ATLAS
ASIAN CLINICAL TRIALS NETWORK FOR CANCERS PROJECT

National Cancer Center Japan

Randomization and Confounding

National Cancer Center, Japan

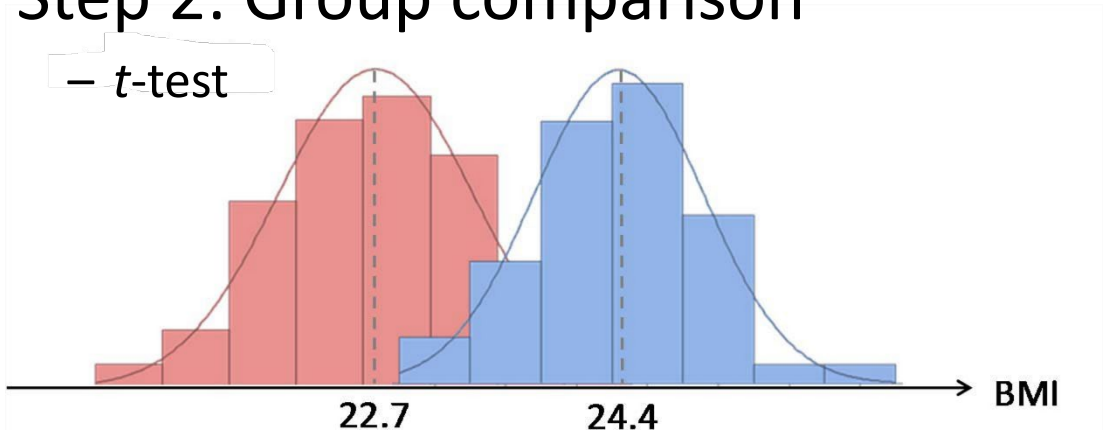
Biostatistics Division/JCOG Data Center

Junki Mizusawa

JCOG
Japan Clinical Oncology Group

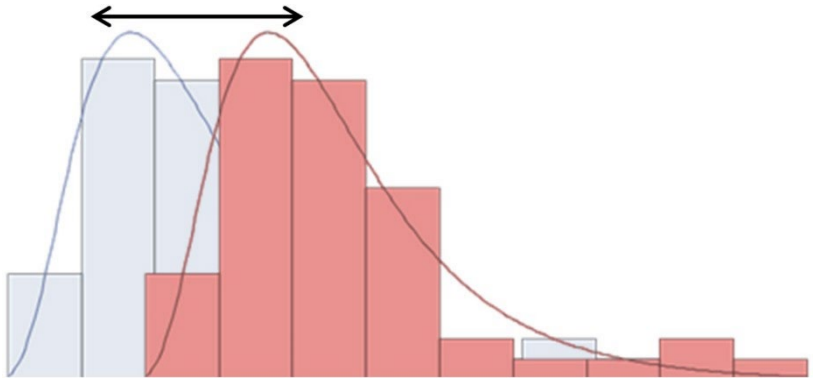
Review: Analysis of Continuous Variable Data

- Step 1: Confirm distribution using histogram
- Step 2: Group comparison
  - *t*-test



$$t = \frac{\overline{X}_A - \overline{X}_B}{\sqrt{V(\frac{1}{n_A} + \frac{1}{n_B})}}$$

  - Wilcoxon rank sum test



	BMI (kg/m^2)				Mean
Group A	22.0 [2]	28.3 [8]	19.4 [1]	22.3 [3]	23.0
Group B	23.3 [4]	25.1 [7]	24.6 [6]	23.5 [5]	24.1

Review: Analysis of Binary Data

- Step 1: Create a (2 × 2) contingency table and count frequencies

		Onset of colorectal cancer		
		yes	no	Total
Aspirin	yes	129	4932	5061
	no	87	2440	2527
	Total	216	7372	7588

- Step 2: Group comparison
  - Fisher's exact test
  - One-sided *P* value = 0.0053 + ・ ・ ・ + 1.2×10^{-106} = 0.0174

216	4845	5061
0	2527	2527
216	7372	7588

Pr = 2.4×10^{-37}

215	4846	5061
1	2526	2527
216	7372	7588

Pr = 1.3×10^{-35}

214	4847	5061
2	2525	2527
216	7372	7588

Pr = 5.0×10^{-34}

. . .

129	4932	5061
87	2440	2527
216	7372	7588

Pr = 0.0053

Same contingency table as observed data

. . .

0	5061	5061
216	2311	2527
216	7372	7588

Pr = 1.2×10^{-106}

Review: Analysis of Binary Data

- Step 1: Create a (2 × 2) contingency table and count frequencies
- Step 2: Group comparison
  – Chi-square test

Observed values		Expected values	
129 (2.5%)	**4932 (97.5%)**	**144 (2.8%)**	**4917 (97.2%)**
87 (3.4%)	**2440 (96.6%)**	**72 (2.8%)**	**2455 (97.2%)**

		Onset of colorectal cancer		
		yes	no	Total
Aspirin	yes	129	4932	5061
	no	87	2440	2527
	Total	216	7372	7588

✓ Test based on the difference between observed and expected frequencies
The larger the difference between the observed and expected frequencies, the rarer the result observed under the null hypothesis.

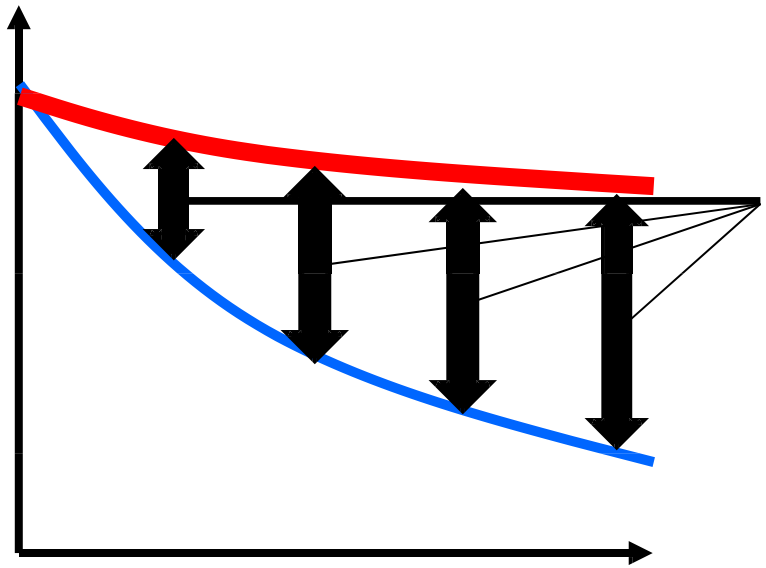
✓ Statistic $\frac{(129-144)^2}{144}+\frac{(4932-4917)^2}{4917}+\frac{(87-72)^2}{72}+\frac{(2440-2455)^2}{2455}$

✓ Approximates a chi-square distribution with one degree of freedom
The higher the expected frequency, the better the approximation.

✓ P = 0.0273 (Significant at the two-sided 5% level!)

Review: Survival Analysis

- Step 1: Confirmation of the survival curve using Kaplan–Meier method
  - Estimation of median survival time and annual survival rate
- Step 2: Group comparison
  - log-rank test



✓ By summing up the deviations in the survival curves at each point of event occurrence, the overall difference in the curves is compared.

✓ In other words, the further apart the two survival curves are, the larger the gap (= lower P value).

Review: Outcome Types and Statistical Methods

	Continuous variable	Binary (0/1)	Survival time
Outcome examples	Blood pressure, laboratory test values	Response rate	Overall survival, Progression-free survival
Data summary	Histogram	Contingency table	Kaplan–Meier method
Group comparison (test)	*t*-test, Wilcoxon test	Chi-square test, Fisher's exact test	Log-rank test
Model fitting	Multiple regression analysis	Logistic regression	Cox regression

Modified from: Ohashi, Yasuo. Statistics for Oncologists

Is Surgery Effective in Patients with Unresectable Colorectal Cancer and Liver Metastases?

Original article

Long-term survival of patients with unresectable colorectal cancer liver metastases following infusional chemotherapy with 5-fluorouracil, leucovorin, oxaliplatin and surgery

S. Giacchetti,[1] M. Itzhaki,[1] G. Gruia,[1] R. Adam,[2] R. Zidani,[1] F. Kunstlinger,[1] S. Brienza,[1] E. Alafaci,[1] F. Bertheault-Cvitkovic,[1] C. Jasmin,[1] M. Reynes,[3] H. Bismuth,[2] J. L. Misset[1] & F. Lévi[1]

[1]*Centre de Chronothérapie, Fédération des Maladies Sanguines Immunitaires et Tumorales, Service d'Oncologie Médicale;* [2]*Centre Hépato-biliaire,* [3]*Service d'Anatomie Pathologique, Hôpital Paul Brousse, Villejuif, France*

Summary

Context: Long-term survival of patients with metastatic colorectal cancer has been achieved only in patients who underwent complete resection of metastases. Such surgery could be performed in a greater proportion of patients if effective chemotherapy could downstage previously unresectable metastases. This approach has been limited by the low tumor response rate achieved with conventional chemotherapy.

Objective: We studied the outcome of patients with initially unresectable liver metastases from colorectal cancer treated with a three-drug chemotherapy regimen followed by liver metastases surgery whenever possible.

Patients and methods: From March 1988 to June 1994, 151 patients with colorectal liver metastases were considered initially unresectable because of large tumor size (>5 cm), multinodular (>4) or ill-located metastases. All patients received fully ambulatory chemotherapy with 5-fluorouracil, leucovorin and oxaliplatin (chronotherapy in 83% of them). They were periodically reassessed for surgery by a joint medico-surgical team.

Results: In 151 patients, the size of liver metastases decreased by >50% in 89 patients (59%) and median overall survival was 24 months (95% confidence interval (95% CI): 19–28 months), with 28% surviving at five years (20%–35%). Surgery with curative intent was attempted in 77 patients (51%), complete resection of liver metastases was achieved in 58 patients (38%). The median survival of the 77 operated patients was 48 months (25–71), with a five-year survival rate of 50% (38–61).

Conclusion: This new strategy of combining effective chemotherapy with surgery apparently altered the natural history of unresectable colorectal cancer metastases.

Key words: chronotherapy, colorectal cancer, liver metastases, oxaliplatin, surgery, survival

Giacchetti, S., et al. Ann Oncol. 10.6 (1999): 663-669.

Summary

- Background
  - ✓ Long-term survival of patients with unresectable colorectal cancer and metastases can be expected only if complete resection of the metastatic site is performed.
  - ✓ However, resection is possible only when chemotherapy has had some effect.
- Objective
  - ✓ To elucidate the prognostic value of liver resection in patients with unresectable colorectal cancer and liver metastasis treated with chemotherapy.
- Subject
  - ✓ Data collected from 151 patients with unresectable colorectal cancer and liver metastases from 1988 to 1994
  - ✓ Approximately 83% of patients received FOLFOX

Background Factor

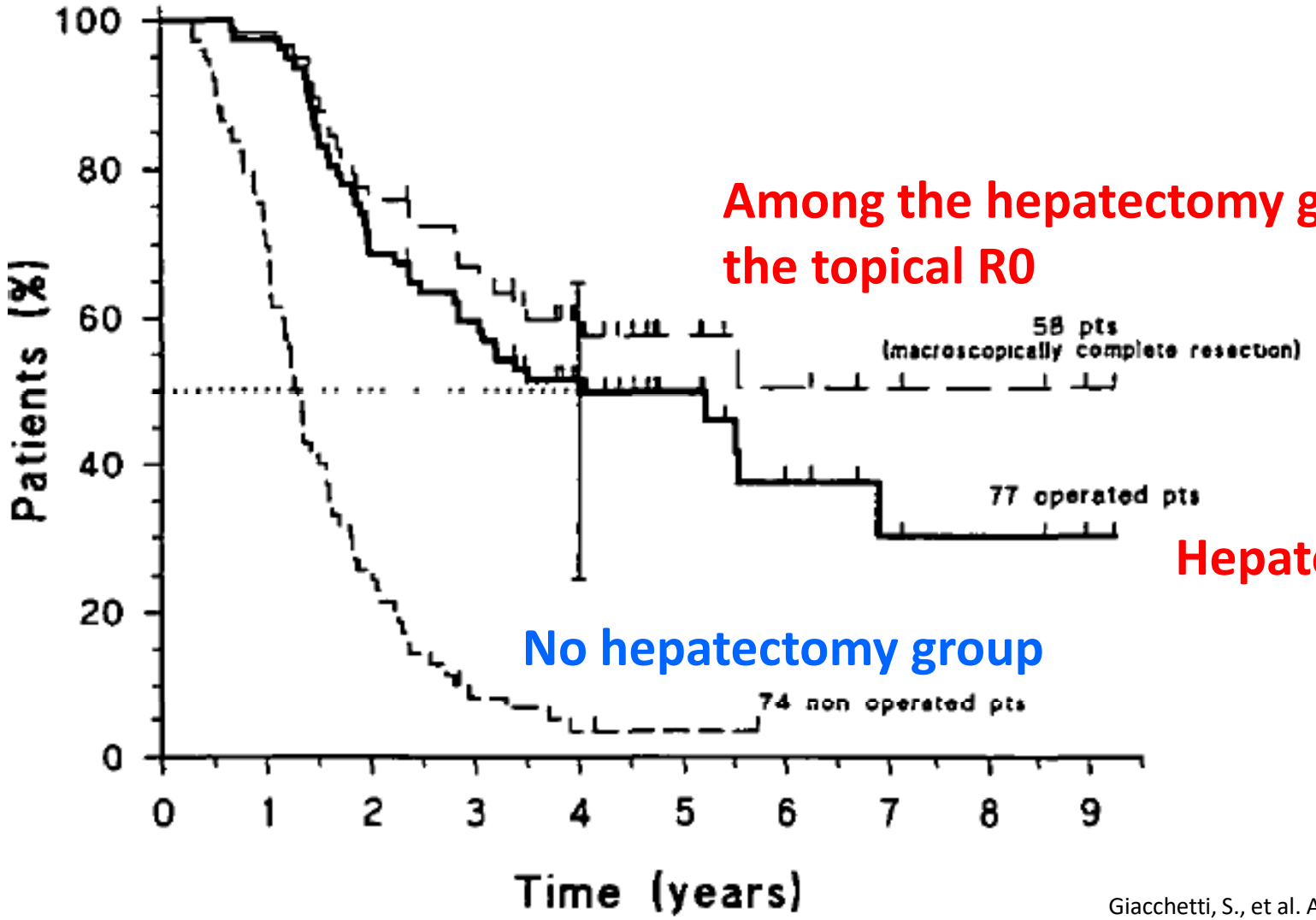
- Common in the hepatectomy group
  - Case of a response
    PR or CR
  - Patients with a small maximum tumor diameter at the metastatic site
  - Patients with one organ with metastasis, etc.

	With hepatectomy N = 77	Without hepatectomy N = 74
Age (years)		
Median (range)	59 (32–79)	58 (27–76)
Best effect (tumor shrinkage by chemotherapy)		
PD or no CX	0	**17**
SD	16	28
PR or CR	**61**	29
Maximum tumor diameter at metastatic site		
≤5 cm	**46**	33
>5 cm	24	**27**
Unknown	7	14
Number of organs with metastasis		
1	**14**	19
2–4	38	16
>4	20	**25**
Unknown	5	14

Best response: CR (complete response), PR (partial response), SD (stable disease), PD (progressive disease), no CX (no chemotherapy)

Giacchetti, S., et al. Ann Oncol. 10.6 (1999): 663-669.

Overall Residency Results



Among the hepatectomy group, the topical R0

Hepatectomy group

No hepatectomy group

Giacchetti, S., et al. Ann Oncol. 10.6 (1999): 663-669.

What Can We Infer from These Results?

1. Hepatectomy is effective
2. Hepatectomy is not considered effective
3. Hepatectomy is not effective

Outline

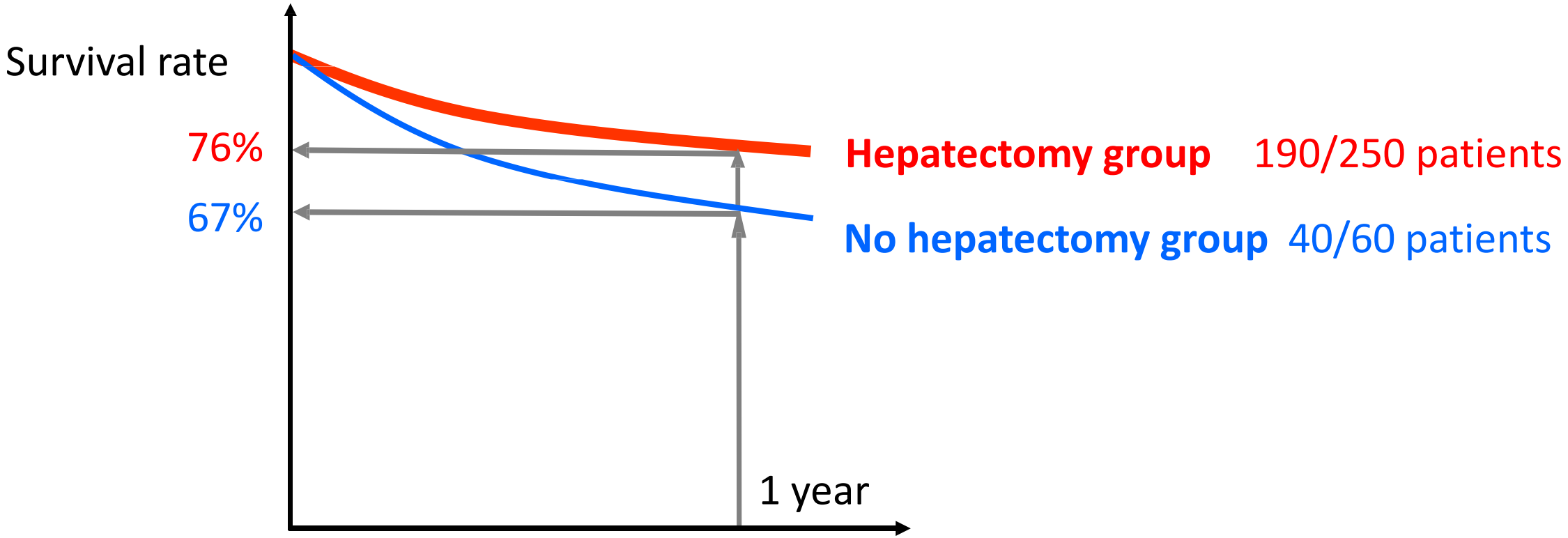
- Confounding and randomization
- Interaction
- Subgroup analysis
- Stratified analysis

Confounding and Randomization

Note: Hypothetical example

Presentations You Will See at Conferences

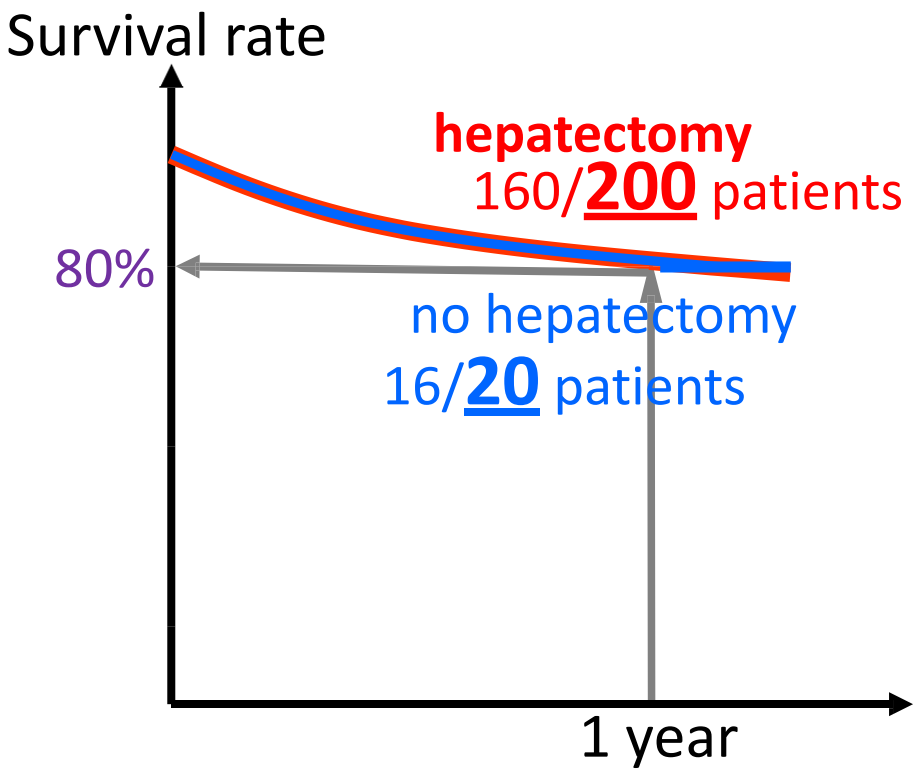
- We retrospectively evaluated patients with unresectable colorectal cancer and liver metastases at our hospital, dividing them into two groups: **hepatectomy group** (250 patients) and **no hepatectomy group** (60 patients).
- The **hepatectomy group** had a better prognosis than the **no hepatectomy group**.



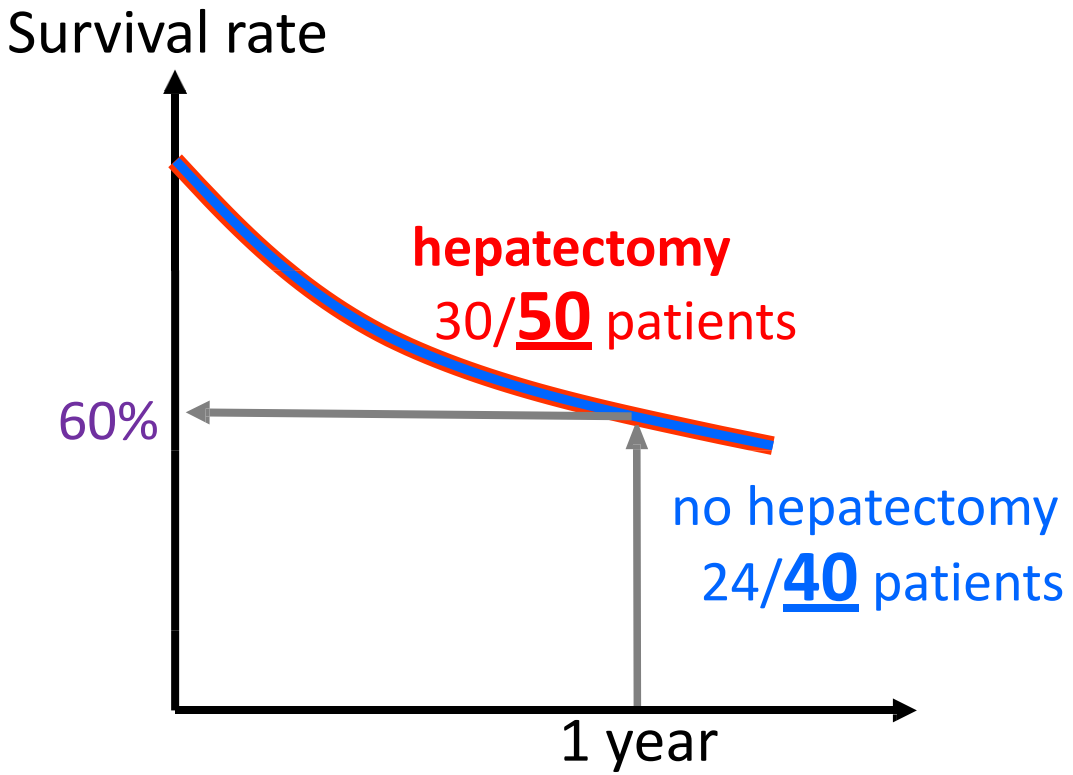
Note: Hypothetical example

Prognosis When Divided by Best Effect

PR or CR
(group responding to chemotherapy)



SD or PD
(group that did not respond to chemotherapy)



The prognosis for **hepatectomy** and **no hepatectomy** groups is the same regardless of the best effect.

Note: Hypothetical example

Since We Want to Compare the Difference Between Hepatectomy and No hepatectomy Groups

- If the conditions of factors affecting prognosis other than the treatment are not the same, it cannot be considered a "comparison"!

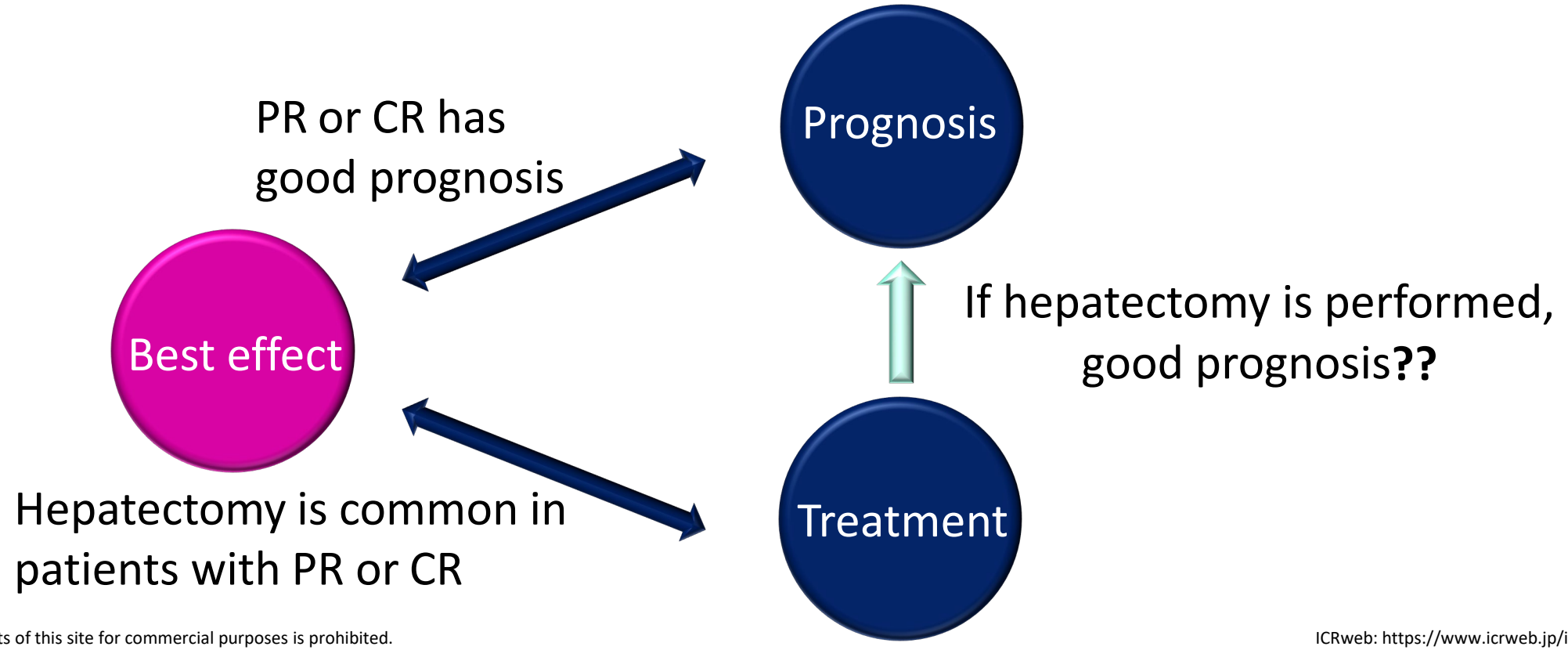
Treatment	PR or CR		SD or PD	Total
hepatectomy	**200 people** (80%)	＞＞	50 people	250 persons
no hepatectomy	20 people (33.3%)	＜＜	**40 people**	60 people

- Hepatectomy group has a higher proportion of "PR or CR" than no hepatectomy group.
- Prognosis depends on best effect (PR or CR has good prognosis)

Note: Hypothetical example

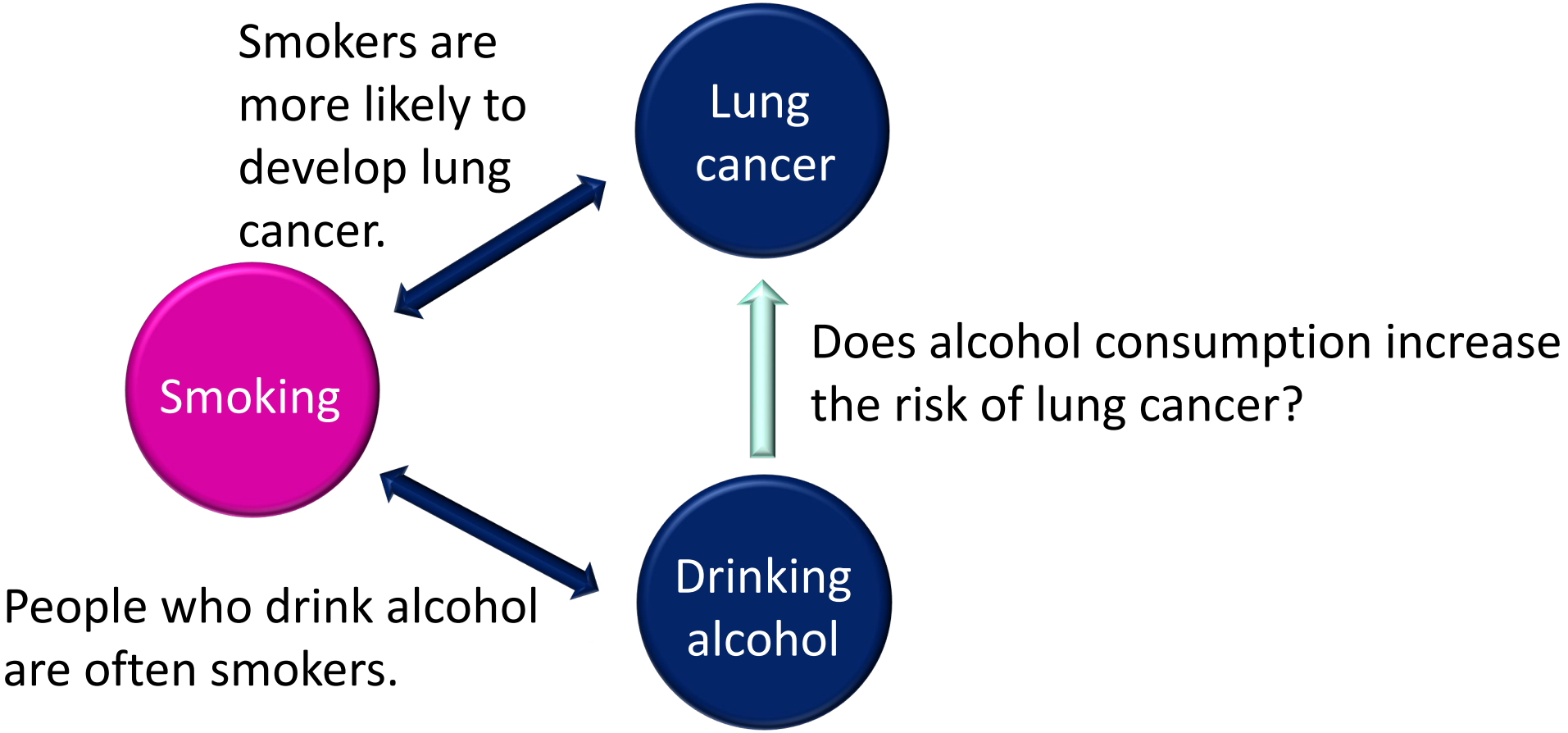
Summary of **Confounding**

- A phenomenon in which a third factor (best effect) related to treatment and prognosis produces an apparent association.
  - The factors that cause confounding (=best effects) **are called confounders**



Other Examples of Confounding

- Alcohol consumption does not cause lung cancer, but it appears to.
  – In this case, smoking is **a confounder**.

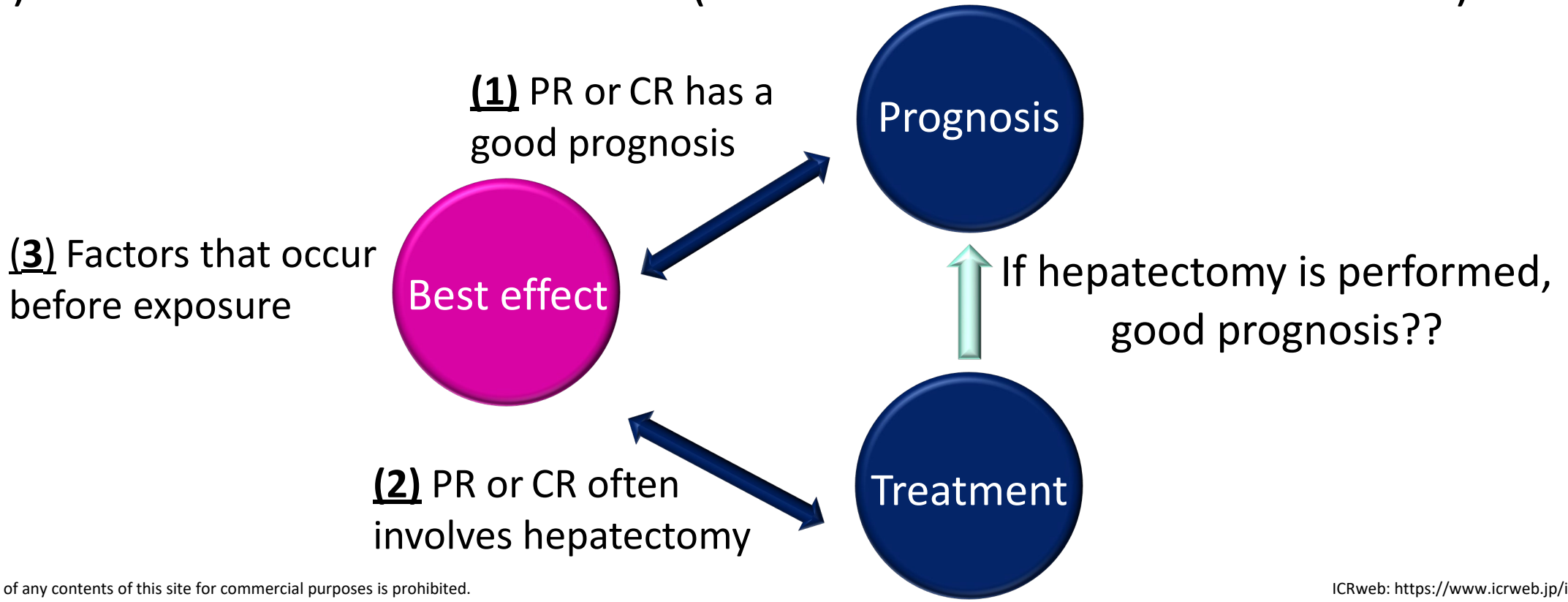


Confounding Requirement

(1) Relates to outcome (prognosis)

(2) Relates to exposure (treatment)

(3) Not an intermediate variable (an event known before treatment)



To Ensure That There Is No Confounding

- Align prognostic background factors across treatment groups
  - Age
  - Stage
  - PS
  - Other (including unknown factors)

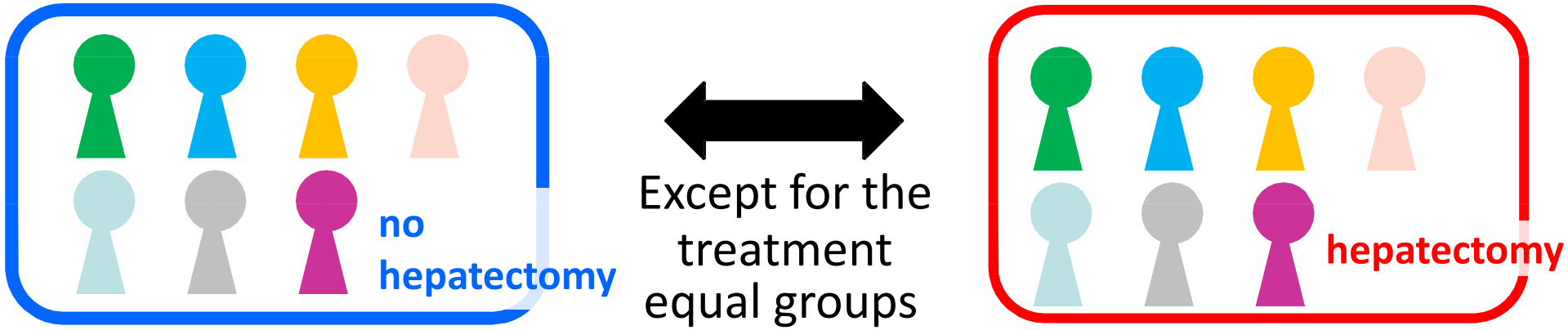
Because there are many known and unknown factors,
it is not possible to consider everything.

Decide **at random**

Note: Hypothetical example

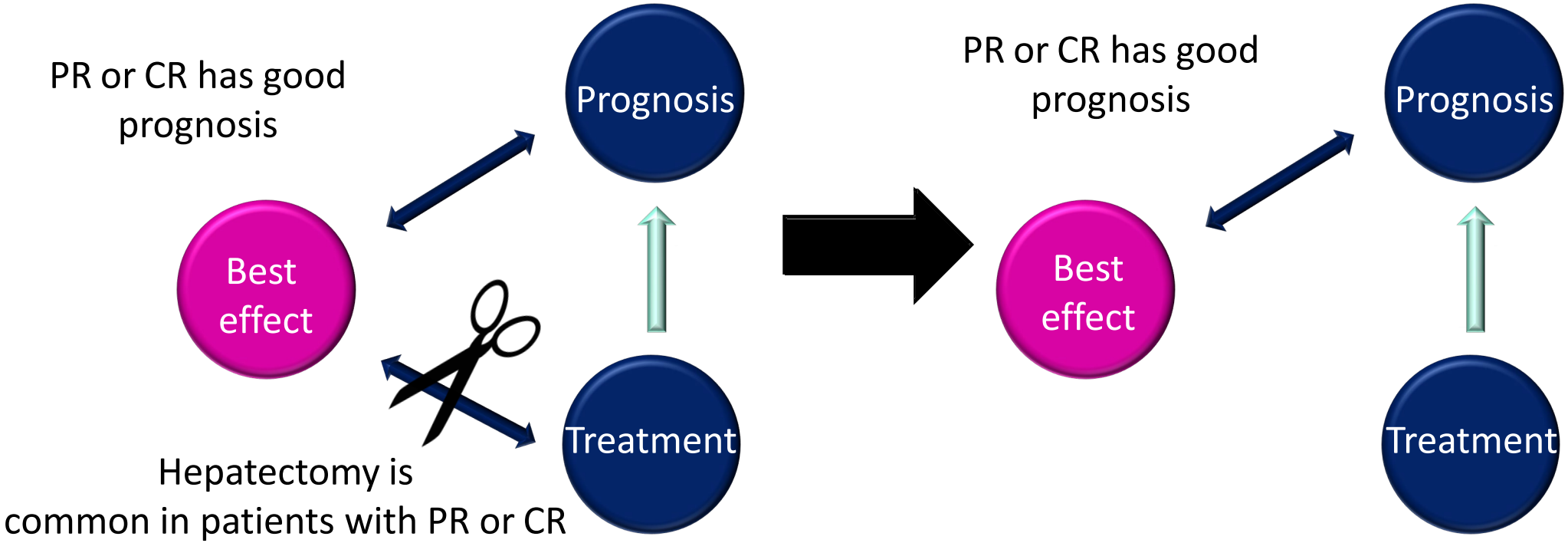
Randomization

- Assign patients to treatment groups based on probability, <u>independent of the physician's or patient's will</u>
- Prevent biased patient selection based on preconceived opinions
  - To prevent patients in good condition from being more likely to be assigned to new drugs, for example
- **<u>The comparability (internal validity)</u> is ensured**
  - Equal groups except for treatment method → If there is a difference in effectiveness, the treatment is different.



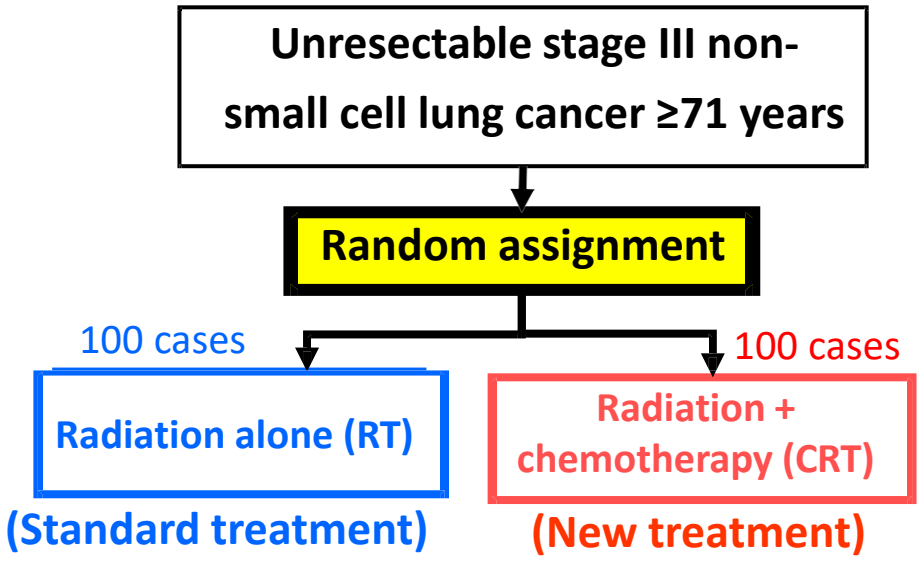
Significance of Randomization

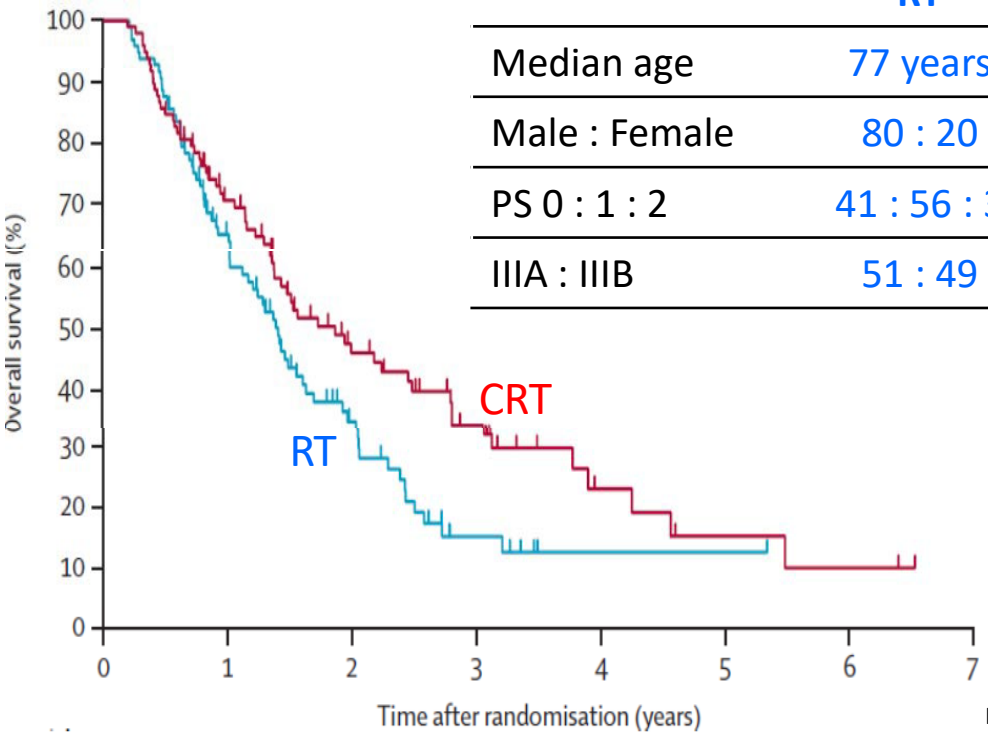
- The association between treatment **and best effects** can be removed
  - Confounding by best effect is eliminated. Thus, the relationship between treatment method and prognosis can be evaluated.
  - Note: The relationship between prognosis and best effect remains



For JCOG0301

- **Randomly assigned to** compare **RT** and **CRT**
  - Background factors other than treatment methods are on average the same between the groups.
  - Differences in survival curves can be expected to be due to differences in treatment methods.



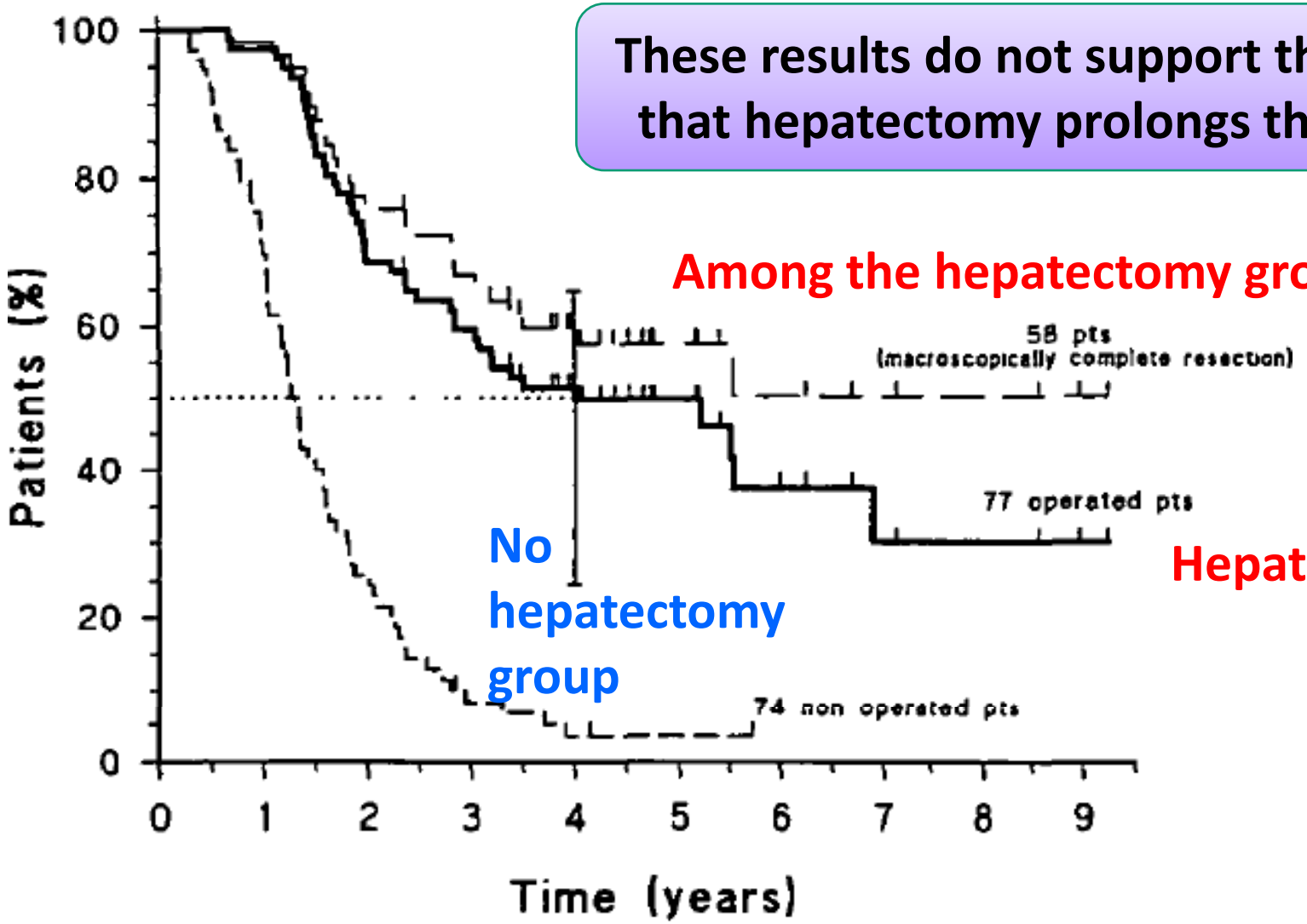


	RT	CRT
Median age	77 years	77 years
Male : Female	80 : 20	84 : 16
PS 0 : 1 : 2	41 : 56 : 3	41 : 55 : 4
IIIA : IIIB	51 : 49	54 : 46

Atagi *et al.* (2012) Lancet Oncol. 13(7): 671-8.

Reposting: Overall Survival Results for Hepatectomy

These results do not support the hypothesis that hepatectomy prolongs the prognosis.



Among the hepatectomy group, the topical R0

Hepatectomy group

No hepatectomy group

Confounding and Misleading Statistical Concepts
– Interaction –

What is the Interaction?

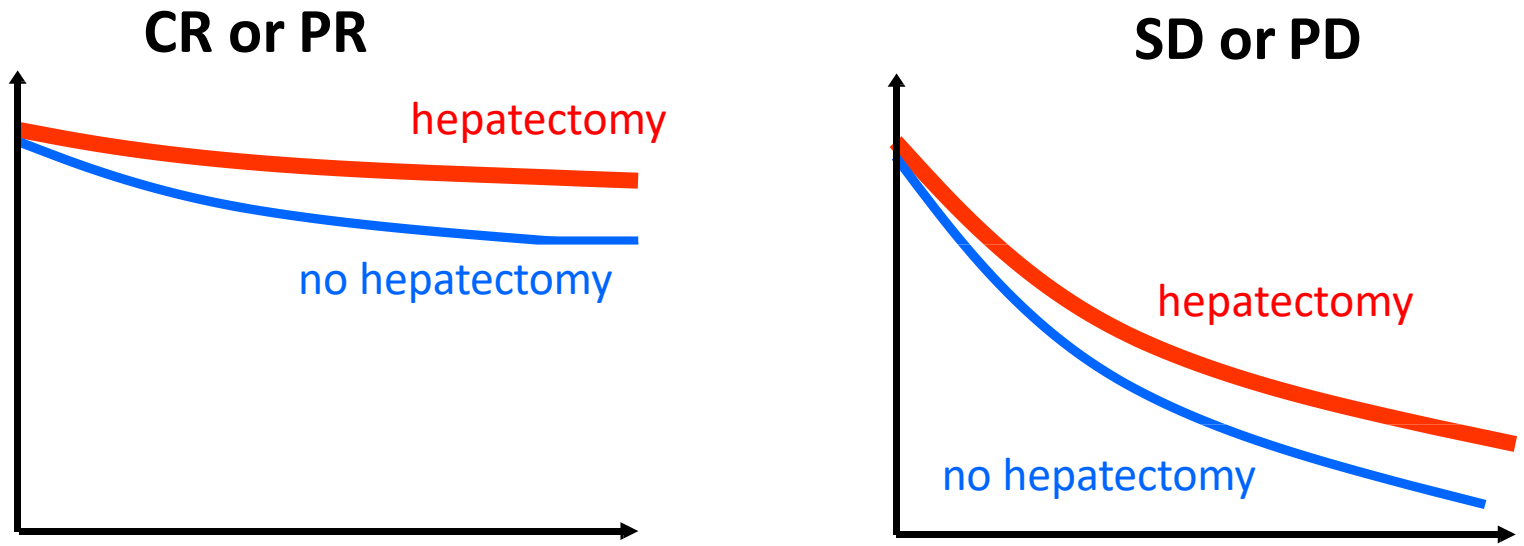
When the magnitude of the treatment effect differs by the subgroups,
“there is an interaction.”

Interaction

- Qualitative interaction
- Quantitative interaction

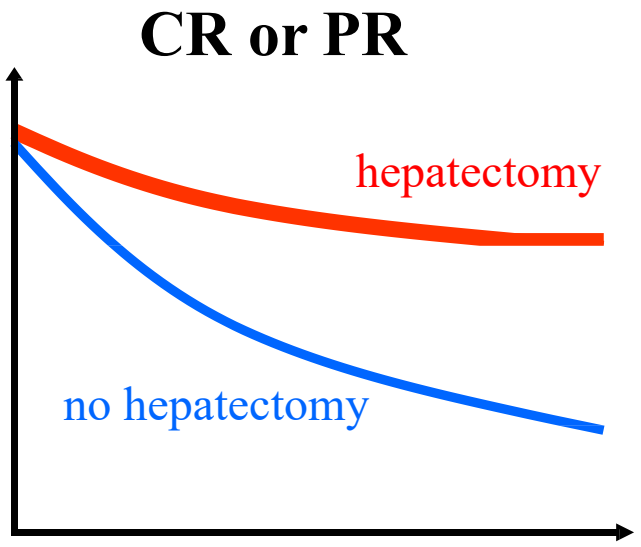
No Interaction

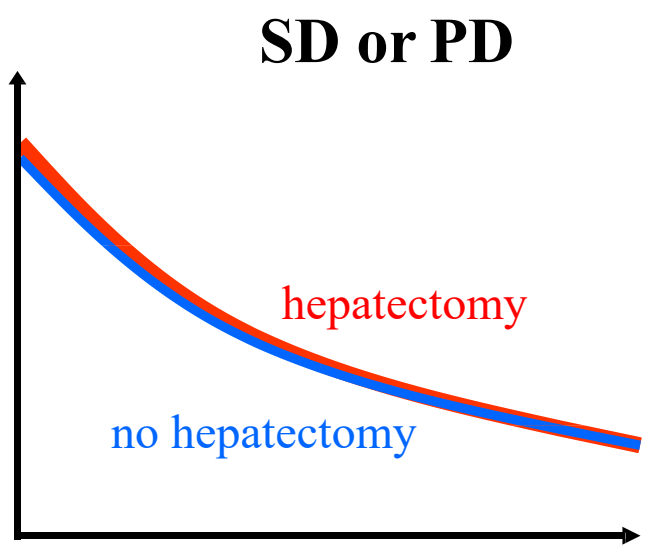
- For the best effect, CR or PR and SD or PD have different prognoses.
- Both CR or PR and SD or PD have a similar level of additional treatment effect with hepatectomy compared with no hepatectomy.
  - No difference in treatment effect in subgroups.
  - The best effect is a prognostic factor



There is an interaction

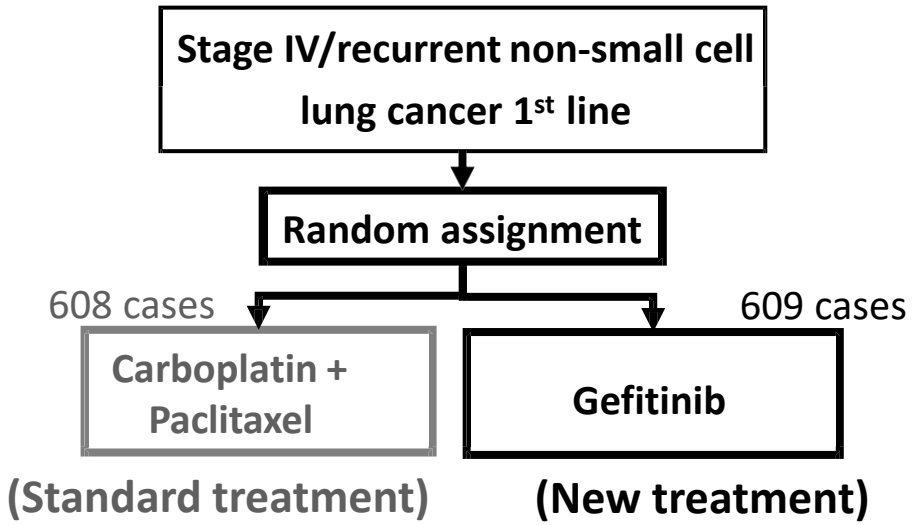
- CR or PR and SD or PD without hepatectomy have the same prognosis
- CR or PR and SD or PD have different treatment effects
  - Only CR or PR shows the effect of hepatectomy
  - Subgroups have different treatment effects.
  - The best effect is a predictive factor for hepatectomy.

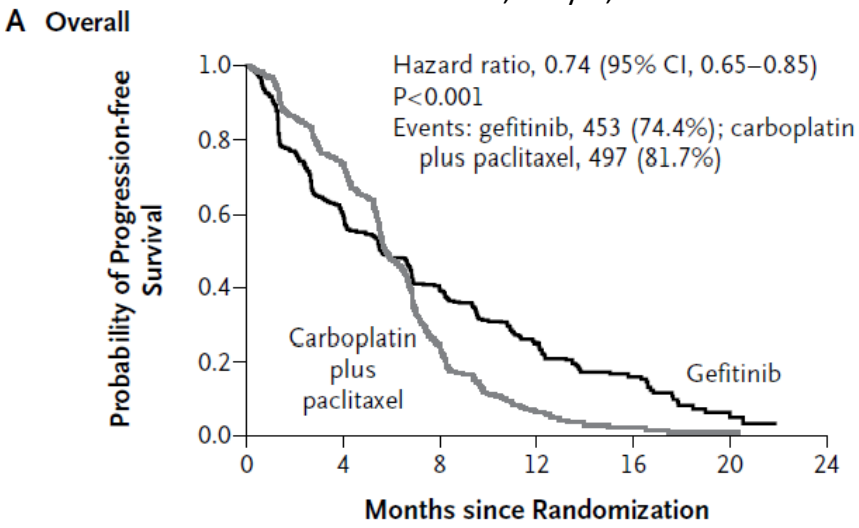


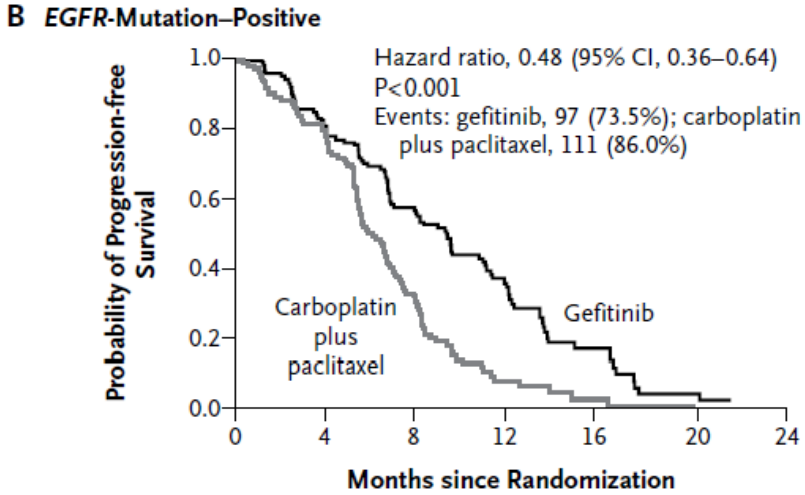


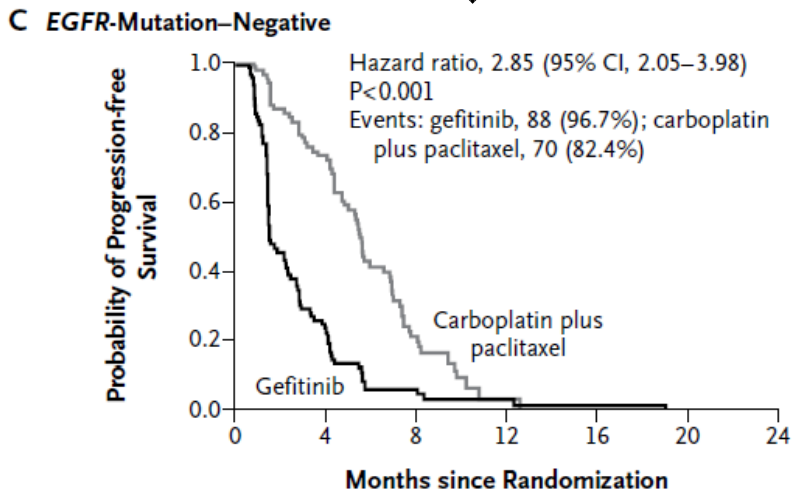
Qualitative Interactions = Predictive Factor for Treatment Selection

Mok, Tony S., et al. *NEJM* 361.10 (2009): 947-957.









Quantitative Interaction = Predictive Factor for Considering Treatment Selection

Post gastrectomy Stage II-IIIB gastric cancer 20–80 years old

Random assignment

608 cases

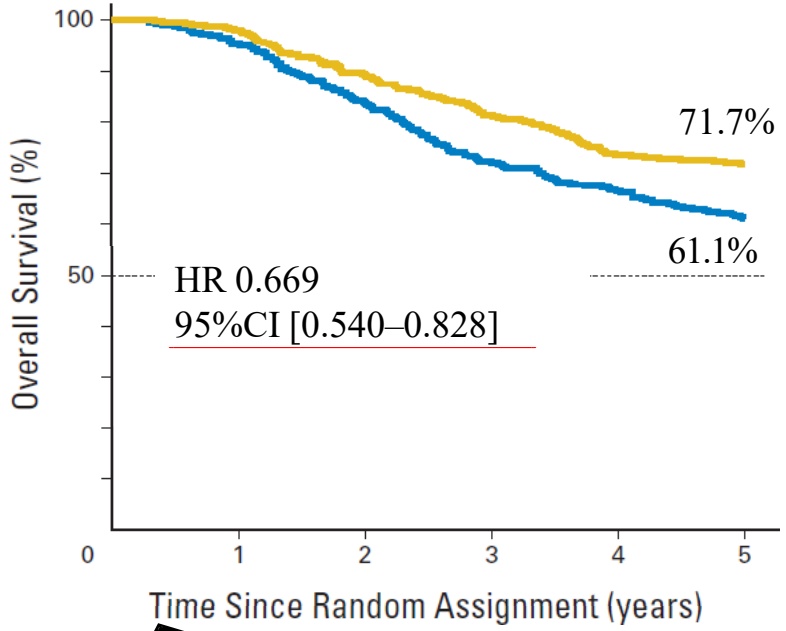
Follow-up (Surgery only)

(Standard treatment)

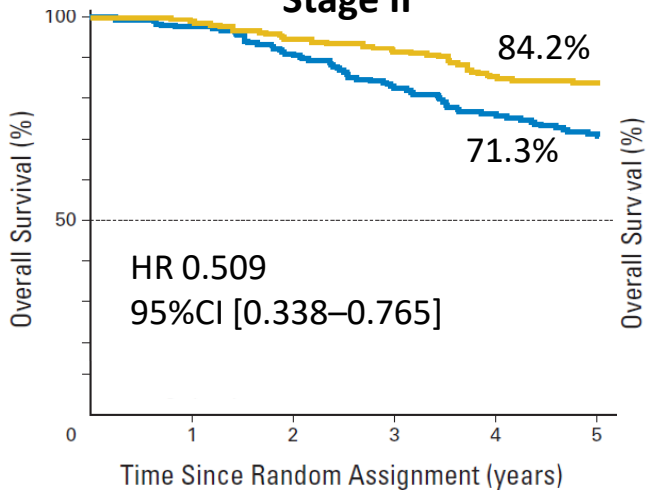
609 cases

S-1 1-year oral administration (S-1)

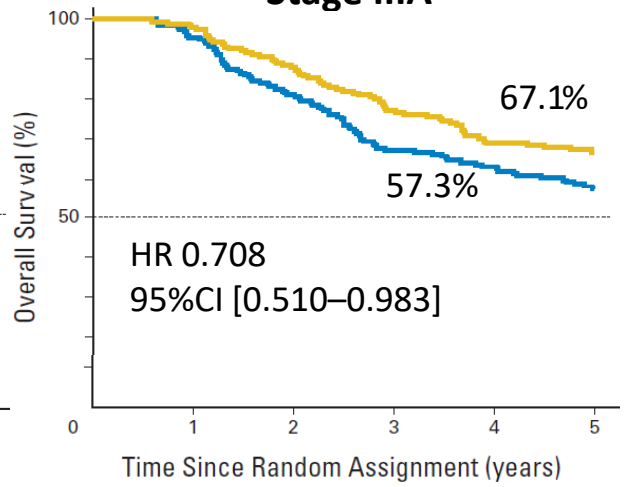
(New treatment)



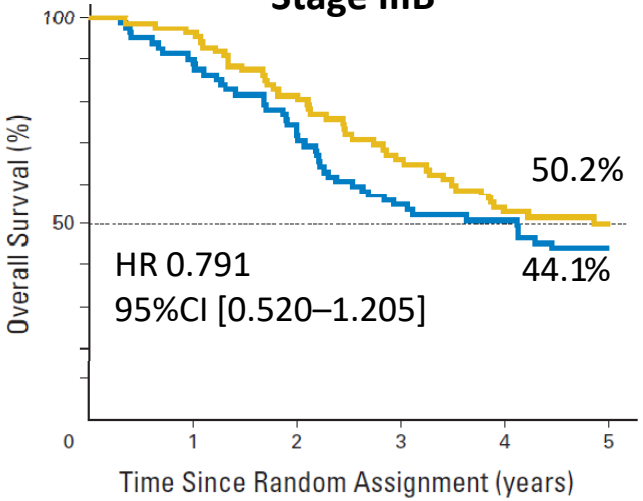
Stage II



Stage IIIA

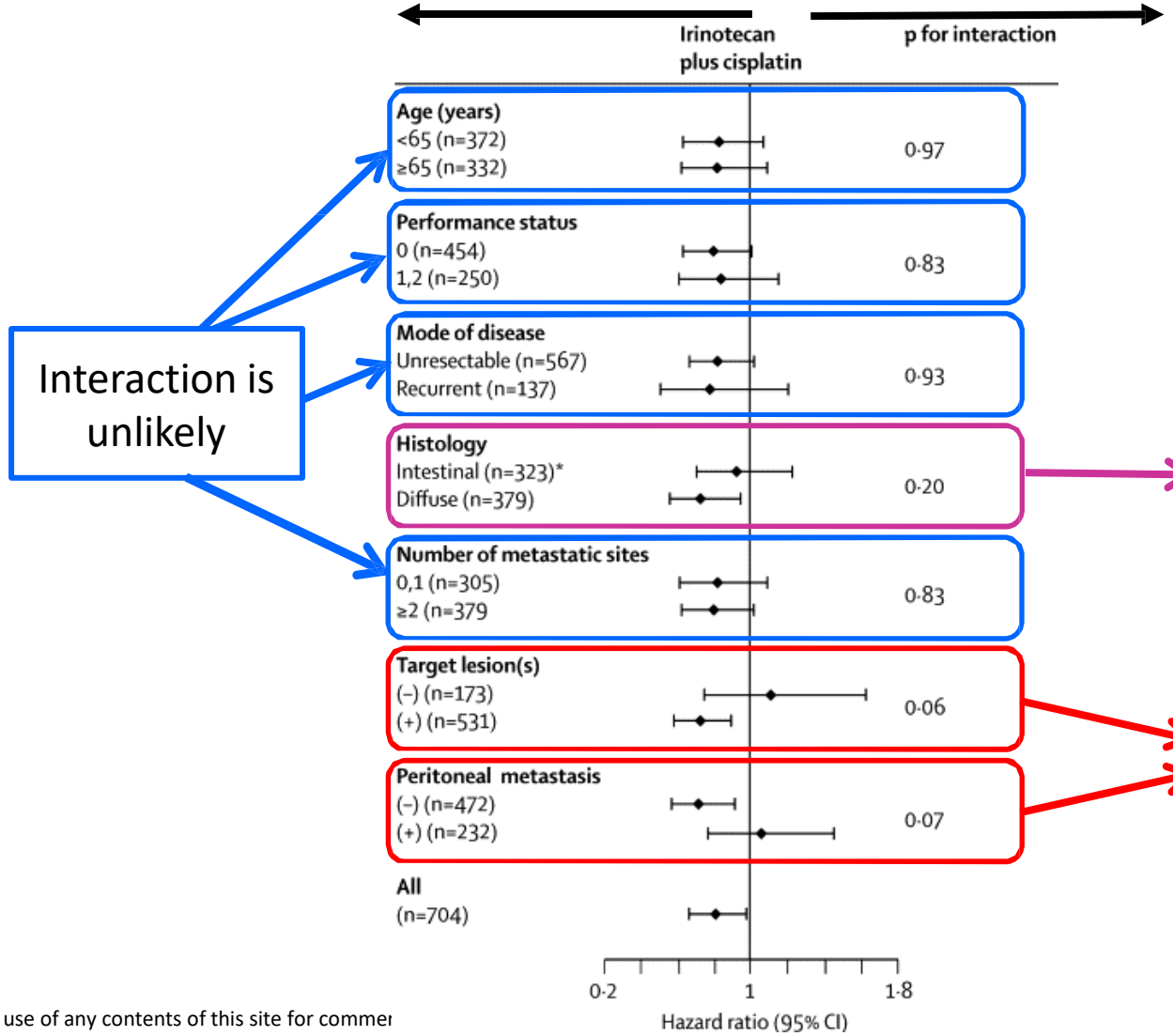


Stage IIIB



Confirmation of Interactions by **Forest Plots**

The new treatment is good ← → Standard treatment is good



Interaction is unlikely

Different from biases, these are not

- something that can be removed
- something that should not be removed.

Therefore, report the result.

Likely to have a quantitative interaction

Likely to have qualitative interaction

Boku N, Lancet Oncol. 2009; 10(11): 1063-9.

Genetic–Environmental Interaction

- Aspirin is effective in preventing adenoma of the large intestine in people with UGT1A6 mutation

Table 1. Risk for colorectal adenoma according to regular aspirin use stratified by UGT1A6 genotype*

Genotype	Nonregular users	Regular users
All UGT1A6 genotypes		
No. of case participants/no. of control participants	373/349	157/183
Age-adjusted OR (95% CI)	1.0 (referent)	0.78 (0.60 to 1.02)
Multivariable OR (95% CI)†	1.0 (referent)	0.76 (0.58 to 0.99)
Wild-type UGT1A6 genotypes		
No. of case participants/no. of control participants	149/154	68/75
Age-adjusted OR (95% CI)	1.0 (referent)	0.90 (0.60 to 1.34)
Multivariable OR (95% CI)†	1.0 (referent)	0.93 (0.60 to 1.44)
Variant UGT1A6 genotypes		
No. of case participants/no. of control participants	224/195	89/108
Age-adjusted OR (95% CI)	1.0 (referent)	0.71 (0.50 to 0.99)
Multivariable OR (95% CI)†	1.0 (referent)	0.66 (0.45 to 0.95)

Chan, Andrew T., et al. *Journal of the National Cancer Institute* 97.6 (2005): 457-460.

Summary of Confounding and Interaction [Clinical Case]

- Confounding
  - **A phenomenon** in which background factors are biased in favor of a treatment group, **causing an appearance or disappearance of a deceptive association**.
  - It is a **type of bias** and should be eliminated as much as possible.
    - It can be eliminated by **randomization**. If this is not possible, multivariate analysis should be performed.
- Interaction
  - The **magnitude of the treatment effect differs among subgroups**.
    - The efficacy of gefitinib differs depending on the presence or absence of EGFR mutations.
  - Check to see if it exists because **it is not a bias**.
    - It cannot be eliminated by randomization.
  - Limitations on the subjects (scope) to whom the results can be applied.
    - Used for examining eligibility criteria and biological considerations.

Summary of Confounding and Interaction [Epidemiology Case]

- Confounding
  - **A phenomenon** in which background factors associated with an outcome bias between the exposure groups, **causing an appearance or disappearance of a deceptive association**.
  - It is a **type of bias** and should be eliminated as much as possible.
    - It can be eliminated by **randomization**. If this is not possible, multivariate analysis should be performed.
- Interaction
  - The **magnitude of the effect of exposure differs among subgroups**.
    - Aspirin's effect differs by UGT1A6 genotype
  - Check to see if it exists because **it is not a bias**.
    - It cannot be eliminated by randomization.
  - Limitations on the subjects (scope) to whom the results can be applied.
    - Used for the examination of subjects for prophylactic intervention and biological considerations.

How Do We Eliminate Confounding As Much As Possible Without Randomization?
– A Method to Increase the Comparability –

[Reposting] Background Factors

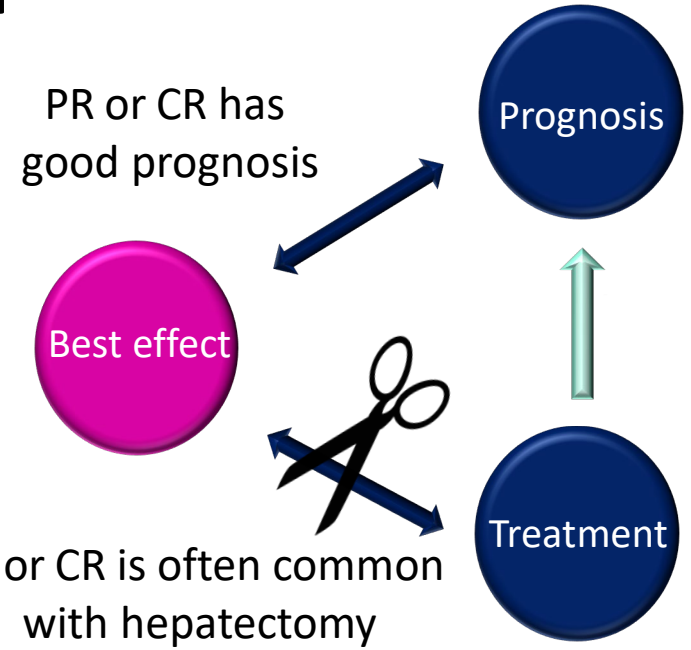
- Common in the hepatectomy group
  - Case of a response
  - Patients with a small maximum tumor diameter at the metastatic site
  - Patients with one organ with metastasis, etc.

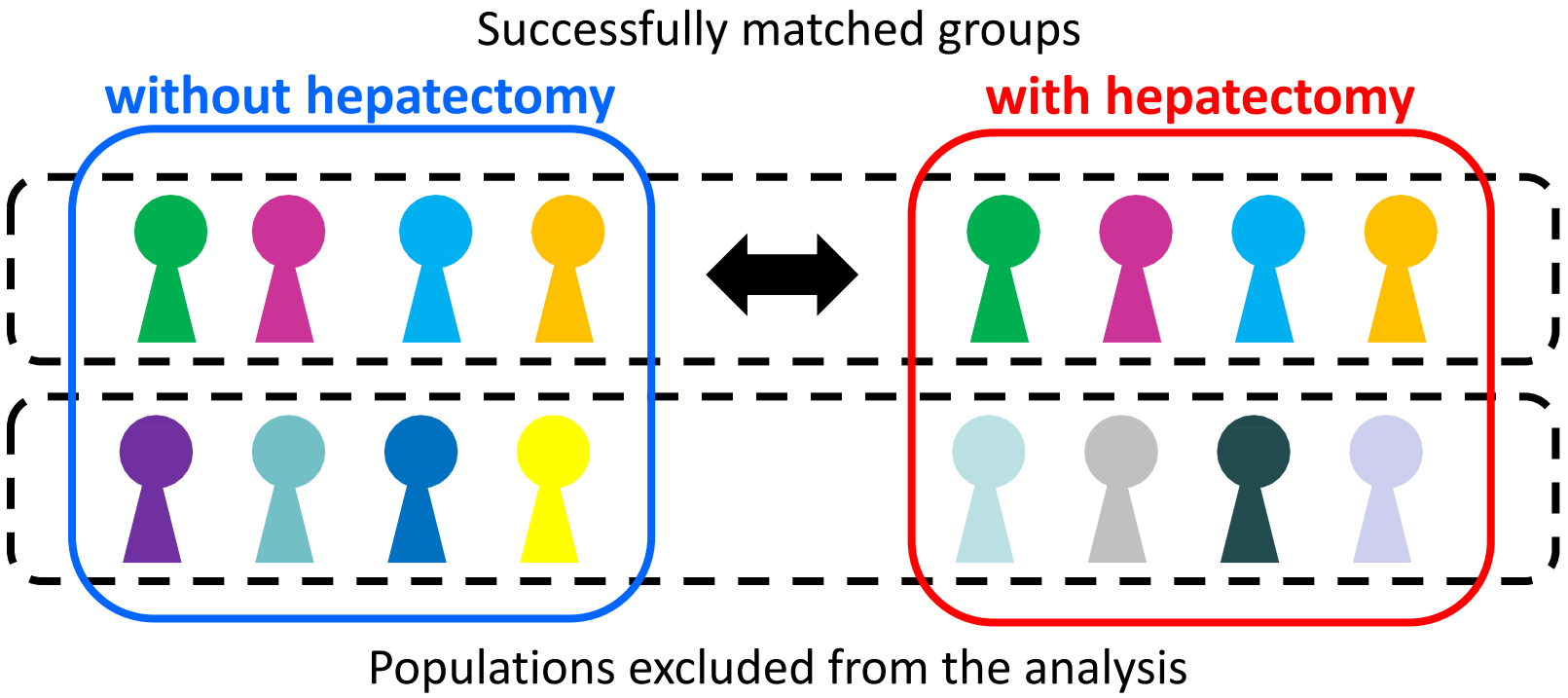
	With hepatectomy N = 77	Without hepatectomy N = 74
Age (years)		
Median (range)	59 (32–79)	58 (27–76)
Best effect (tumor shrinkage by chemotherapy)		
PD or no CX	0	**17**
SD	16	28
PR or CR	**61**	29
Maximum tumor diameter at metastatic site		
≤5 cm	**46**	33
>5 cm	24	**27**
Unknown	7	14
Number of organs with metastasis		
1	**14**	19
2–4	38	16
>4	20	**25**
Unknown	5	14

Giacchetti, S., et al. Ann Oncol. 10.6 (1999): 663-669.

How to address this during the design phase

- Matching
  - Match patients with matching background factors that are considered to have a strong impact on outcomes.





Methods to Increase Comparability at the Analysis Stage

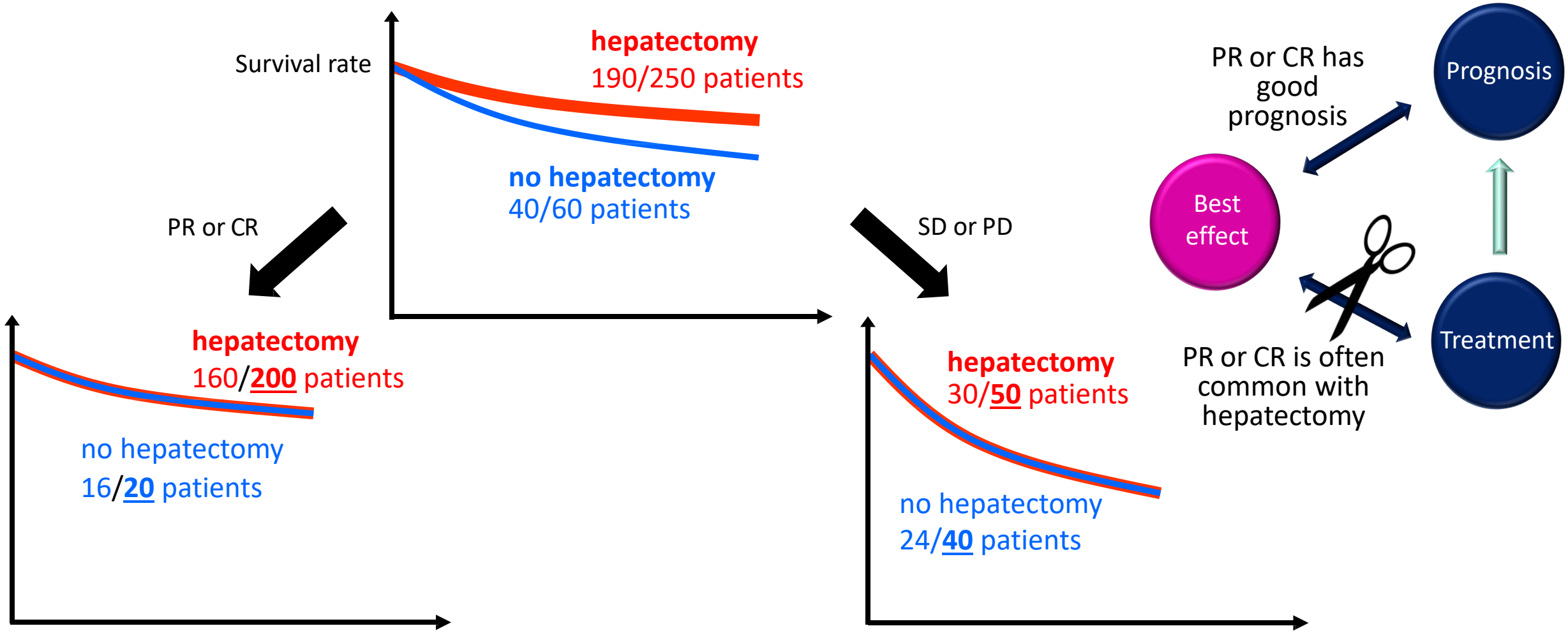
Topics this time!

- Subgroup analysis
  - Subset analysis, subpopulation analysis
  - Examine treatment effects for each subgroup
- Stratified analysis
  - Integrate (weighted average) the results for each subgroup, and determine one P value and one treatment effect
- Multivariate analysis using models
  - Perform Cox regression and logistic regression
- Analysis using propensity score (propensity score analysis)
  - Find and adjust the probability that a given patient will be assigned to a given treatment

Note: Hypothetical example

Subgroup Analysis

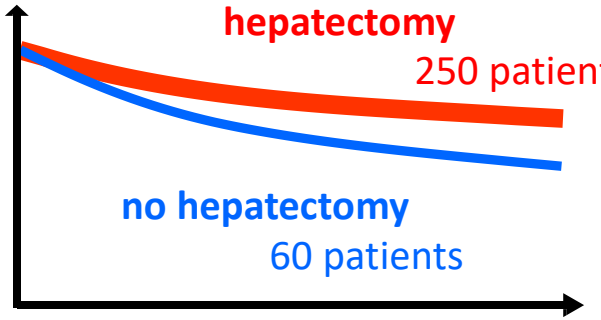
Examine the relationship between treatment and prognosis by best effect



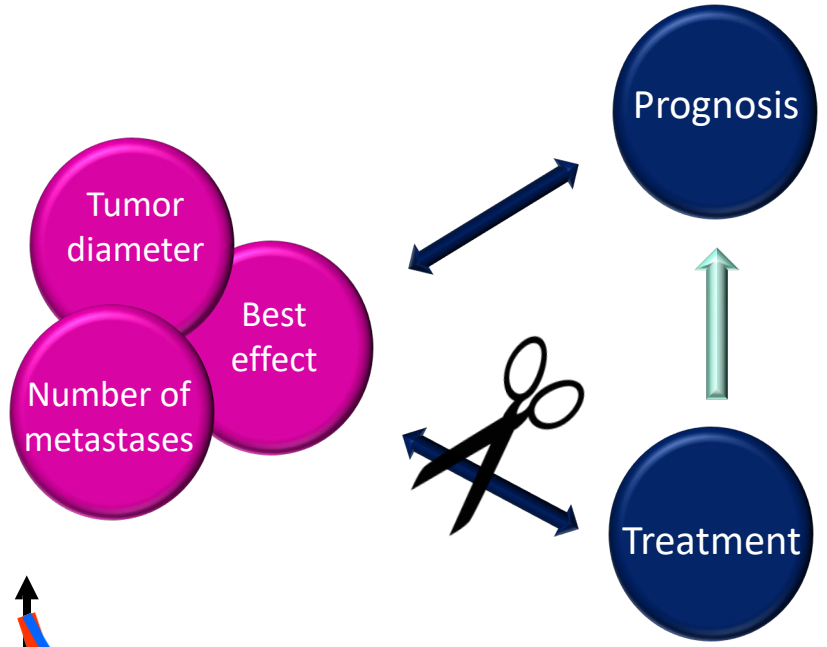
Note: Hypothetical example

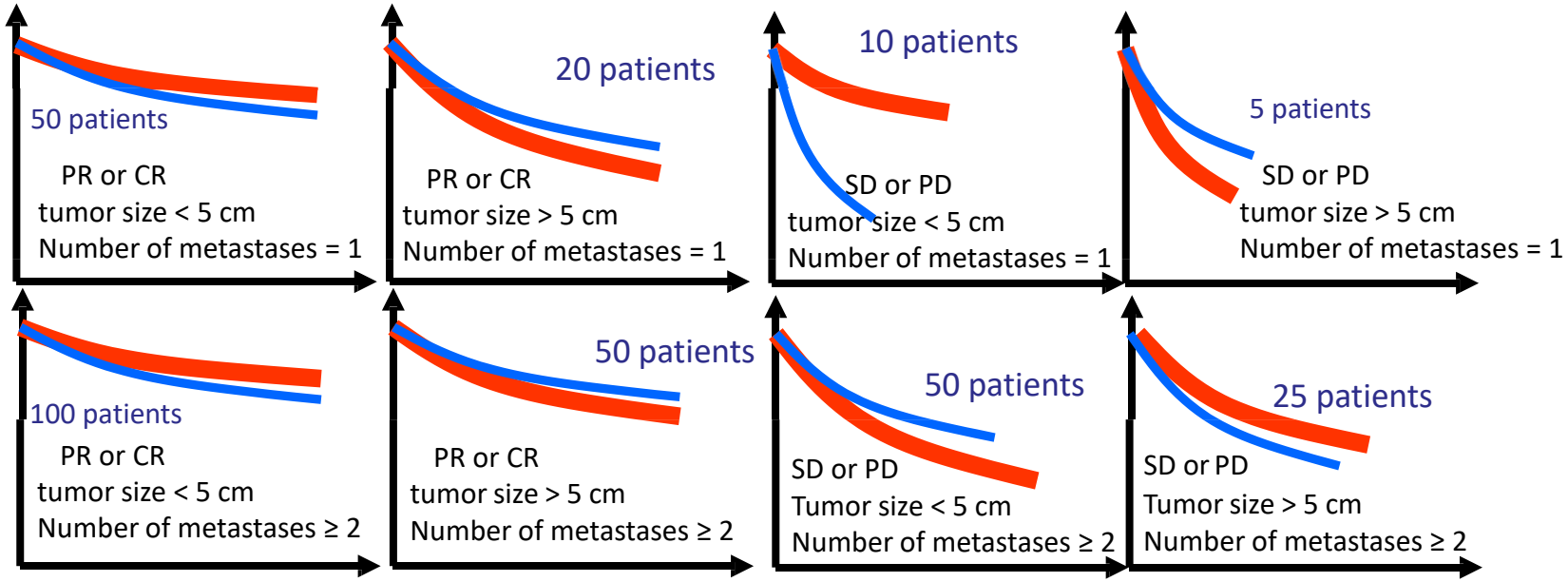
Drawbacks of Subgroup Analysis

To examine the relationship between treatment and prognosis by best effect, tumor size, and number of metastatic organs



The sample size for each subgroup becomes too small when there are multiple confounders!





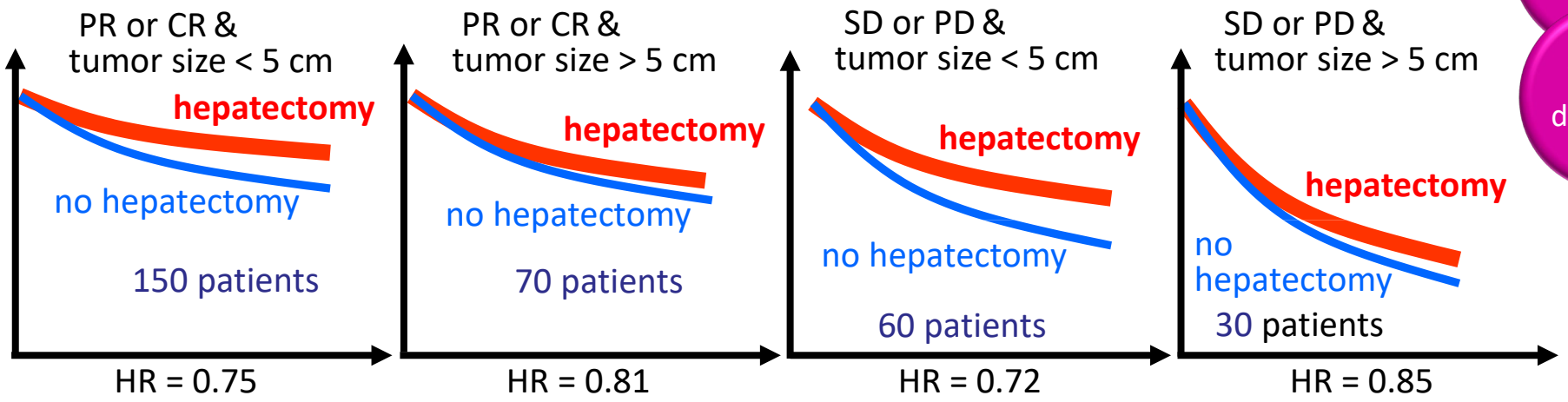
Advantages and Disadvantages of Subgroup Analysis

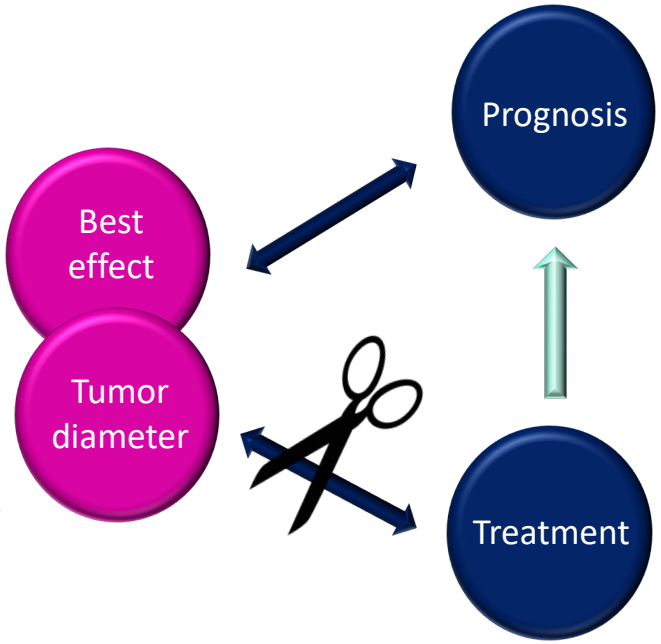
- Advantages
  - Easy and straightforward
    - We only need to examine treatment effects in each subgroup.
    - Fewer statistical assumptions
- Disadvantages
  - Cannot determine the effect of treatment on the population as a whole
  - If there are too many subgroups, the sample size for each subgroup is too small
    - If there are five confounders, at least $2^5 = 32$ subgroups
    - When confounders are continuous variables, subgroup analysis can only be performed after categorization
  - We do not know the magnitude of the effect of the confounders themselves (best effect is PR or CR for SD or PD)

Note: Hypothetical example

Startified Analysis

Integrating the treatment effect determined for each best effect and tumor size subgroups





Combine individual subgroups' treatment effects (HRs) into one weighted average

$$\text{Overall Effects } (HR) = 0.75 \times \frac{150}{310} + 0.81 \times \frac{70}{310} + 0.72 \times \frac{60}{310} + 0.85 \times \frac{30}{310} = 0.77$$

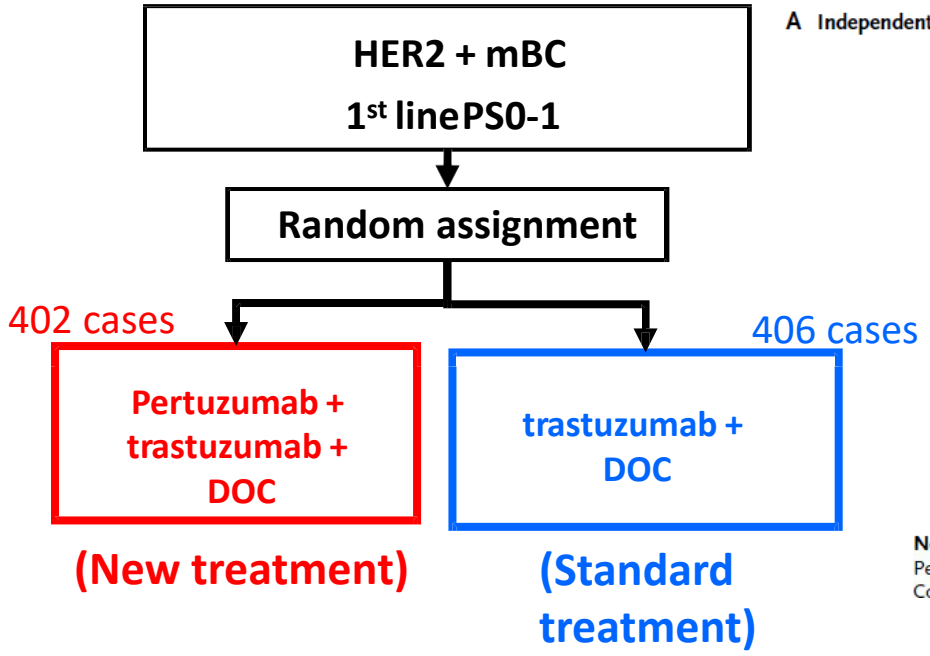
The weights include the sample size and the variability of the estimated values.

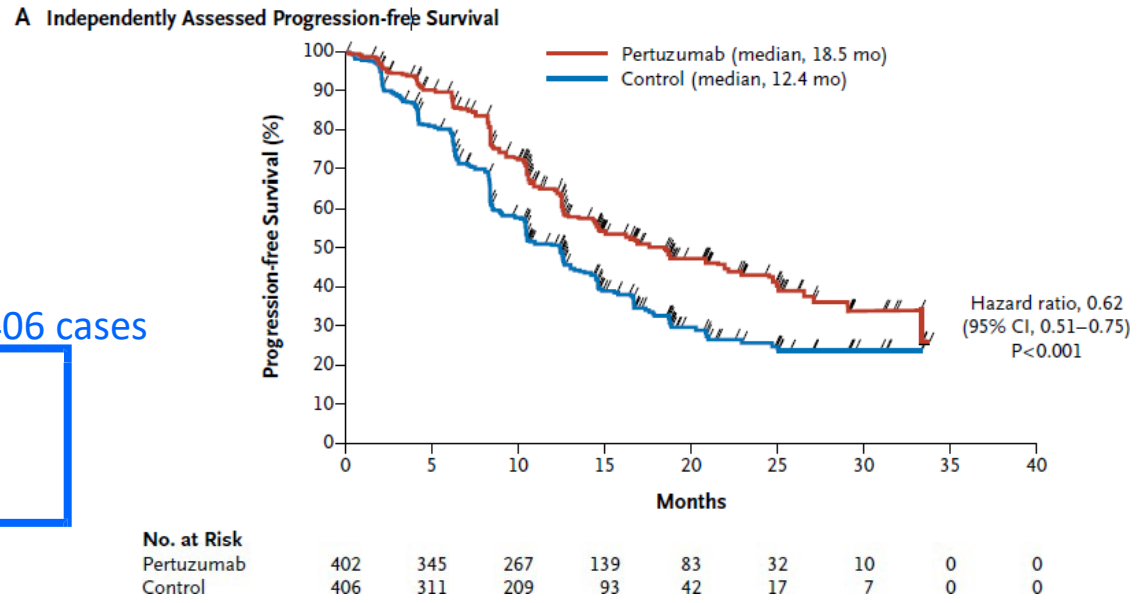
Example of Use in a Stratified Analysis Article or Literature

Baselga J, Cortés J, Kim S-B, et al:. N Engl J Med. 366:109-119, 2012

ORIGINAL ARTICLE

Pertuzumab plus Trastuzumab plus Docetaxel for Metastatic Breast Cancer





Statistical Statement on Stratified Analysis

Baselga J, Cortés J, Kim S-B, et al:. N Engl J Med. 366:109-119, 2012

tention-to-treat population (all patients who underwent randomization). The log-rank test, with stratification according to prior treatment status and region, was used to compare independently assessed progression-free survival between the two groups. The Kaplan–Meier approach was used to estimate the median independently assessed progression-free survival in each group. A Cox proportional-hazards model, with stratification according to prior treatment status and region, was used to estimate the hazard ratio and 95% confidence intervals.

free survival. The objective response rate was analyzed on data from patients who had independently assessed measurable disease at baseline and was compared between the groups with the use of the Mantel–Haenszel test, with stratification according to prior treatment status and region. Adverse events were evaluated descriptively

Previous treatment (no/yes), region (North America/Europe/South America/Asia)

↓

Stratified analysis combining the results of 8 subgroup analyses into one weighted average.

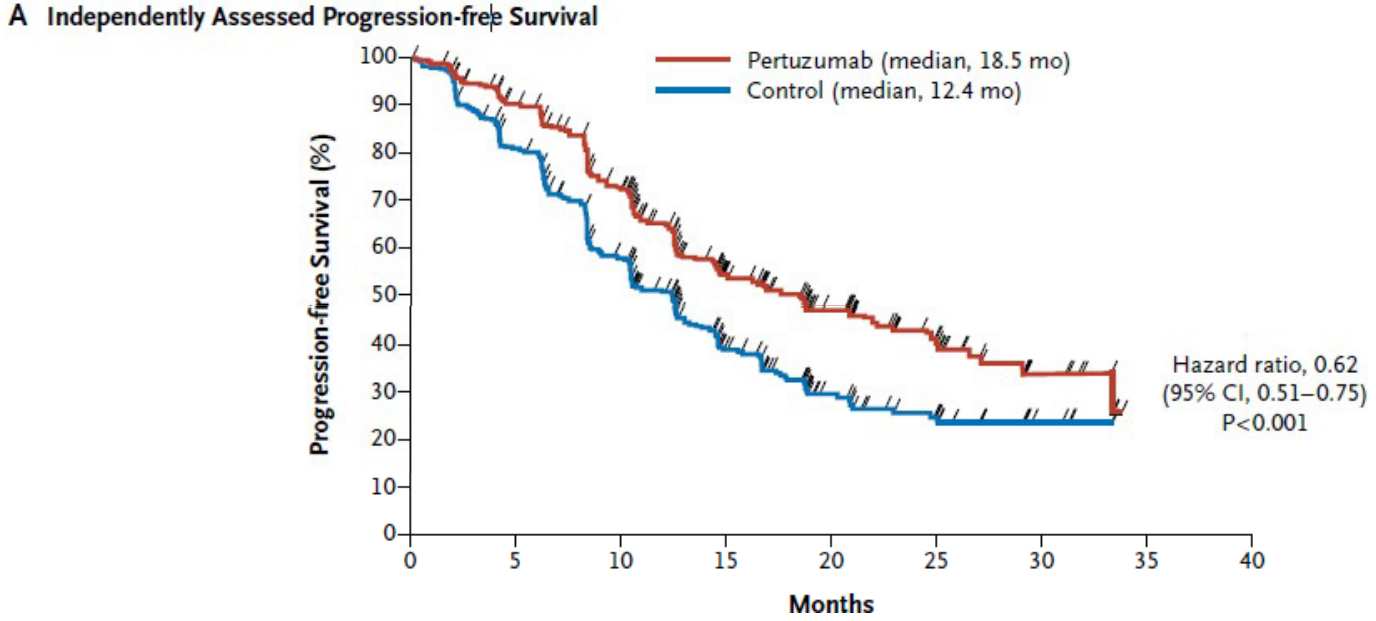
About PFS

- P value: stratified log-rank test
- HR and 95% CI: Stratified Cox regression

About response rate

- Mantel–Haenszel test

Confirmation of Stratified Analysis Results



Baselga J, Cortés J, Kim S-B, et al:. N Engl J Med. 366:109-119, 2012

- ***P* value: stratified log-rank test**
- **HR and 95%CI:**
  Stratified Cox regression

Table 2. Overall Response, as Assessed at an Independent Review Facility.*

Response	Placebo plus Trastuzumab plus Docetaxel (N=336)	Pertuzumab plus Trastuzumab plus Docetaxel (N=343)
	number (percent)	
Objective response	233 (69.3)	275 (80.2)
Complete response	14 (4.2)	19 (5.5)
Partial response	219 (65.2)	256 (74.6)
Stable disease	70 (20.8)	50 (14.6)
Progressive disease	28 (8.3)	13 (3.8)
Not assessable	2 (0.6)	2 (0.6)
No assessment performed	3 (0.9)	3 (0.9)

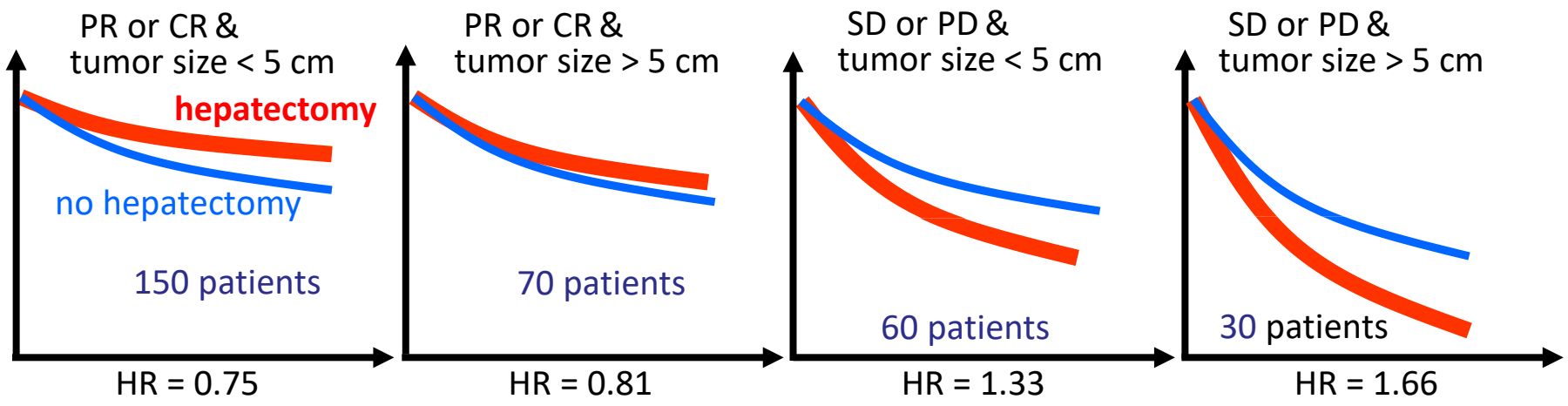
Mantel–Haenszel test

The difference in response rates was 10.8 percentage points (95% CI, 4.2 To 17.5; **P = 0.001**)

Note: Hypothetical example

Stratified Analysis Assumptions

Treatment effects appear comparable in each subgroup (No significant **interaction**)



$$\text{Overall Effects } (HR) = 0.75 \times \frac{150}{310} + 0.81 \times \frac{70}{310} + 1.33 \times \frac{60}{310} + 1.66 \times \frac{30}{310} = 0.96\ (??)$$

Advantages and Disadvantages of Stratified Analysis

- Advantages
  - The effect of treatment on the entire population can be determined
  - Fewer assumptions (Compared with analysis using models)
- Disadvantages
  - If there are too many subgroups, the sample size for each subgroup becomes too small
    - If there are five confounders, at least $2^5 = 32$ subgroups
    - When confounders are continuous variables, subgroup analysis can only be performed after categorization
  - The magnitude of the effect of the confounders themselves (PR or CR for the best effect of SD or PD) is unknown

Summary of Analytical Methods to Eliminate Confounding

Topics this time!

- Subgroup analysis
  - Subset analysis, subpopulation analysis
  - Examine treatment effects for each subgroup
- Stratified analysis
  - Integrate (weighted average) the results for each subgroup, and determine one P value and one treatment effect

Simple, but limited in the number of factors that can be handled simultaneously. Cannot handle continuous variable factors.

- Multivariate analysis using models
  - Perform Cox regression and logistic regression
- Analysis using propensity score (propensity score analysis)
  - Find and adjust the probability that a given patient will be assigned to a given treatment

Summary

- **Confounding** is a phenomenon in which background factors associated with prognosis are biased between treatment groups, causing an appearance or disappearance of apparent association
  - Confounding can be eliminated by **randomization**
- Different treatment effects among subgroups are called interactions
  - Interaction cannot be eliminated by **randomization**
- Methods to increase comparability without randomization (introduced today)
  - Matching
  - Subgroup analysis
  - Stratified analysis