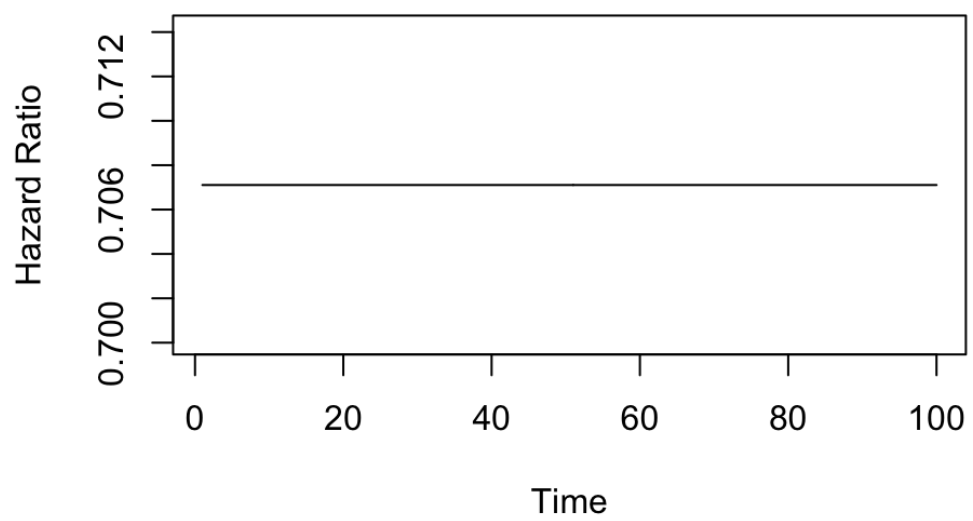

Hazard Ratio
0.6
1.0
1.4
0
20
40
60
80
100
Time


Hazard Ratio
0.1
0.3
0.5
0
20
40
60
80
100
Time


Probability of Survival
0.1
0.3
0.5
0.7
0
20
40
60
80
100
Time


Probability of Survival
0.1
0.3
0.5
0.7
0
20
40
60
80
100
Time


Hazard Ratio
1.0
2.0
3.0
0
20
40
60
80
100
Time


Hazard Ratio
0.700
0.706
0.712
0
20
40
60
80
100
Time

Hazard Ratio

- Hazard function itself is not used very much but ratio of two hazard functions is commonly used to “compare” survival functions
- Constant hazard ratio → Proportional hazards (PH)
- Crossing survival curves → Not PH
- Converse is not true; PH can fail without survival curves crossing
- Eyeballing the survival curves to figure out hazard ratios is futile

The NEW ENGLAND JOURNAL of MEDICINE

ESTABLISHED IN 1812 JULY 17, 2025 VOL. 393 NO. 3

Perioperative Durvalumab in Gastric and Gastroesophageal Junction Cancer

Y.Y. Janjigian,[1] S.-E. Al-Batran,[2] Z.A. Wainberg,[3] K. Muro,[4] D. Molena,[5] E. Van Cutsem,[6] W.J. Hyung,[7] L. Wyrwicz,[8] D.-Y. Oh,[9] T. Omori,[10] M. Moehler,[11] M. Garrido,[12] S.C.S. Oliveira,[13] M. Liberman,[14] V.C. Oliden,[15] E.C. Smyth,[16] A. Stein,[17] M. Bilici,[18] M.L. Alvarenga,[19] V. Kozlov,[20] F. Rivera,[21] A. Kawazoe,[22] O. Serrano,[23] E. Heilbron,[24] A. Negro,[24] J.F. Kurland,[24] and J. Tabernero,[25] for the MATTERHORN Investigators*

METHODS

In a phase 3, multinational, double-blind, randomized trial, we assigned participants with resectable gastric or gastroesophageal junction adenocarcinoma, in a 1:1 ratio, to receive durvalumab at a dose of 1500 mg or placebo every 4 weeks plus FLOT for 4 cycles (2 cycles each of neoadjuvant and adjuvant therapy), followed by durvalumab or placebo every 4 weeks for 10 cycles. The primary end point was event-free survival; secondary end points included overall survival and pathological complete response.

RESULTS

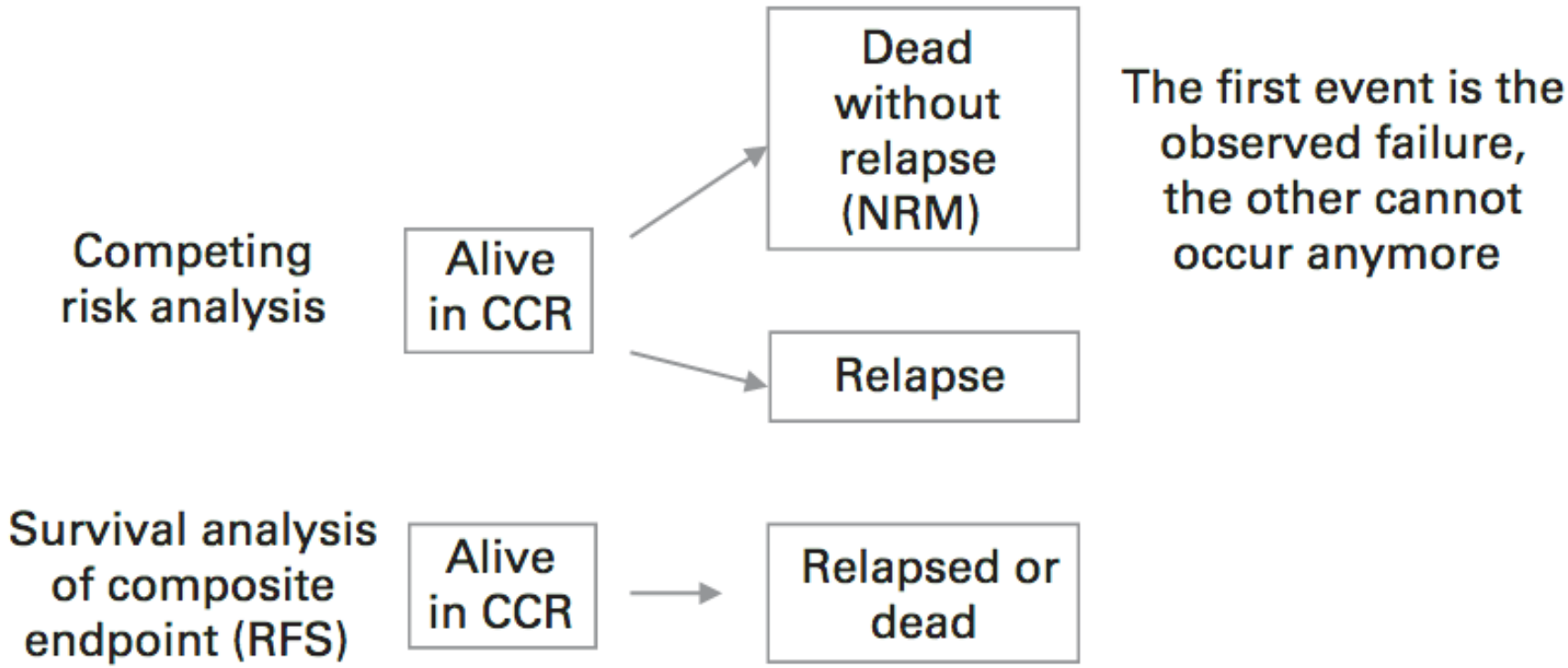
A total of 474 participants were randomly assigned to the durvalumab group, and 474 to the placebo group (median follow-up, 31.5 months; interquartile range, 26.7 to 36.6). Two-year event-free survival (Kaplan–Meier estimate) was 67.4% among the participants in the durvalumab group and 58.5% among those in the placebo group (hazard ratio for event or death, 0.71; 95% confidence interval [CI], 0.58 to 0.86; P<0.001). Two-year overall survival was 75.7% in the durvalumab group and 70.4% in the placebo group (piecewise hazard ratio for death during months 0 to 12, 0.99 [95% CI, 0.70 to 1.39], and during the period from month 12 onward, 0.67 [95% CI, 0.50 to 0.90]; P=0.03 by a stratified log-rank test [exceeding the significance threshold of P<0.0001]). The percentage of participants with a pathological complete response was 19.2% in the durvalumab group and 7.2% in the placebo group (relative risk, 2.69 [95% CI, 1.86 to 3.90]). Adverse events with a maximum grade of 3 or 4 were reported in 340 participants (71.6%) in the durvalumab group and in 334 (71.2%) in the placebo group. The percentage of participants with delayed surgery was 10.1% and 10.8%, respectively, and the percentage with delayed initiation of adjuvant treatment was 2.3% and 4.6%.

Composite Endpoints

- Survival analysis we have learned so far considers one type of event (Death in Overall Survival)
- What if we are analyzing time to progression and we have patients who dies without progression?
- One approach is to include deaths as an event (progression free survival comes from this: one has to be progression free AND surviving to be censored)
- These are called composite endpoints: multiple ways to have an event

Competing Risks

- Composite endpoints can be analyzed using K-M, HR etc since there is not ambiguity on who is an event and who is censored
- What if we want to isolate progression and do not want to count death as an event?
  - Can we think of examples why we may want to do this?
- Then death without progression is a competing event and needs to be handled separately.

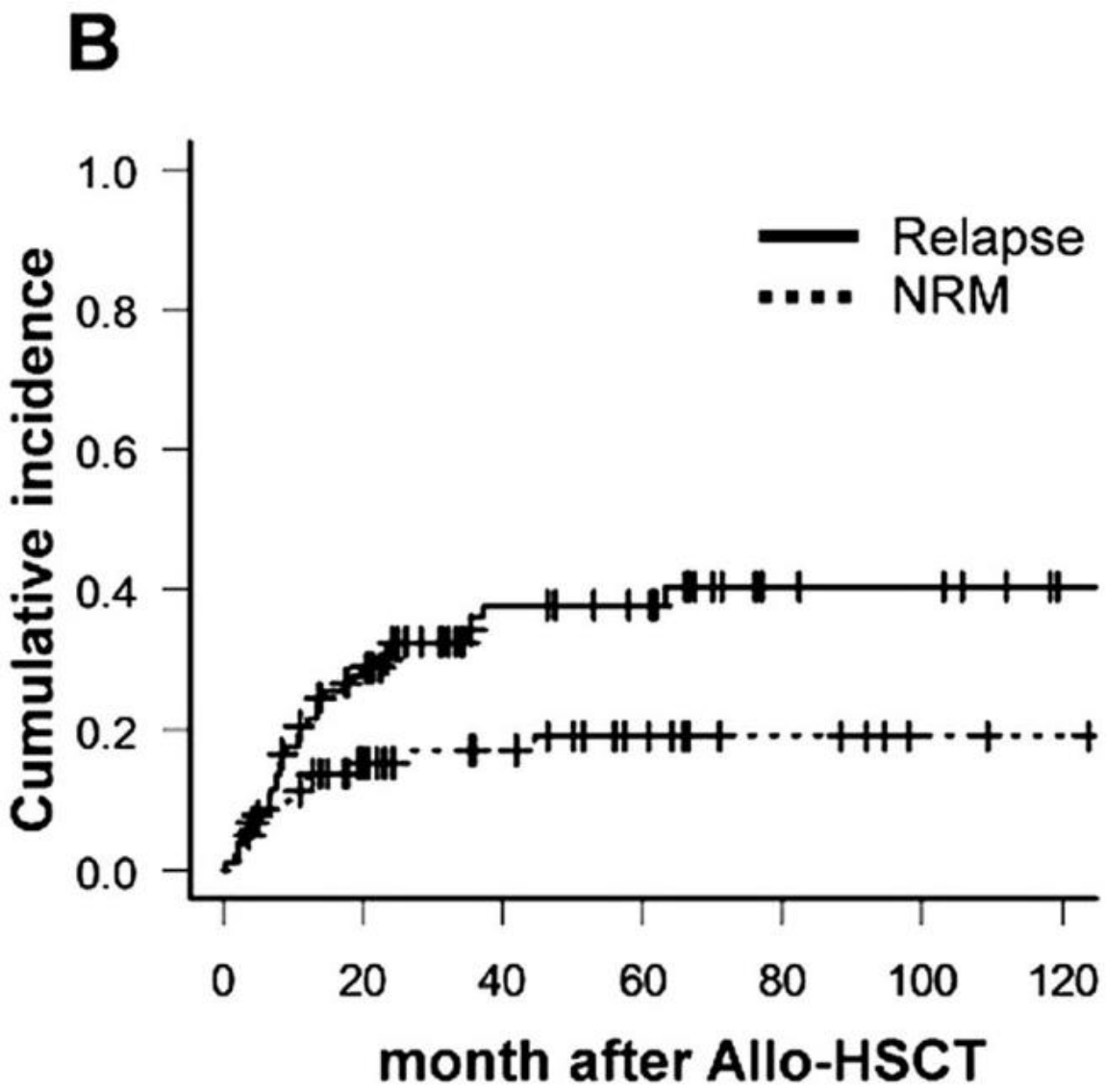

Competing risk analysis
Alive in CCR
Dead without relapse (NRM)
Relapse
The first event is the observed failure, the other cannot occur anymore
Survival analysis of composite endpoint (RFS)
Alive in CCR
Relapsed or dead

Looking under the hood

- 100 patients transplanted
  - 50 relapse within 2 years
  - 10 die between 2 and 5 years
  - No relapse between 2-5 years
- What is the (cumulative) probability of death without relapse at 5 years?
  - 10/100 = 10%; 10 of 100 to begin with
  - 10/50 = 20%; 25 of 50 who did not relapse

Joint versus Conditional Probability

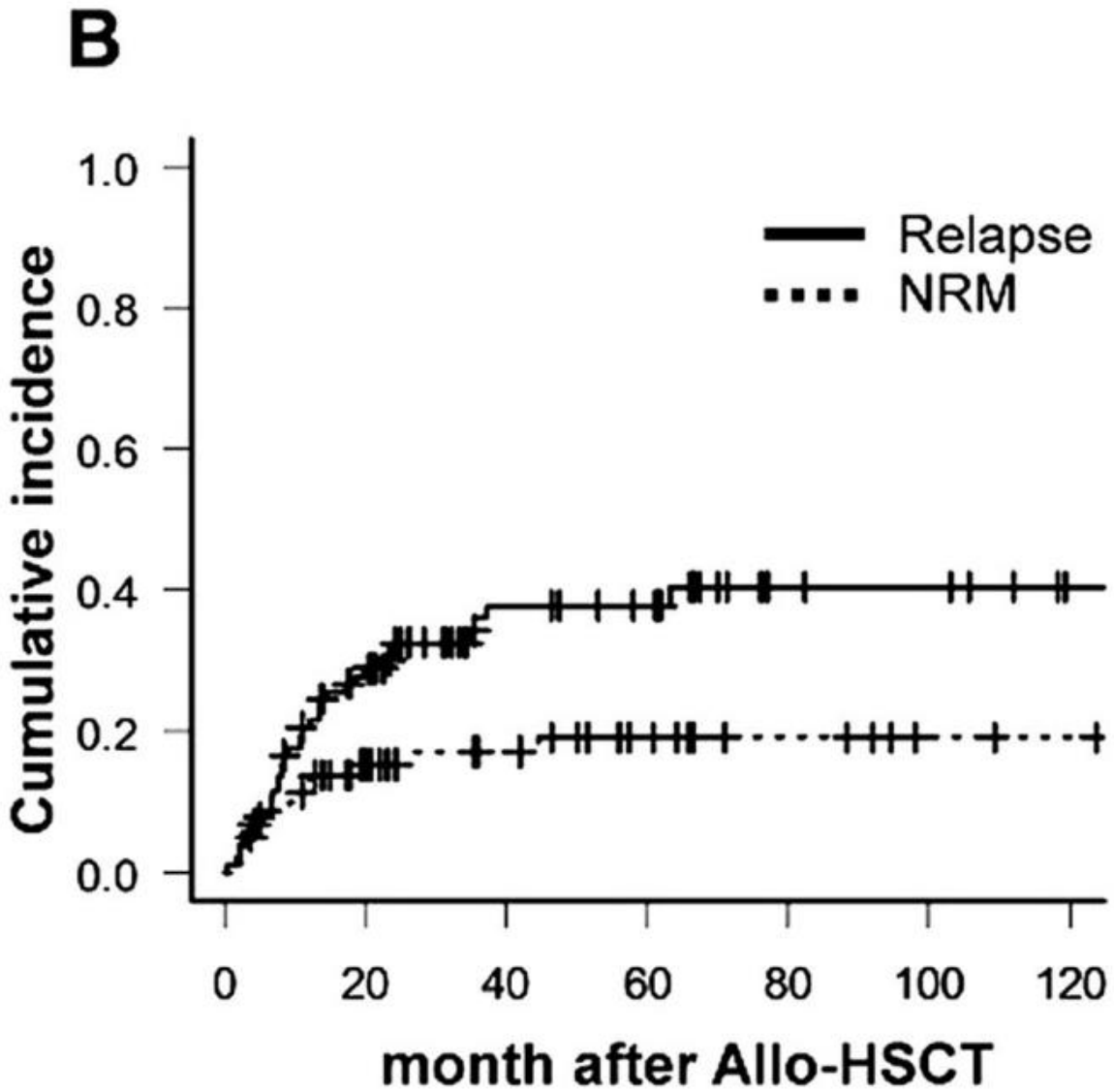
- 10/50 = 20% is the probability of death without relapse CONDITIONAL on surviving the first 5 years (known as cause specific survival)
  - Can be estimated using K-M by censoring competing events
- 10/100 = 10% is the JOINT probability of second malignancy AND not dying before 5 years (known as cumulative incidence)
  - Needs specific methods
- Bayes Rule:
  - Joint = Conditional * Marginal
  - 10/100 = 10/50 * 50/100



Annals of Hematology (2019) 98:1743–1753
https://doi.org/10.1007/s00277-019-03714-x

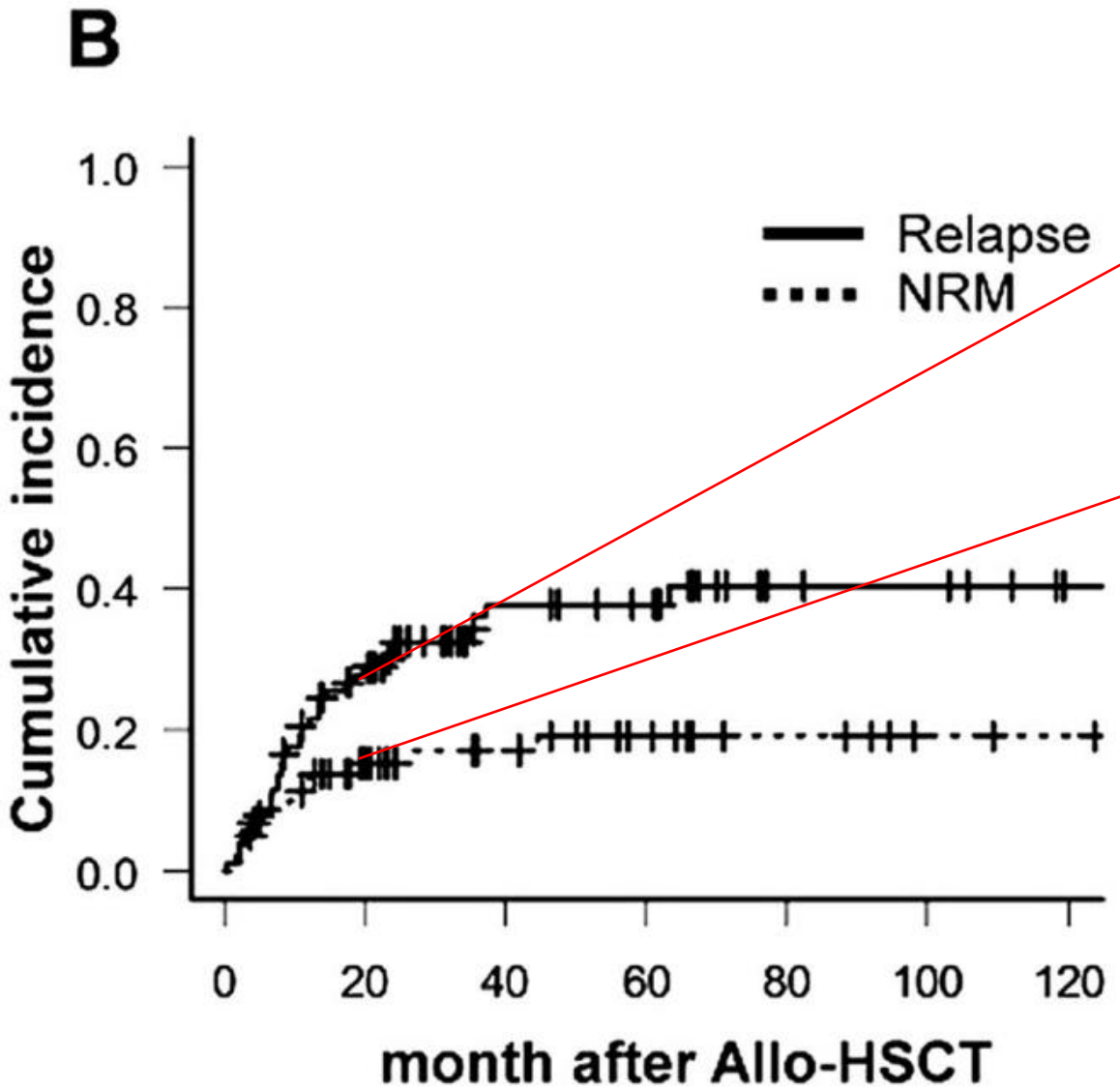
ORIGINAL ARTICLE

Risk factors predicting graft-versus-host disease and relapse-free survival after allogeneic hematopoietic stem cell transplantation in relapsed or refractory non-Hodgkin's lymphoma



What is the 20-month probability of relapse?

What is the 20-month probability of NRM?

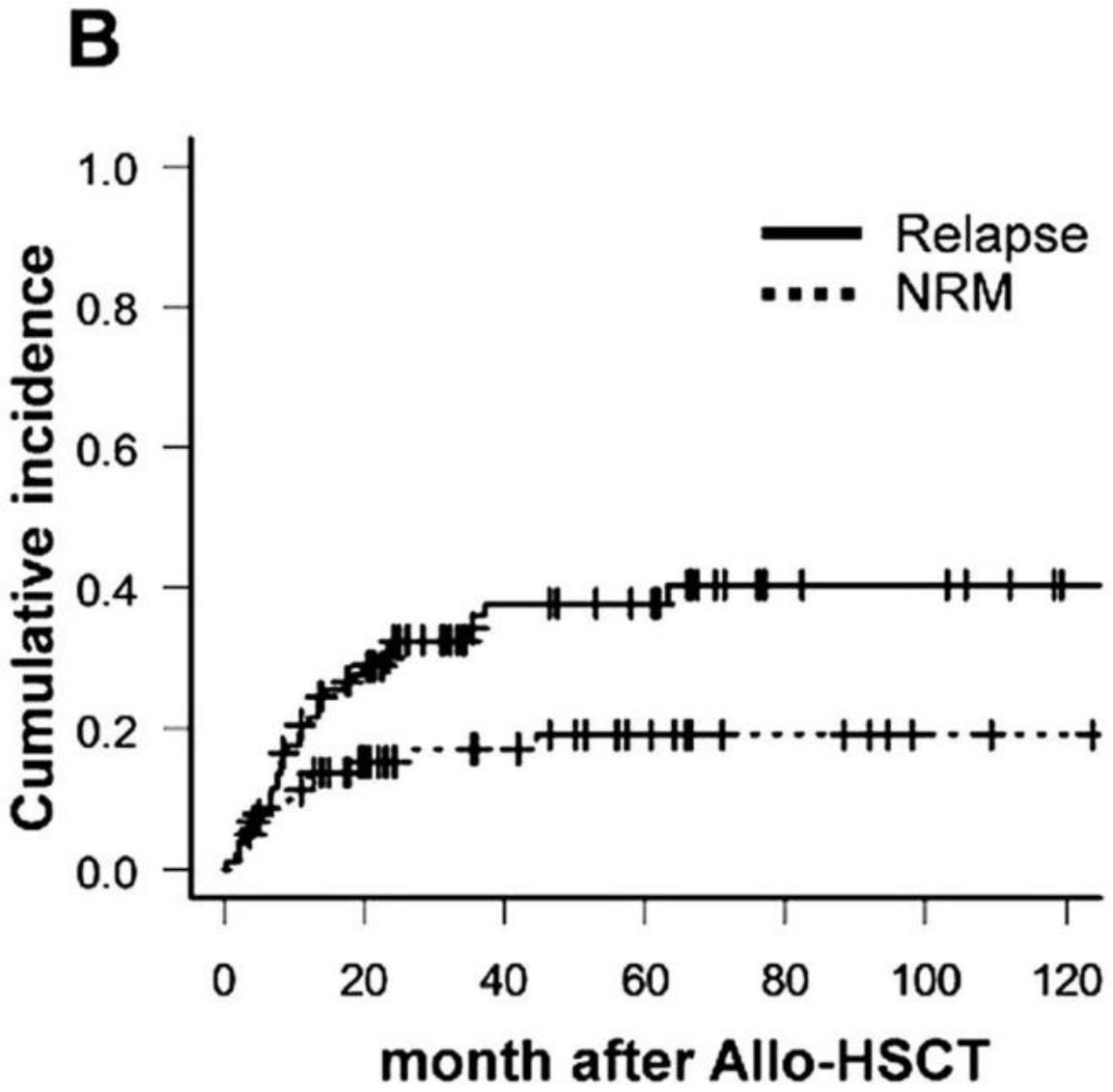


What is the 20-month probability of relapse?

~25%

What is the 20-month probability of NRM?

~15%



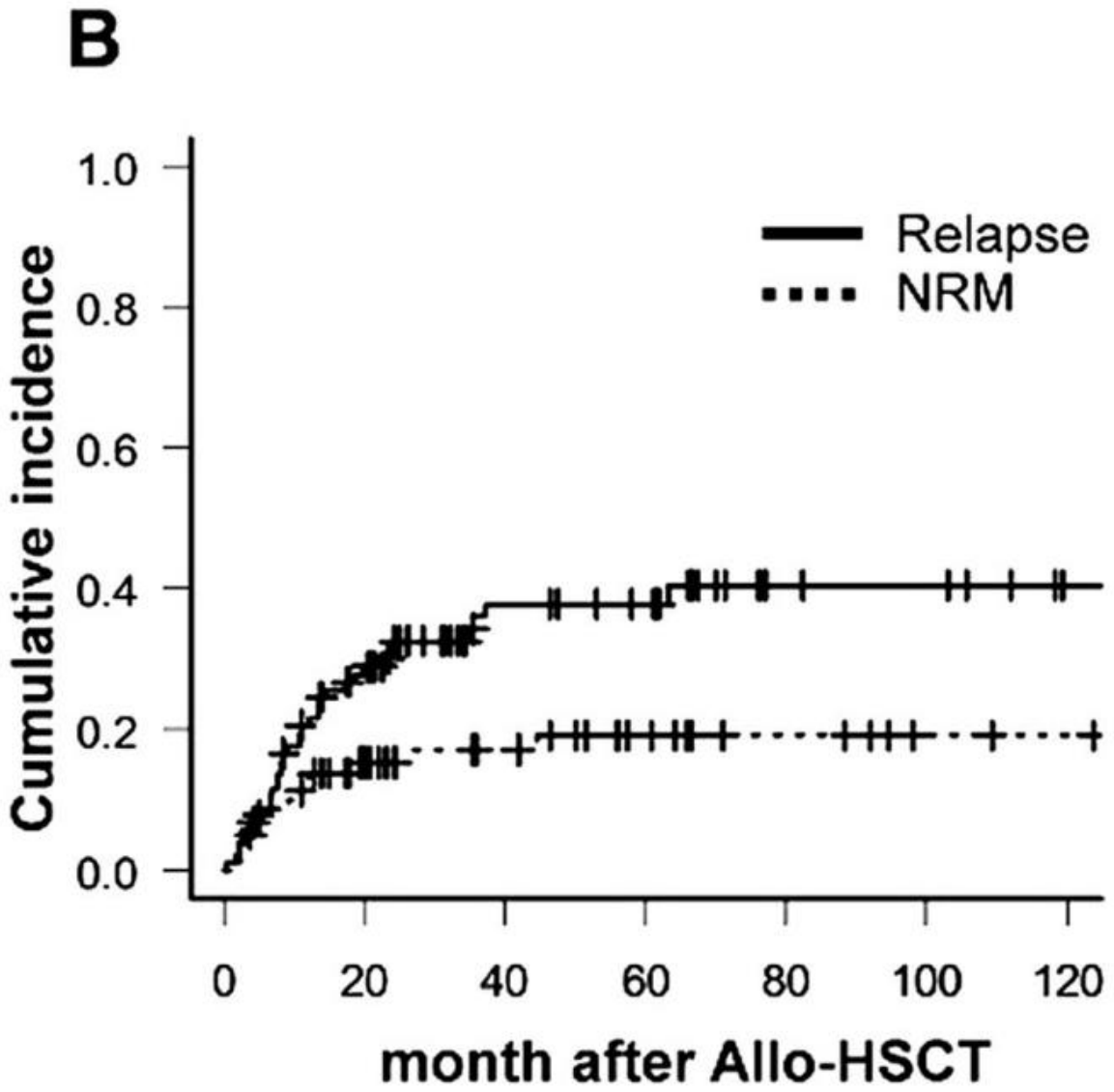
What is the 20-month probability of relapse?

~25%

What is the 20-month probability of NRM?

~15%

What is the probability of neither at 20-months? Do we have a name for this probability?



What is the 20-month probability of relapse?

~25%

What is the 20-month probability of NRM?

~15%

What is the probability of neither at 20-months? Do we have a name for this probability?

60% (1-0.25-0.15)
Relapse-free survival!

Lessons Learned

- One minus all cumulative incidence curves added is the probability of being free of all events
- Consequences
  - Cumulative incidence curves will not reach 1 (they are not real probability distributions, sometimes called sub-distributions)
  - Showing all cumulative incidence curves is important even if you want to focus on only one of them
  - Does the median time to relapse even mean anything? Report probabilities at selected time points.

Multiple Competing Events

- Intrahepatic recurrence
- Extrahepatic recurrence
- Simultaneous recurrence
- Death without recurrence
- Can be analyzed with the same principle

Comparing treatments or risk factors in presence of competing risks

- Fundamental principle: compare like-to-like
- Compare the cumulative incidence of relapse for treatment A vs treatment B
- Do not forget to compare comparing the deaths without relapse, otherwise you could be favoring a very toxic treatment
- You should always look at the comparisons of all cumulative incidence curves
- This is one of the reasons RFS is favored as a clinical trial endpoint

What happened to the cause specific analyses

- They can be useful to understand mechanisms of failure
- But many traps along the way
- Do not perform one or even interpret one without your trusted statistician

Post-Baseline Covariates

- Do responders live longer than non-responders?
- Does relapse change the prognosis?
- Does a local recurrence portend death?
- These questions have two things in common
  - They are all important and need to be answered
  - They all require special methods

Immortal time bias

- We want to analyze a post-baseline covariate such as response to treatment
- But not everyone survives up to the point of response evaluation
- One common strategy is to include them as non-responders. But then by definition all responders are “immortal” until response evaluation
- Probably the most common statistical mistake I see in cancer journals

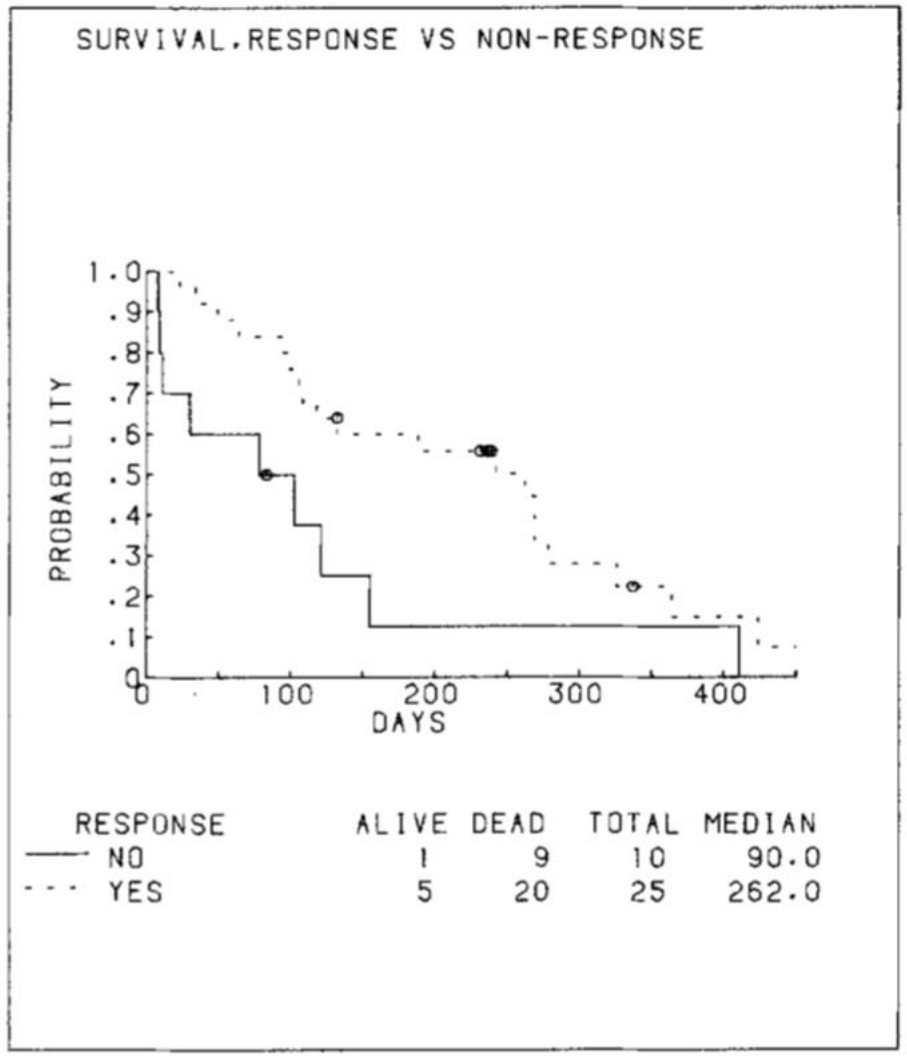


Fig. 2. Evaluation of survival for responders versus nonresponders by the usual method. Data from Eastern Cooperative Oncology Group Study EST 3477: Phase II master protocol for evaluation of agents in patients with multiple myeloma.

Anderson JR, Cain KC, Gelber RD. Analysis of survival by tumor response. J Clin Oncol. 1983 Nov;1(11):710-9. doi: 10.1200/JCO.1983.1.11.710. PMID: 6668489.

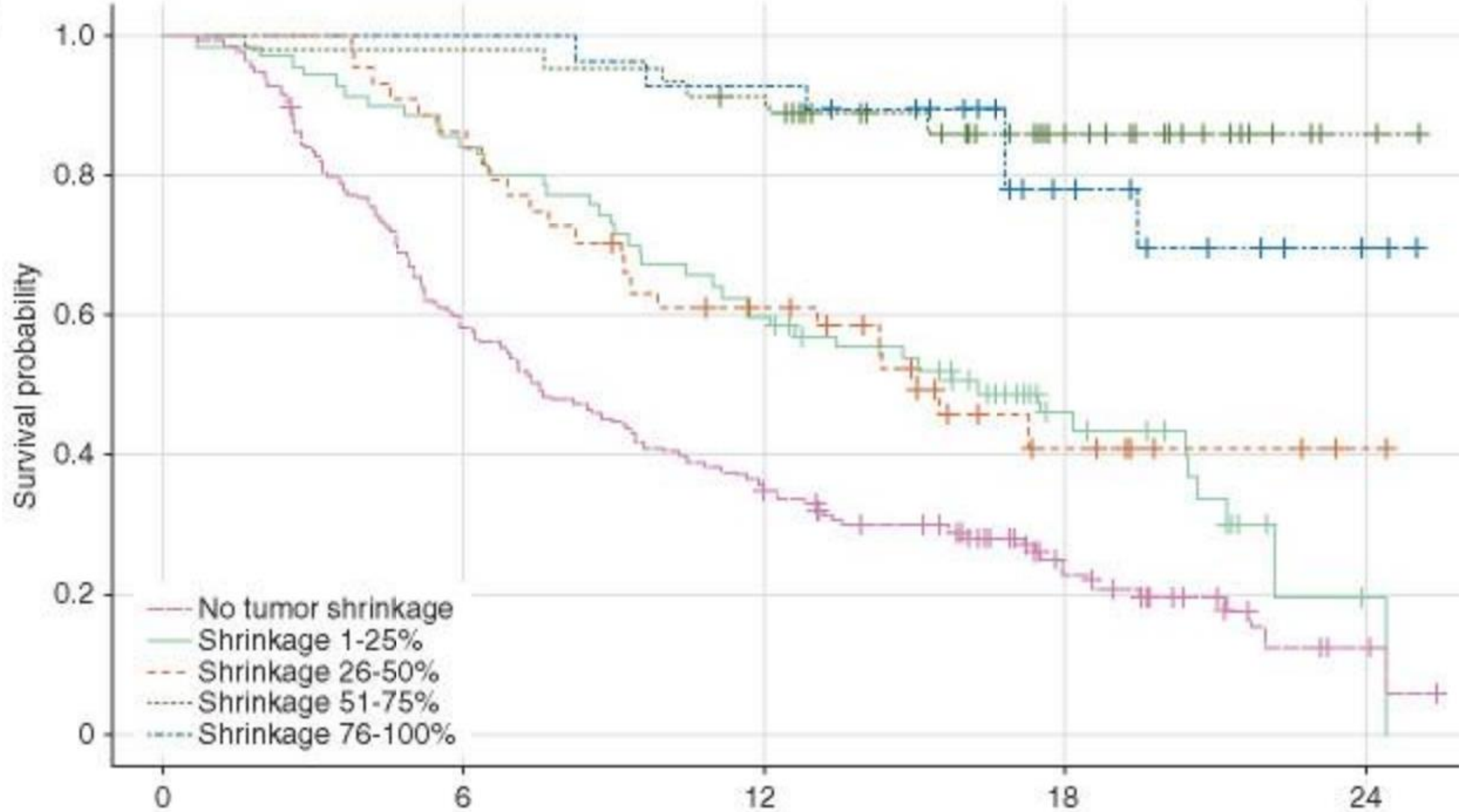
ESMO GOOD SCIENCE BETTER MEDICINE BEST PRACTICE

Annals of Oncology 28: 2707–2714, 2017
doi:10.1093/annonc/mdx414
Published online 2 August 2017

ORIGINAL ARTICLE

Exploratory analysis of the association of depth of response and survival in patients with metastatic non-small-cell lung cancer treated with a targeted therapy or immunotherapy

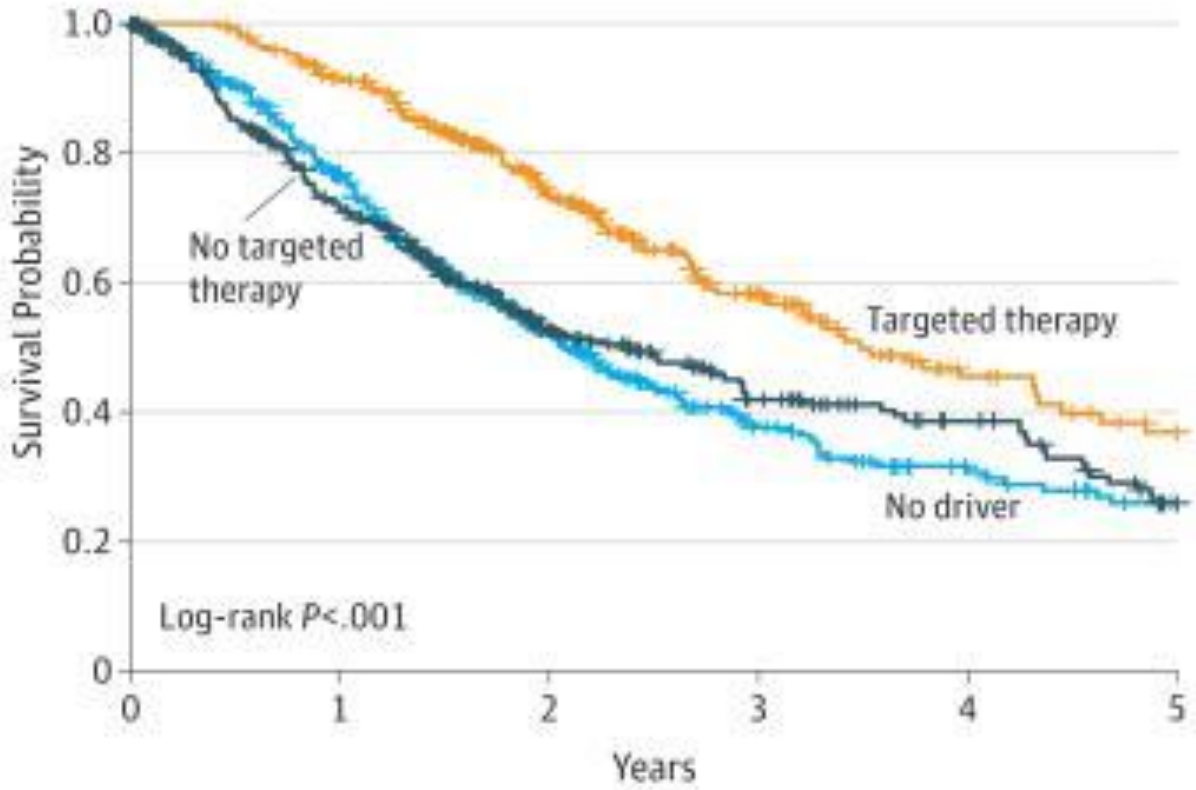
C. E. McCoach[1], G. M. Blumenthal[2], L. Zhang[3], A. Myers[2], S. Tang[3], R. Sridhara[3], P. Keegan[2], R. Pazdur[2], R. C. Doebele[1] & D. Kazandjian[2,4*]


1.0
0.8
0.6
0.4
0.2
0
Survival probability
0
6
12
18
24
No tumor shrinkage
Shrinkage 1-25%
Shrinkage 26-50%
Shrinkage 51-75%
Shrinkage 76-100%

Landmark analysis

- Choose a time point (landmark time) as the new baseline
- Create categories of responder vs non-responder at the landmark time
  - Beware: if a patient responded after the landmark, then he is or she is a non-responder for this analysis
- Estimate by KM survival from the time of landmark
- If response time is variable
  - Choice of landmark time becomes crucial
  - One needs to balance the number of patients wo will be excluded (died before landmark) against those responded after the landmark

A Patients with an oncogenic driver mutation who did and did not receive targeted therapy, and patients without an ocogenic driver



No. at risk

Patients with oncogenic driver						
No targeted therapy	318	205	110	64	43	20
Targeted therapy	260	225	143	72	36	23
Patients with no driver	360	250	122	59	36	23

Time-dependent covariates

- Another way to deal with this
- Allow the covariate to vary over time
- One can be a non-responder for the first four months than become a responder
- Use the risk set to track the appropriate denominators
- One can derive a test but no KM for this methodology
- There are also “dynamic models”, i.e. every time a new scan is taken the model is updated.

Summary of issues

- Response is not known at baseline (start of treatment)
- Comparing survival by a post-baseline covariate is tricky
  - Some people never get the response assessment (Death, clinical progression etc)
  - Those who respond, by definition, must have lived up to the time of response assessment
- Landmark analysis
- Time dependent covariates

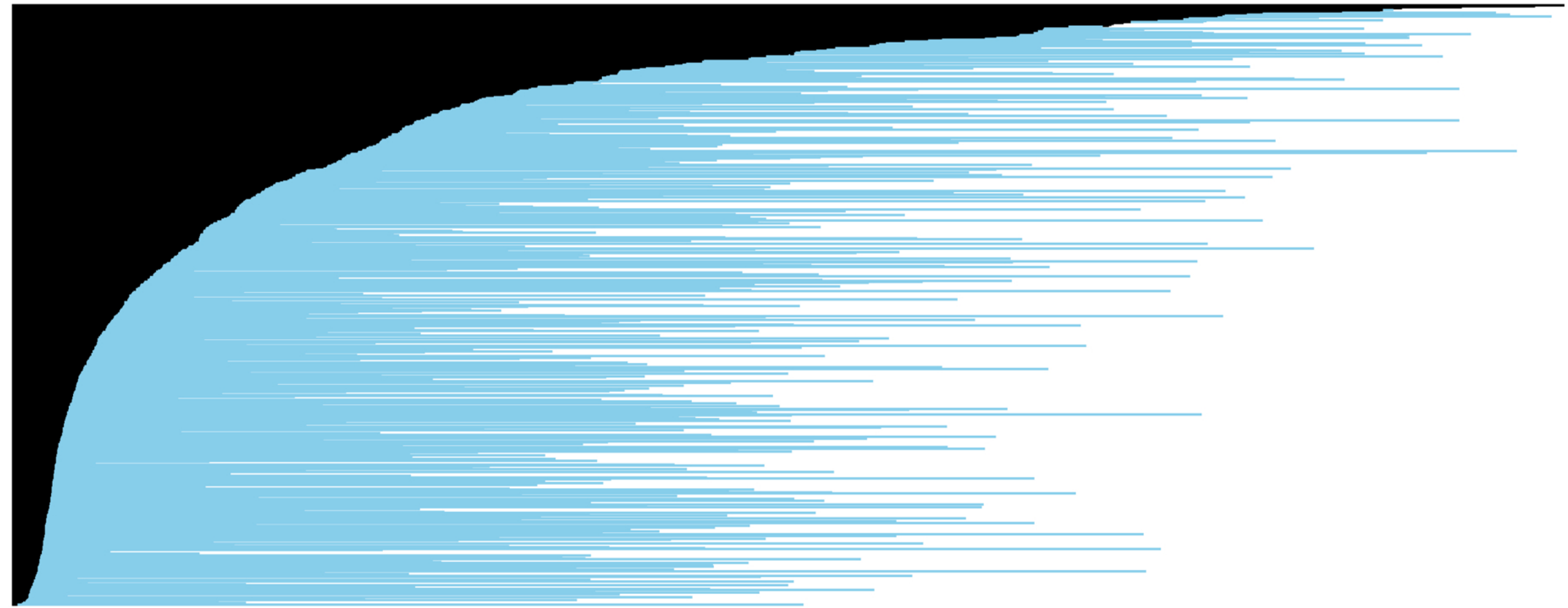
Left Truncation

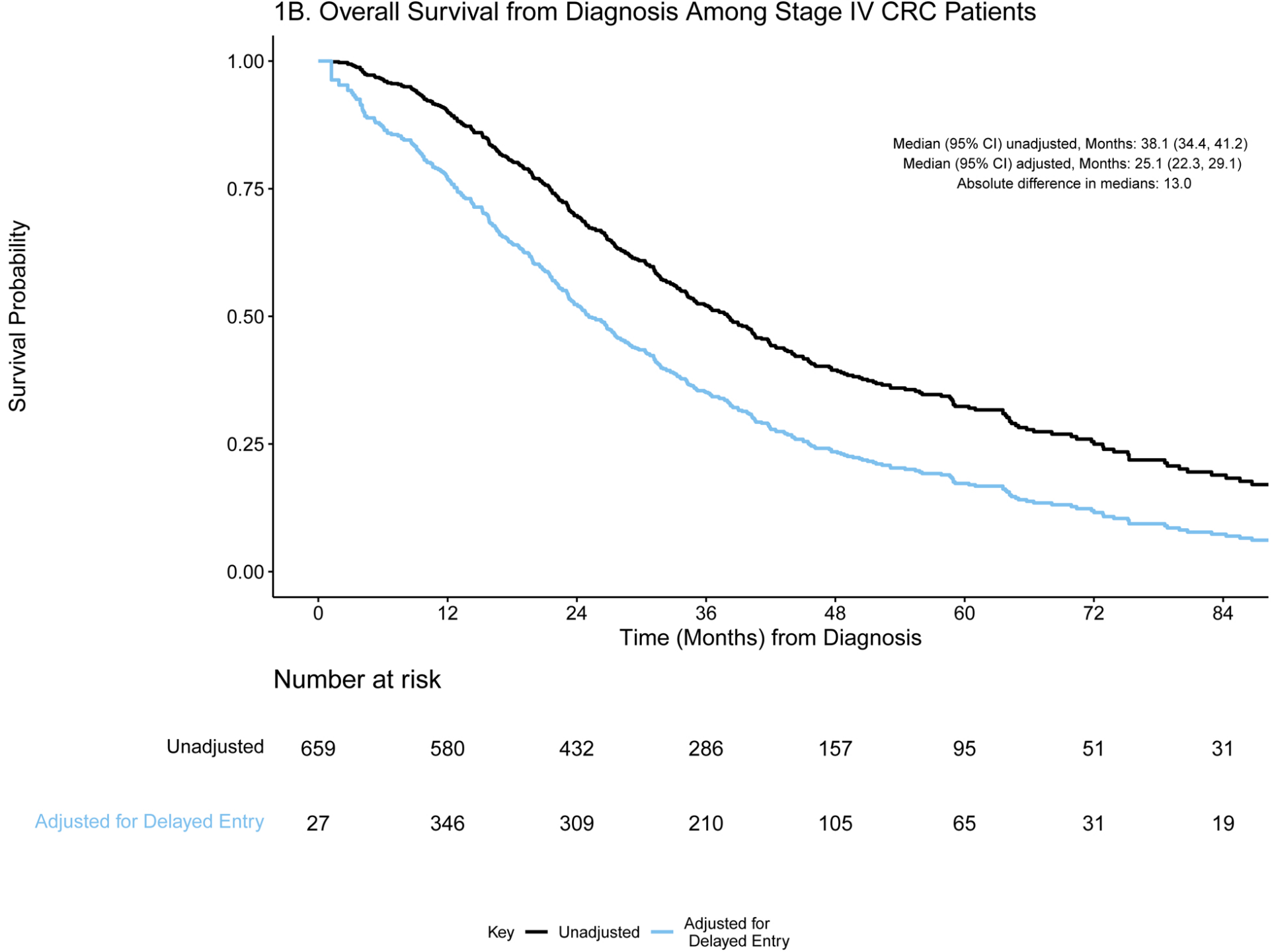
- Common in clinical genomic studies
- Most obvious when some patients' genomic material is harvested at progression, but we want to analyze time from diagnosis
- By definition, that patient was not at risk of death during the period between diagnosis and progression
- This is called left truncation (late entry into the risk set)
- We need to account for left truncation in these studies

Selection Bias

- When there is left truncation there is usually selection bias; some patients do not make it to sequencing at all
- This is mostly a problem of selective sequencing. If selection is for clinical reasons than it could add to bias as well.
- You can see if there is selection bias by plotting the KM curve of unsequenced patients to those sequenced (adjusted for left truncation)
  - Brown S, Lavery JA, Shen R, Martin AS, Kehl KL, Sweeney SM, Lepisto EM, Rizvi H, McCarthy CG, Schultz N, Warner JL, Park BH, Bedard PL, Riely GJ, Schrag D, Panageas KS; AACR Project GENIE Consortium. Implications of Selection Bias Due to Delayed Study Entry in Clinical Genomic Studies. JAMA Oncol. 2022 Feb 1;8(2):287-291. doi: 10.1001/jamaoncol.2021.5153. PMID: 34734967; PMCID: PMC9190030.

1A. Event History for Overall Survival from Diagnosis Among Stage IV CRC Patients (N=659)




1B. Overall Survival from Diagnosis Among Stage IV CRC Patients
Survival Probability
1.00
0.75
0.50
0.25
0.00
Median (95% CI) unadjusted, Months: 38.1 (34.4, 41.2)
Median (95% CI) adjusted, Months: 25.1 (22.3, 29.1)
Absolute difference in medians: 13.0
0
12
24
36
48
60
72
84
Time (Months) from Diagnosis
Number at risk
Unadjusted 659 580 432 286 157 95 51 31
Adjusted for Delayed Entry 27 346 309 210 105 65 31 19
Key Unadjusted Adjusted for Delayed Entry