Lecture 3

Monday, September 18, 2023 10:06 AM

BIOS522_Slides3

EMORY ROLLINS SCHOOL OF PUBLIC HEALTH | Department of Biostatistics and Bioinformatics

BIOS 522: Survival Analysis Methods

Lecture 3:

Log-rank test

Previously

- Defined the survival function
- Estimated the survival function via the empirical CDF or Kaplan-Meier estimator
- Characterized whether censoring is informative or non-informative

2

Today's learning objectives

- *Define the null and alternative hypotheses for the log-rank test*
- *Calculate the log-rank test statistic*
- *Calculate a weighted log-rank test statistic*
- *Calculate the stratified log-rank test statistic*
- *Interpret the results of a log-rank statistic in the context of an*

Today's learning objectives

- *Define the null and alternative hypotheses for the log-rank test*
- *Calculate the log-rank test statistic*
- *Calculate a weighted log-rank test statistic*
- *Calculate the stratified log-rank test statistic*
- *Interpret the results of a log-rank statistic in the context of an applied problem*

3

ORIGINAL ARTICLE

Ticagrelor or Prasugrel in Patients with Acute Coronary Syndromes

Stefanie Schüpke, M.D., Franz-Josef Neumann, M.D., Maurizio Menichelli, M.D., Katharina Mayer, M.D., Isabell Bernlochner, M.D., Jochen Wöhrle, M.D., Gert Richardt, M.D., Christoph Liebetrau, M.D., Bernhard Witzenbichler, M.D., David Antoniucci, M.D., Ibrahim Akin, M.D., Lorenz Bott-Flügel, M.D., et al., for the ISAR-REACT 5 Trial Investigators*

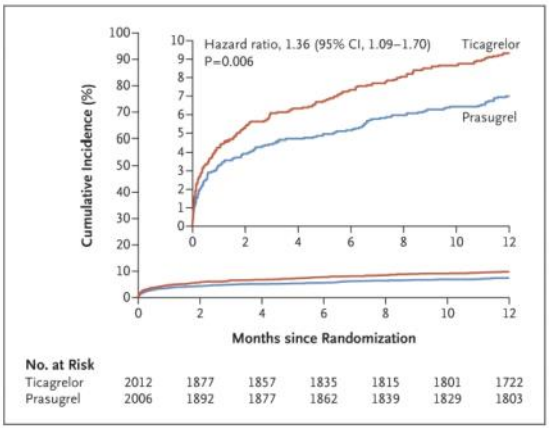


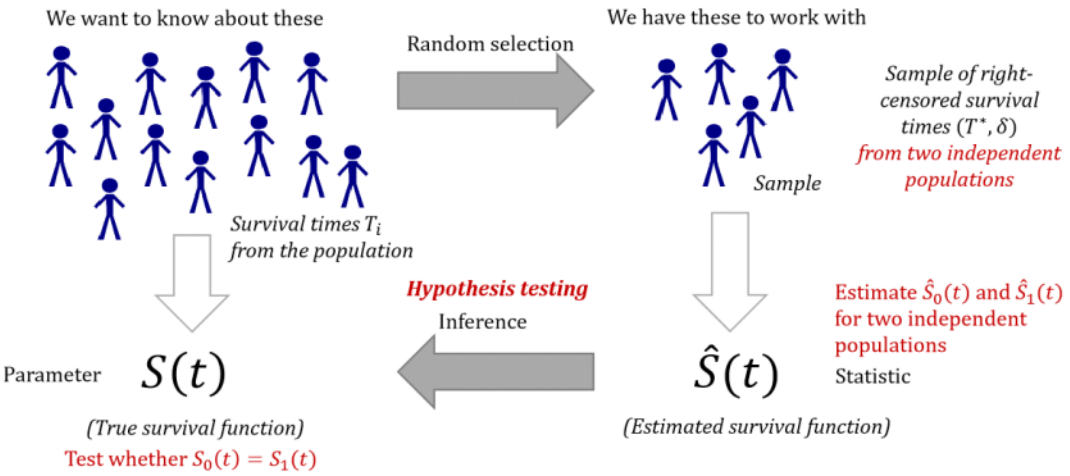
Figure 2. Cumulative Incidence of the Primary End Point at 1 Year.

The Kaplan–Meier curves show the cumulative incidence of the primary end point, which was a composite of death, myocardial infarction, or stroke at 1 year. The inset shows the same data on an enlarged y axis.

4

Source: Schüpke et al. (2019) NEJM doi: 10.1056/NEJMoa1908973

Statistical inference



Adapted from: https://www.cliffsnotes.com/study-guides/statistics/sampling/populations-samples-parameters-and-statistics

5

Hypothesis testing for survival data

Null hypothesis H_0

- $S_0(t) = S_1(t)$ for all t
- $S_0(\cdot) = S_1(\cdot)$
- The survival probabilities are equal at all times t

Alternative hypothesis H_A or H_1

- $S_0(t) \neq S_1(t)$ for some t
- $S_0(\cdot) \neq S_1(\cdot)$
- The survival probabilities are different at some time(s) t

6

Log-rank test

- Most commonly used statistical test for comparing the survival functions of two or more independent groups

- At each distinct failure time, we compare the failure rates between the two groups
- The test adds up the differences in failure rates over time and standardizes these by the variance under the null hypothesis

Calculating the log-rank test statistic

- Consider each distinct failure time j

	At risk just before t_j	Fails at t_j	Survives past t_j
Group 0	n_{0j}	d_{0j}	$n_{0j} - d_{0j}$
Group 1	n_{1j}	d_{1j}	$n_{1j} - d_{1j}$
Total	$n_j = n_{0j} + n_{1j}$	$d_j = d_{0j} + d_{1j}$	$n_j - d_j$

Calculating the log-rank test statistic

- For each distinct failure time j, we observe O_j and calculate E_j, V_j

$$O_j = d_{0j}$$

$$E_j = \frac{n_{0j}}{n_j} d_j$$

$$V_j = \frac{n_{0j} n_{1j} d_j (n_j - d_j)}{n_j^2 (n_j - 1)}$$

9

Calculating the log-rank test statistic

- For J total unique failure times, calculate:

$$O = \sum_{j=1}^{J} O_j, \quad E = \sum_{j=1}^{J} E_j, \quad V = \sum_{j=1}^{J} V_j$$

- The **log-rank test statistic** is:

$$Z = \frac{O - E}{\sqrt{V}}$$

10

Calculating the log-rank test statistic

- At each distinct failure time, we create a table of this form:

	Fail	Survive	At risk
Group 0	0	12	**12**
Group 1	2	8	**10**
	2	**20**	**22**

- We observed 0 failures in group 0 and 2 failures in group 1
- Under H_0 and given that 2 failures were observed, we expect 1.09 failures in group 0
  - $12/22 = 54.5\%$ of the population at risk is in group 0
  - $2 \times 0.545 = 1.09$ expected failures

- $12/22 = 54.5\%$ of the population at risk is in group 0
- $2 \times 0.545 = 1.09$ expected failures

Interpreting the log-rank test statistic

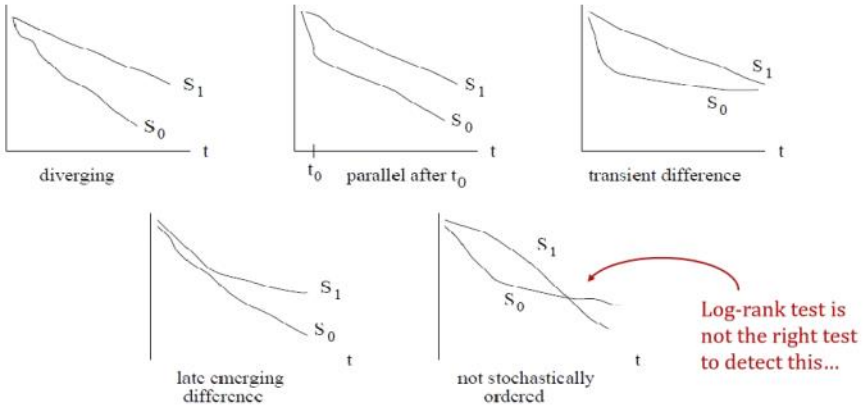
- The Z test statistic follows a standard normal distribution, so we reject the null hypothesis if $Z > 1.96$ or $Z < -1.96$

- Equivalently, the log-rank test statistic can be calculated as:

$$Z^2 = \frac{(O-E)^2}{V}$$

- The Z^2 test statistic follows a chi-squared distribution, so we reject the null hypothesis if $Z^2 > 3.84$

The log-rank test is best suited to capture differences where one group has a consistently higher failure rate than the other



Other tests available later will learn

Example: Hemophiliacs

We examine times from primary AIDS diagnosis until death for hemophiliacs. We wish to compare survival for patients who were at most 40 years of age at the time of diagnosis to survival for hemophiliacs who were over the age of 40 at diagnosis.

Consider the right-censored survival times (in months) for the two age groups:

Group 0 (Age <40): 2, 3+, 6, 6, 8, 10+, 15, 15, 16, 27, 30, 32

Group 1 (Age ≥ 40): 1, 1, 2, 4, 4, 5, 5+, 7, 14+, 22

Example: Hemophiliacs

- For the first group (age < 40):

$$O = 10$$

$$E = 13.46$$

$$Z^2 = \frac{(O-E)^2}{V} = 4.23$$

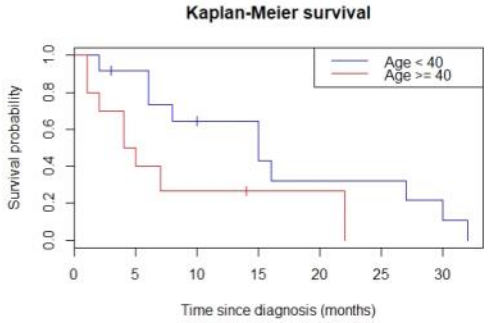
```
Call:
survdiff(formula = Surv(time, event) ~ agegt40, data = dat)

             N Observed Expected (O-E)^2/E (O-E)^2/V
agegt40=0 12       10    13.46     0.891      4.23
agegt40=1 10        8     4.54     2.645      4.23

 Chisq= 4.2  on 1 degrees of freedom, p= 0.0398
```

- $Z^2 > 3.84$
- p=0.0398 < 0.05
- Reject H_0 and conclude that the two groups have different survival

- *Survival is better in the first group because the number of observed deaths is lower than the number of expected deaths*

Example: Hemophiliacs

Or... we can examine a plot

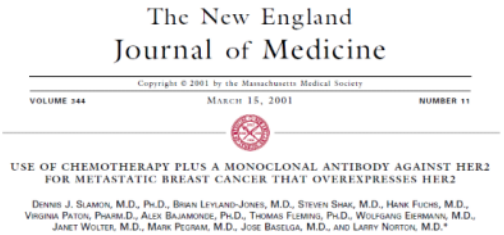


16

Example: Metastatic breast cancer

- Goal: Investigators conducted a **randomized trial** to assess the efficacy and safety of **adding a monoclonal antibody, trastuzumab**, to standard chemotherapy for the treatment of **metastatic breast cancers** in which the HER2 gene is overexpressed. Cancers with overexpression of this gene are more aggressive.

- Population: **469 women** with metastatic breast cancer that overexpressed HER2 who had not previously received chemotherapy for metastatic disease were randomized to receive either **standard chemotherapy** alone or **standard chemotherapy plus trastuzumab.**


The New England
Journal of Medicine

Copyright © 2001 by the Massachusetts Medical Society

VOLUME 344 MARCH 15, 2001 NUMBER 11

USE OF CHEMOTHERAPY PLUS A MONOCLONAL ANTIBODY AGAINST HER2 FOR METASTATIC BREAST CANCER THAT OVEREXPRESSES HER2

DENNIS J. SLAMON, M.D., PH.D., BRIAN LEYLAND-JONES, M.D., STEVEN SHAK, M.D., HANK FUCHS, M.D., VIRGINIA PATON, PHARM.D., ALEX BAJAMONDE, PH.D., THOMAS FLEMING, PH.D., WOLFGANG EIERMANN, M.D., JANET WOLTER, M.D., MARK PEGRAM, M.D., JOSE BASELGA, M.D., AND LARRY NORTON, M.D.*

17

Source: Slamon et al. (2001) NEJM https://www.nejm.org/doi/full/10.1056/nejm200103153441101

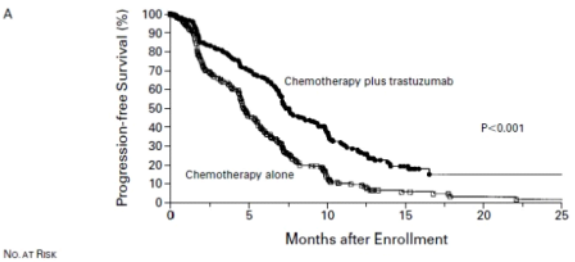
Example: Metastatic breast cancer

- Outcome variable: The primary outcome was **disease-free survival**. This is a **composite endpoint** defined as **time from randomization** until the first occurrence of the following: local, regional, and distant tumor recurrence, contralateral breast cancer, including ductal carcinoma in situ, other second primary cancers, or death.
- Predictor variables: For the primary efficacy analysis, the only predictor variable considered was **trial arm** (with or without trastuzumab).
- Statistical analysis: Researchers fit **Kaplan-Meier estimators** to estimate progression-free survival in each of the two trial arms.

18

Example: Metastatic breast cancer

- Results: Figure 1A summarizes the Kaplan-Meier curves for the primary efficacy analysis. The median time to disease progression was 7.4 months in the trastuzumab group and 4.6 months in the comparator group.
- Participants randomized to trastuzumab have **longer progression-free survival** than participants randomized to placebo. The difference is statistically significant ($p<0.001$) by the log-rank test.



No. at Risk				
Chemotherapy plus trastuzumab	235	152	63	15
Chemotherapy alone	234	103	25	6

19

Weighted log-rank tests

- The standard log-rank test sums over all distinct failure times; all times receive equal weight:

$$Z = \frac{\left(\sum_{j=1}^{J} O_j - \sum_{j=1}^{J} E_j\right)}{\sqrt{\sum_{j=1}^{J} V_j}}$$

- Because the numbers at risk are larger at *earlier* failure times, we may prefer to give these times more weight

20

Weighted log-rank tests

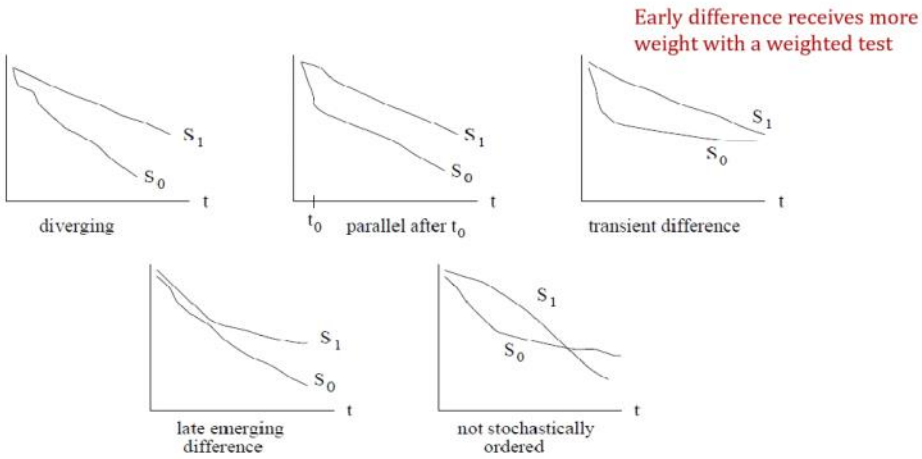
- **Generalized Wilcoxon test**

$$Z_w = \frac{\sum_{j=1}^{J} n_j\left(O_j - E_j\right)}{\sqrt{\sum_{j=1}^{J} n_j^2 V_j}}$$

- This will tend to emphasize early differences in survival
- Other weighted tests include the **Tarone-Ware test**, the **Peto-Prentice test**, the **Efron test**, and the **Harrington-Fleming test**

21

Early diff: Bigger teset statistic



Late diff: Smaller test statistic

22

Example: Lagged treatment effects

- In various settings, one expects some lag period before an intervention is fully effective
- Assuming a threshold lag, the weights for times before that threshold could be set to zero (not counted)
- In the Women's Health Trial, the designers expected a linear lag, and proposed a log-rank test with corresponding linearly increasing weights

23

Source: Zucker and Lakatos (1990) Biometrika https://doi.org/10.1093/biomet/77.4.853

Extending the log-rank test

- So far, we have described tests for comparing survival of two independent groups
- The standard log-rank test and weighted log-rank test can be easily extended to test that survival is equal for **three or more groups**

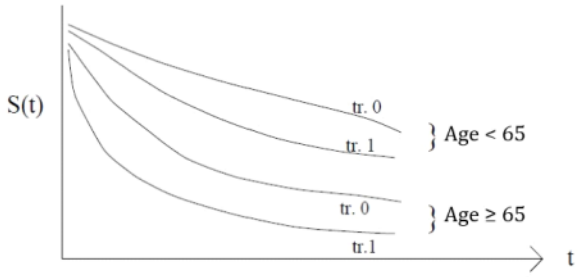
- *Examples*:
  - Survival on Drug A = Survival on Drug B = Survival on Drug C
  - H_0: $S_0(\cdot) = S_1(\cdot) = S_2(\cdot)$

24

Stratified log-rank test

- In some cases, we may want to compare survival across groups but **adjusting/controlling** for another categorical covariate

- In the example shown, we are interested in differences in survival across treatment groups
- We are *not* interested in differences in survival across ages



25

Stratified log-rank test formula

Cosndier

$$H_0: S_0^{(k)}(\cdot) = S_1^{(k)}(\cdot)$$

- Because the numbers at risk are larger at *earlier* failure times, we may prefer to give these times more weight

Stratified log-rank test formula

- Imagine the population is divided into $k = 1, \dots, K$ strata

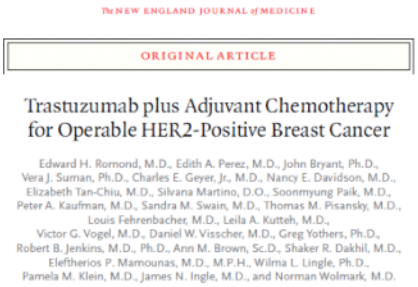
$$O_k = \sum_{j=1}^{J} O_{kj} \qquad E_k = \sum_{j=1}^{J} E_{kj} \qquad V_k = \sum_{j=1}^{J} V_{kj}$$

$$Z_S = \frac{\sum_{k=1}^{K} O_k - \sum_{k=1}^{K} E_k}{\sqrt{\sum_{k=1}^{K} V_k}}$$

$$H_0: S_0^{(k)}(\cdot) = S_1^{(k)}(\cdot)$$

Example: Metastatic breast cancer

- Goal: Investigators analyzed the results of **two randomized trials** designed to compare adjuvant chemotherapy **with or without concurrent trastuzumab** in women with surgically removed HER2-positive **breast cancer**.
- Population: The National Surgical Adjuvant Breast and Bowel Project trial B-31 included **2043 women**, and the North Central Cancer Treatment Group trial N9831 included **1633 women** who contributed to the combined analysis (some trial arms were excluded). Median follow-up was 2.0 years.


The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

Trastuzumab plus Adjuvant Chemotherapy for Operable HER2-Positive Breast Cancer

Edward H. Romond, M.D., Edith A. Perez, M.D., John Bryant, Ph.D., Vera J. Suman, Ph.D., Charles E. Geyer, Jr., M.D., Nancy E. Davidson, M.D., Elizabeth Tan-Chiu, M.D., Silvana Martino, D.O., Soonmyung Paik, M.D., Peter A. Kaufman, M.D., Sandra M. Swain, M.D., Thomas M. Pisansky, M.D., Louis Fehrenbacher, M.D., Leila A. Kutteh, M.D., Victor G. Vogel, M.D., Daniel W. Visscher, M.D., Greg Yothers, Ph.D., Robert B. Jenkins, M.D., Ph.D., Ann M. Brown, Sc.D., Shaker R. Dakhil, M.D., Eleftherios P. Mamounas, M.D., M.P.H., Wilma L. Lingle, Ph.D., Pamela M. Klein, M.D., James N. Ingle, M.D., and Norman Wolmark, M.D.

Source: Romond et al. (2005) NEJM https://www.nejm.org/doi/full/10.1056/nejmoa052122

Example: Metastatic breast cancer

- Outcome variable: The primary outcome was **disease-free survival**. This is a composite endpoint defined as time from randomization until the first occurrence of the following: local, regional, and distant tumor recurrence, contralateral breast cancer, including ductal carcinoma in situ, other second primary cancers, or death.
- Predictor variables: The main predictor of interest for the primary efficacy analysis is trial arm (**with or without trastuzumab**). Researchers also considered the **study** (trial B-31 vs. trial N9831), intended chemotherapy **schedule** (every three weeks vs. weekly), **nodal status** (0, 1-3, 4-9, or 10+ positive nodes), and **hormone-receptor status** (estrogen- or progesterone-receptor positive vs. estrogen- and progesterone-receptor negative).

Example: Metastatic breast cancer

- Statistical analysis: Comparison of the two groups was based on a log-rank test, stratified according to the study trial, intended chemotherapy schedule, nodal status, and hormone-receptor status. Thus, there were 2x2x4x2=32 strata.

30

Example: Metastatic breast cancer

- Results: Figure 2A below summarizes the Kaplan-Meier curve for the primary efficacy analysis of disease-free survival. The percentage of patients alive and disease-free at three years was 75.4% in the control group and 87.1% on the trastuzumab group. The **stratified log-rank test p-value is <0.0001**, indicating significant improvement in disease-free survival in the trastuzumab group **adjusting for key covariates**.

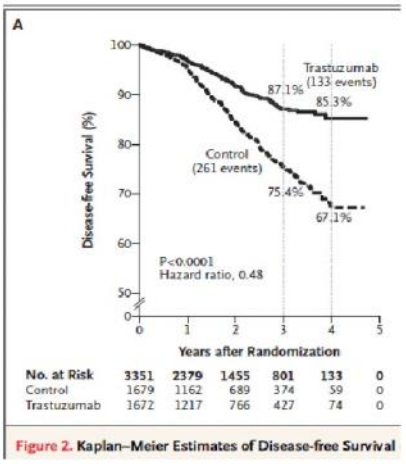
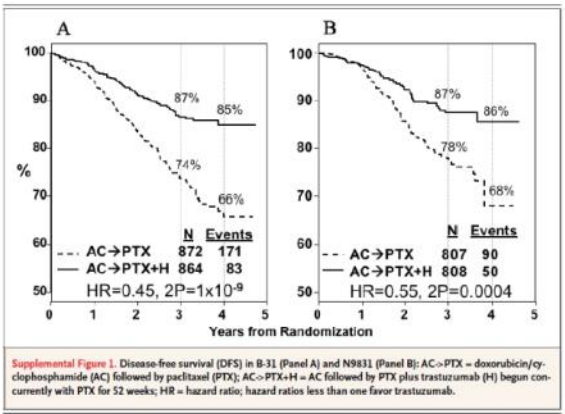


Figure 2. Kaplan–Meier Estimates of Disease-free Survival

31

Example: Metastatic breast cancer

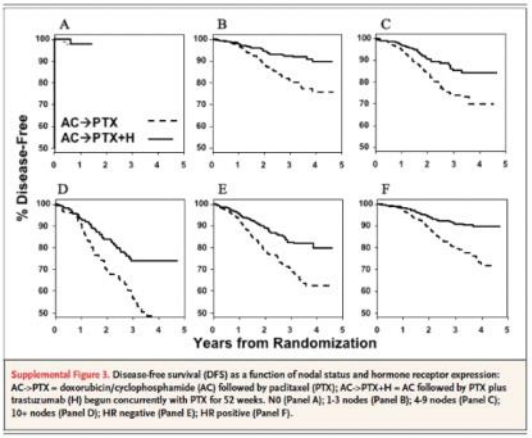


Supplemental Figure 1. Disease-free survival (DFS) in B-31 (Panel A) and N9831 (Panel B); AC->PTX = doxorubicin/cyclophosphamide (AC) followed by paclitaxel (PTX); AC->PTX+H = AC followed by PTX plus trastuzumab (H) begun concurrently with PTX for 52 weeks; HR = hazard ratio; hazard ratios less than one favor trastuzumab.

- Stratified Kaplan-Meier curves for all 32 strata are not provided, but the Supplementary Appendix includes several helpful plots. Supplemental Figure 1 summarizes disease-free survival stratified by trial (B-31 in Panel A, N9831 in Panel B).

32

Example: Metastatic breast cancer



Supplemental Figure 3. Disease-free survival (DFS) as a function of nodal status and hormone receptor expression: AC->PTX = doxorubicin/cyclophosphamide (AC) followed by paclitaxel (PTX); AC->PTX+H = AC followed by PTX plus trastuzumab (H) begun concurrently with PTX for 52 weeks. N0 (Panel A); 1-3 nodes (Panel B); 4-9 nodes (Panel C); 10+ nodes (Panel D); HR negative (Panel E); HR positive (Panel F).

- Supplemental Figure 3 summarizes the Kaplan-Meier curves for patients with 0 positive nodes (Panel A), 1-3 positive nodes (Panel B), 4-9 positive nodes (Panel C), 10+ positive nodes (Panel D), hormone receptor negative (Panel E), and hormone receptor positive (Panel F). Across all subgroups, disease-free survival is higher for the participants randomized to trastuzumab.

33

Benefits of stratification

- Survival may vary a lot across levels of the stratifying factors
- For non-randomized studies, the two groups may be imbalanced with respect to these factors, leading to confounding
  - Stratification can address this confounding
- For randomized trials, the two groups are expected to be balanced so confounding is not a concern and stratification is not necessary for validity
  - But adjusting for other covariates that are predictive of survival may decrease variability in the population and improve our ability to detect differences across groups

34

Looking ahead

- So far, we have talked about inference on the survival function, but this is not the only function used to characterize time-to-event random variables
- Next week, we will introduce the **hazard function** and **cumulative hazard function**
- The hazard function is the basis of the most popular regression model for time-to-event data – the **Cox proportional hazards model**

35

Today's activity

- R computing