

Non-Inferiority Trial 실전

KSIC Summer Conference 2023

Jinseob Kim (차라투)

6/24/23

Contents

1. 샘플수 계산
2. 비열등성 마진(margin)
3. 분석결과 제시

References

1. LOADSTAR(JAMA `23): YUHS
2. RACING(Lancet `22): YUHS
3. FLAVOUR(NEJM `22): SNU
4. RENOVATE-COMPLEX-PCI(NEJM `23): SMC

샘플 수 계산

Proportion vs Survival based approach

Proportion

- Event rate 만 이용.
- Racing, LOADSTAR, FLAVOUR

Survival

- Event rate, accrual/FU time 고려
- RENOVATE-COMPLEX-PCI

Racing

revealed primary endpoint rates of 32·7% and 34·7% in the simvastatin–ezetimibe and simvastatin groups, respectively, with a 6-year median follow-up duration, the primary endpoint rate in this study was presumed to be 20%, which was lower than in the same duration of the IMPROVE-IT trial because of our strategy to use a more potent statin.^{2,22} Therefore, the expected 3-year event rate was 13% in the combination therapy group and 14% in the high-intensity statin monotherapy group. A non-inferiority margin of 2·0 percentage points was primarily chosen because it was considered to be clinically not different between the two groups. A previous meta-analysis showed a 19% relative risk increase in the moderate-statin therapy versus high-intensity statin therapy for coronary death or any cardiovascular events (MI, stroke, hospitalisation for unstable angina, or any revascularisation) in similar patients.²³ Taking a conservative approach, 7·2% of relative risk increase (38% of the high-intensity statin effect relative to the moderate-intensity statin) in the moderate-intensity statin and ezetimibe combination therapy was thought to be clinically no difference, which corresponds to a 2·0 percentage-point difference between the groups in our study. On the basis of the non-inferiority hypothesis of a 2·0% margin, a total of 1605 patients

were required for each group considering a 5% one-sided alpha error rate and 80% power. Considering a 15% loss to follow-up, a total of 3780 patients were required to prove our hypothesis.

For the secondary objective, the achievement of LDL cholesterol of less than 70 mg/dL at 1 year was 31% in the simvastatin and 51% in the simvastatin–ezetimibe groups, in the IMPROVE-IT trial, and the LDL cholesterol-lowering effect of simvastatin–ezetimibe 40–10 mg was known to be similar to that of rosuvastatin 20 mg.^{22,24} Therefore, as a clinically important difference, the proportion of participants achieving LDL cholesterol of less than 70 mg/dL was presumed to be 70% in the combination therapy group and 50% in the high-intensity statin monotherapy group. On the basis of the superiority hypothesis with a 5% two-sided alpha error rate, 80% power, and an estimated 15% loss to follow-up, a total of 220 patients were required, which was fulfilled sufficiently by the sample size of 3780 patients for the primary objective of this trial.

Categorical data on demographic, medication, and procedural characteristics are described as numbers (percentages). Continuous variables are expressed as mean (SD) or median (IQR) for normal or skewed distributions. Kaplan-Meier curves for time-to-event

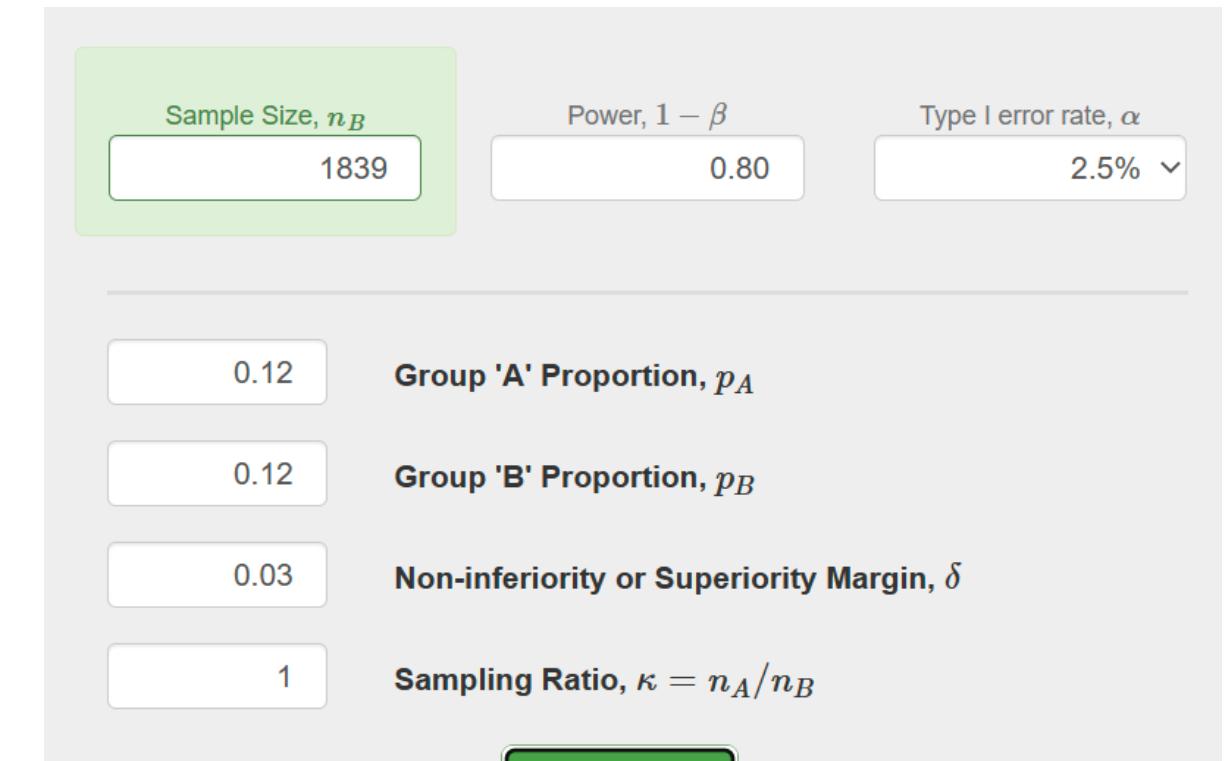
The screenshot shows a sample size calculator with the following input fields:

Parameter	Value
Sample Size, n_B	1603
Power, $1 - \beta$	0.80
Type I error rate, α	5%
Group 'A' Proportion, p_A	0.13
Group 'B' Proportion, p_B	0.14
Non-inferiority or Superiority Margin, δ	0.02
Sampling Ratio, $\kappa = n_A/n_B$	1

A green "Calculate" button is located at the bottom right of the form.

LOADSTAR

ized. Based on previous studies that compared different statin intensities in patients with CAD, the expected event rate of the primary end point was 4% per year in the high-intensity statin group.^{14,17} Assuming that the 2 strategies had equivalent efficacy, the expected event rate of the primary end point at 3 years was estimated to be 12% in each group. A noninferiority margin of 3.0 percentage points was primarily chosen with a consideration that this was not clinically different between the 2 groups. A total of 3686 patients was required, with a 2.5% 1-sided α error rate and 80% power. Considering a 15% loss to follow-up and balancing the 2 types of statins (rosuvastatin and atorvastatin), a total of 4400 patients (2200 patients in each group) was required.



FLAVOUR

The sample-size calculation was based on the assumption that the event rate for the primary outcome at 24 months would be 10% in the FFR group and 12% in the IVUS group.¹⁷⁻¹⁹ We determined that the enrollment of 1700 patients would provide 90% power to test the hypothesis that the event rate in the FFR group would be noninferior to that in the IVUS group according to a noninferiority margin of 2.5 percentage points, with the use of a one-sided 95% confidence interval and with a type I error rate of 5%. Although a one-sided type I error rate of 2.5% is considered to be robust for a noninferiority assessment, we used a one-sided error rate of 5% that is sometimes used for evaluating medical devices.^{20,21} We report results for the assessment of noninferiority on the basis of both a one-sided 95% confidence interval and a one-sided 97.5% confidence interval. Details regarding this power calculation are provided in the [Supplementary Appendix](#).

A screenshot of a sample size calculator interface. The top section shows three input fields: 'Sample Size, n_B ' with value '827' (highlighted in green), 'Power, $1 - \beta$ ' with value '0.9', and 'Type I error rate, α ' with value '5%' (with a dropdown arrow). Below these are four parameter inputs: 'Group 'A' Proportion, p_A ' with value '0.1', 'Group 'B' Proportion, p_B ' with value '0.12', 'Non-inferiority or Superiority Margin, δ ' with value '0.025', and 'Sampling Ratio, $\kappa = n_A/n_B$ ' with value '1'. At the bottom is a large green 'Calculate' button.

Input Parameter	Value
Sample Size, n_B	827
Power, $1 - \beta$	0.9
Type I error rate, α	5%
Group 'A' Proportion, p_A	0.1
Group 'B' Proportion, p_B	0.12
Non-inferiority or Superiority Margin, δ	0.025
Sampling Ratio, $\kappa = n_A/n_B$	1

RENOVATE-COMPLEX-PCI

STATISTICAL ANALYSIS

We estimated that a sample size of 1620 would provide the trial with at least 90% power, at a two-sided significance level of 5%, to reject the null hypothesis. The null hypothesis was that there would be no between-group difference for the primary composite end point as assessed by the log-rank test given an anticipated enrollment period of 3 years, follow-up of 1 year after the enrollment of the last patient, and withdrawal by 5% of the patients. The annual incidence of the primary end point was expected to be 3.6% in the intravascular imaging group and 6.0% in the angiography group. These estimates were based on the results of previous studies, as described in the Supplementary Appendix.^{3,5,9} No interim analysis was planned.

Inputs		Results																																											
Survival time (year or month)	1	Sample size of standard group	520																																										
Survival probability of standard group	0.94	Sample size of test group	1040																																										
Survival probability of test group	0.964	Total sample size	1560																																										
Allcation ratio	2	Expected event numbers of standard group	75																																										
Accrual time (year or month)	3	Expected event numbers of test group	92																																										
Follow-up time (year or month)	1	Total expected event numbers	167																																										
α	0.05	Actual power	0.901																																										
$1 - \beta$	0.9																																												
Method	log-rank test Gehan rank test Tarone-Ware rank test	Conditions																																											
Hypothesis	Two-sided One-sided	<input type="button" value="Submit"/> <input type="button" value="Reset"/>		Survival time	1			Survival probability of standard group	0.94			Survival probability of test group	0.964			HR (Test / Standard)	0.593			Allcation ratio	2			Accrual time	3			Follow-up time	1			α	0.05			$1 - \beta$	0.9			Method	Gehan rank test			Hypothesis	Two-sided
<input type="button" value="Submit"/> <input type="button" value="Reset"/>		Survival time	1																																										
		Survival probability of standard group	0.94																																										
		Survival probability of test group	0.964																																										
		HR (Test / Standard)	0.593																																										
		Allcation ratio	2																																										
		Accrual time	3																																										
		Follow-up time	1																																										
		α	0.05																																										
		$1 - \beta$	0.9																																										
		Method	Gehan rank test																																										
		Hypothesis	Two-sided																																										

Use gsdesign packages

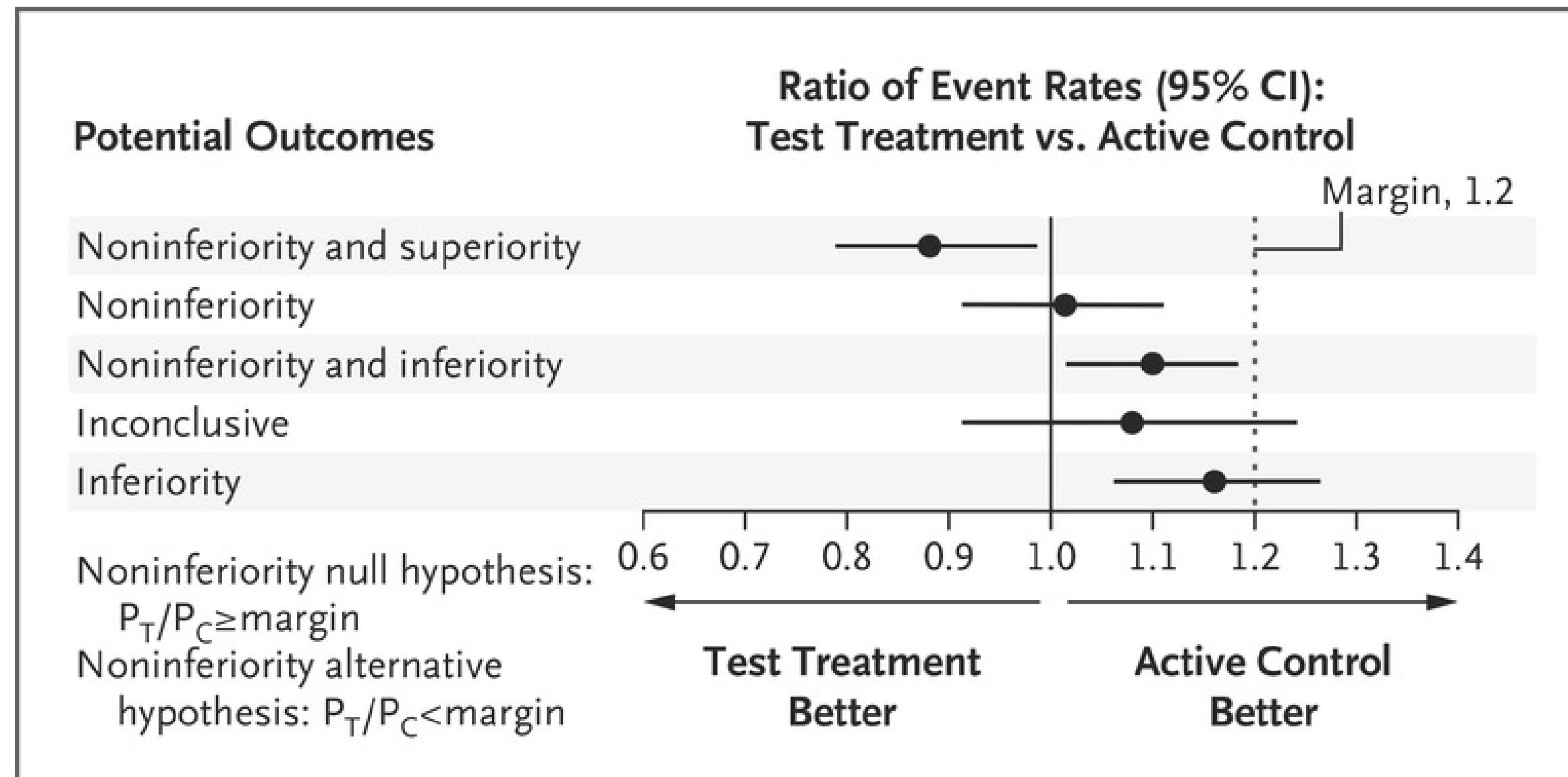
```
1 library(gsDesign)
2 ## Annual incidence
3 Pc <- 0.06; Pt <- 0.036
4 ## Hazard rate:  $S(t) = \exp(-\lambda * t)$ 
5 lambda_c <- -log(1 - Pc); lambda_t <- -log(1 - Pt)
6 hr <- lambda_t/lambda_c
7
8 nSurv(
9   lambdaC = lambda_c, # Hazard rate of control group
10  hr = hr,           # Alternative hypothesis
11  hr0 = 1,            # Null hypothesis
12  T = 4,              # Total study period
13  R = 3,              # Accrual period
14  minfup = 1,         # 4- 3
15  beta = 0.1,
16  alpha = 0.025,      # 1-sided
17  eta = 0,             # Annual drop out rate
18  ratio = 2)          # Randomization ratio, experimental/control
```

Fixed design, two-arm trial with time-to-event outcome (Lachin and Foulkes, 1986).
Solving for: Accrual rate
Hazard ratio H1/H0=0.5925/1
Study duration: T=4
Accrual duration: 3
Min. end-of-study follow-up: minfup=1
Expected events (total, H1): 165.1735
Expected sample size (total): 1609.307
Accrual rates:
Stratum 1
0-3 536.4356
Control event rates (H1):
Stratum 1
0-Inf 0.0619

비열등성 마진(margin)

Statistical & Clinical

- 차이가 없을 때 $p < 0.05$ 나와야 함.
- 임상적으로 허용 가능한 범위



<https://www.nejm.org/doi/full/10.1056/NEJMra1510063>

LOADSTAR 리비전

An unclear rational for a non-inferiority trial at all, i.e., **why is the targeted approach considered less burdensome and therefore advantageous if equally effective?**

- “target0| burden0이 작기때문에 non-inferiority를 증명하는 것만으로 충분하다는 논리”에 대한 의문.
- targeted approach 가 왜 burden0이 작냐는 질문에 답변 필요

Poor justification for the 3% non-inferiority margin, since that would correspond to a NNT of about 33, which would seem to justify the use of fixed high-intensity therapy if that were the true difference

- margin 3% 너무 큰거 아니냐?
- 실제로 3% 차이면 기존 fixed therapy가 더 나은것이라는 의견. 임상적 설명 필요.

분석결과 제시

Difference with upper 97.5(or 95)% CI

```
1 library(survival)
2 fit <- summary(survfit(Surv(time, status) ~ sex, data = colon), times = 365)
3
4 ## Surv2 - Surv1 = Inci1 - Inci2
5 kmdiff <- diff(1 - fit$surv)
6
7 ## Var(Surv1 + Surv2) = Var(Surv1) + Var(Surv2)
8 sediff <- sqrt(sum(fit$std.err^2))
9
10 c(Diff = kmdiff, LCI = kmdiff - 1.96 * sediff, UCI = kmdiff + 1.96 * sediff)
```

Diff	LCI	UCI
-0.024406273	-0.058041994	0.009229448

Upper limit < margin 이면 유의한 결과.

```
1 margin <- 3
2 pv <- 1 - pnorm(abs((margin - kmdiff)/sediff))
3 pv
[1] 0
```

KM estimate vs Proportion

KM 발생률을 대부분 썼으나 proportion 을 요구하는 경우도 있음.

- RACING: Kaplan-meier estimate 쓰지말고 그냥 proportion 써라.

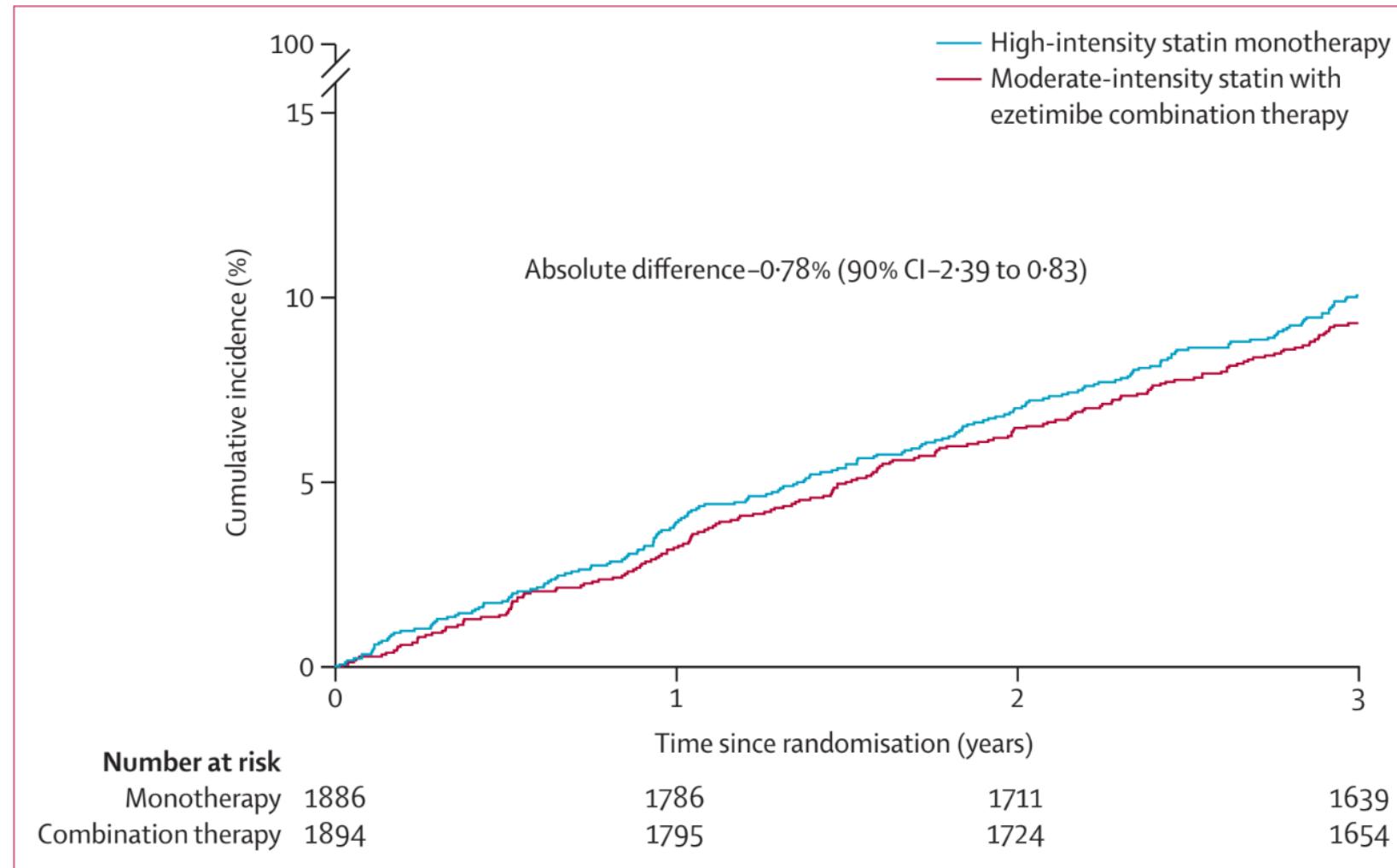


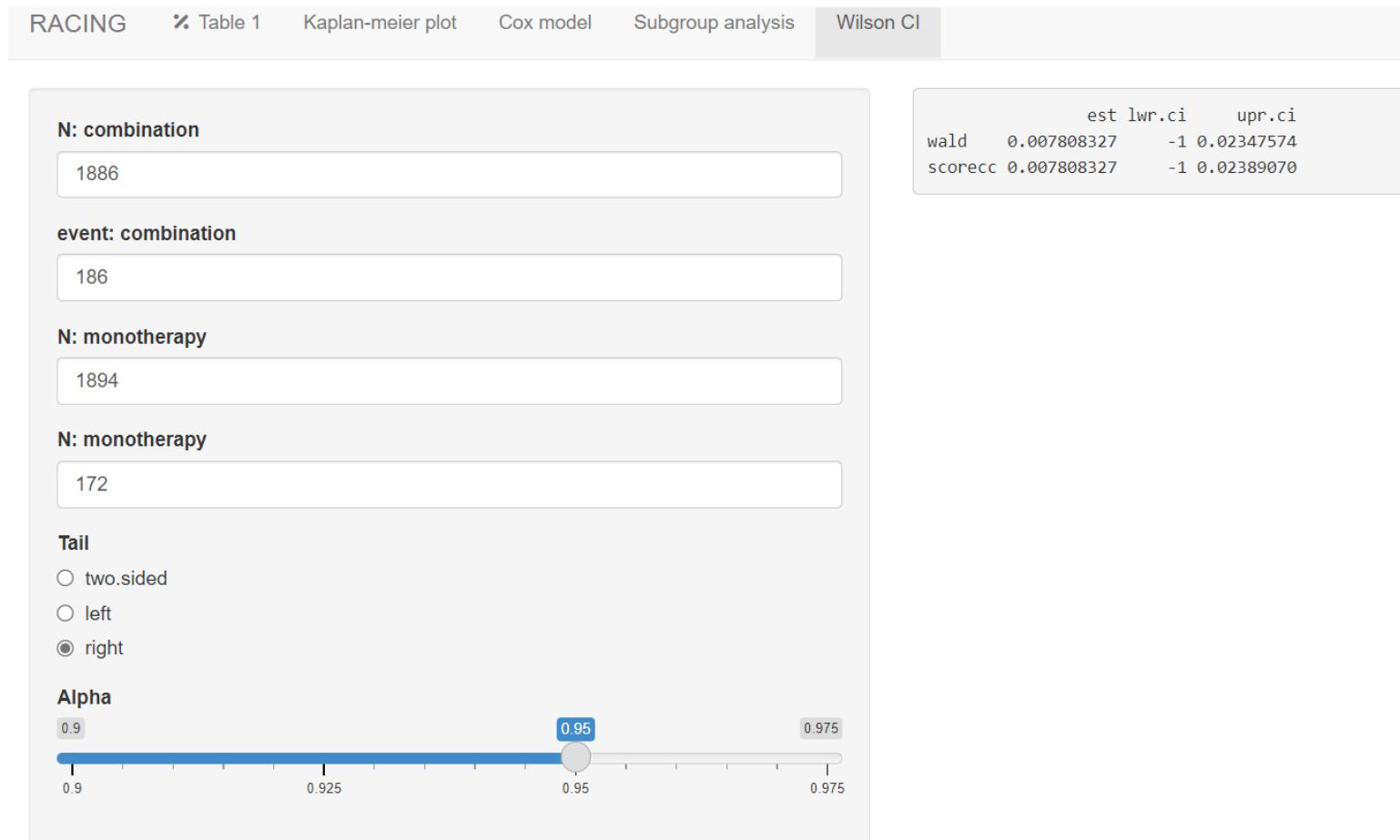
Figure 2: Kaplan-Meier curves of the primary endpoint of the intention-to-treat population

- KM: 9.3 vs 10.3% \rightarrow 9.1 vs 9.9% (Event/N)

CI for proportion difference

RACING: Wilson CI 써라

- For the primary and key secondary outcome I would question if the **Normal approximation does not hold.**
- 보수적인 지표



Primary outcome만 비열등성 검정

Table 2. Primary and Secondary End Points at 3 Years After Randomization^a

Outcome	Patients, No. (%)		Absolute difference, % (95% CI) ^b	P value
	Treat-to-target group (n = 2200)	High-intensity statin group (n = 2200)		
Primary end point				
Death, myocardial infarction, stroke, or coronary revascularization	177 (8.1)	190 (8.7)	-0.6 (-∞ to 1.1) ^c	<.001 ^d
Components of primary end point				
Death	54 (2.5)	54 (2.5)	<0.1 (-0.9 to 0.9)	.99
Cardiac death	16	13		
Myocardial infarction	34 (1.6)	26 (1.2)	0.4 (-0.3 to 1.1)	.23
Stroke	17 (0.8)	27 (1.3)	-0.5 (-1.1 to 0.1)	.13
Ischemic	12	20		
Hemorrhagic	5	7		
Coronary revascularization ^e	112 (5.2)	114 (5.3)	-0.1 (-1.4 to 1.2)	.89
Secondary end points				
New-onset diabetes	121 (5.6)	150 (7.0)	-1.3 (-2.8 to 0.1)	.07
Initiation of antidiabetic medication	73	105		
Cataract operation	43 (2.0)	42 (1.9)	0.1 (-0.8 to 0.9)	.90
Discontinuation of statin therapy	31 (1.5)	46 (2.2)	-0.7 (-1.5 to 0.1)	.09
Composite of laboratory abnormalities ^f	18 (0.8)	30 (1.3)	-0.5 (-1.1 to 0.1)	.11
Aminotransferase elevation	8	12		
Creatine kinase elevation	3	8		
Creatinine elevation	7	11		
Peripheral artery revascularization	12 (0.6)	17 (0.8)	-0.2 (-0.8 to 0.3)	.35
Hospitalization due to heart failure	13 (0.6)	7 (0.3)	0.3 (-0.1 to 0.7)	.17
End-stage kidney disease	3 (0.1)	10 (0.5)	-0.3 (-0.7 to 0.0)	.05
Deep vein thrombosis or pulmonary embolism	4 (0.2)	5 (0.2)	<0.1 (-0.3 to 0.2)	.74
Deep vein thrombosis	2	5		
Pulmonary embolism	3	0		
Aortic intervention or surgery	2 (0.1)	3 (0.1)	NR	
Endovascular therapy	1	2		
Surgical therapy	1	1		
Composite of new-onset diabetes, aminotransferase or creatine kinase elevation, or end-stage kidney disease (post hoc)	132 (6.1)	177 (8.2)	-2.1 (-3.6 to -0.5)	.009

Abbreviation: NR, not reported.

^a Primary and secondary end points were evaluated as randomized 3 years after randomization. The listed percentages were estimated using the Kaplan-Meier method, so the values might not calculate mathematically. Differences in event rates are not reported for aortic intervention because of the low numbers of events.

^b The between-group difference was measured in the treat-to-target group compared with the high-intensity statin group. The widths of the confidence intervals have not been adjusted for multiplicity and cannot be used to infer treatment effects.

^c A 1-sided 97.5% CI was calculated for the primary end point.

^d The P value for noninferiority is for an upper boundary of the 97.5% CI of the

between-group difference in the primary end point, which was 1.1 percentage points. Other P values were 2-sided.

^e All coronary revascularizations were clinically indicated by an invasive angiographic percent diameter stenosis of 50% or greater with ischemic symptoms or signs or 70% or greater even without symptoms or signs.

^f Aminotransferase elevation was defined as greater than baseline level and more than 3 times the upper limit of reference. Creatine kinase elevation was defined as greater than baseline level and more than 5 times the upper limit of reference. Creatinine level elevation was defined as greater than 50% increase from baseline and greater than the upper limit of reference. Reference values may vary based on laboratory and location.

Analyses of secondary endpoints were not adjusted for multiplicity, and findings should be interpreted as exploratory because of the potential for type I error

LOADSTAR 리비전

Some lack of clarity regarding the exact statistical test used to establish the confidence interval around the difference in event rate

- Kaplan meier 방법으로 두 군의 Incidence 값과 standard error 를 각각 얻습니다.
- 이것을 이용해 difference 값과 upper 97.5% 신뢰구간을 구합니다.
- 본 연구에서 Cox 분석은 없습니다.

Uncertainty regarding the responsiveness of the endpoint over the duration of follow up.

- 3년 F/U 기간 이후엔 어떠냐. 더 오래 관찰해야 할 수도 있다는 의견.
- 3년 이후 outcome이 혹시 있다면 보여주고, 아니면 3년 관찰로 충분히 의미 있다고 설명.

HR 필요?

LOADSTAR: HR 지표가 아예 없음. Rate difference 만 보여줌.

- 따라서 Cox 분석도 없음. **simple is the best**

RENOVATE-COMPLEX-PCI: 반대로 고급분석요구

- Stratified Cox: 각 병원의 고유 특성을 고려
- Competing risk analysis: Other cause death 고려

RENOVATE-COMPLEX-PCI 리비전

1. P10 L53. Earlier you stated that the randomization was stratified by clinical presentation and by treatment center. Did you apply a stratified logrank test to incorporate this design feature, and include these covariates in the Cox model estimation? Please clarify.

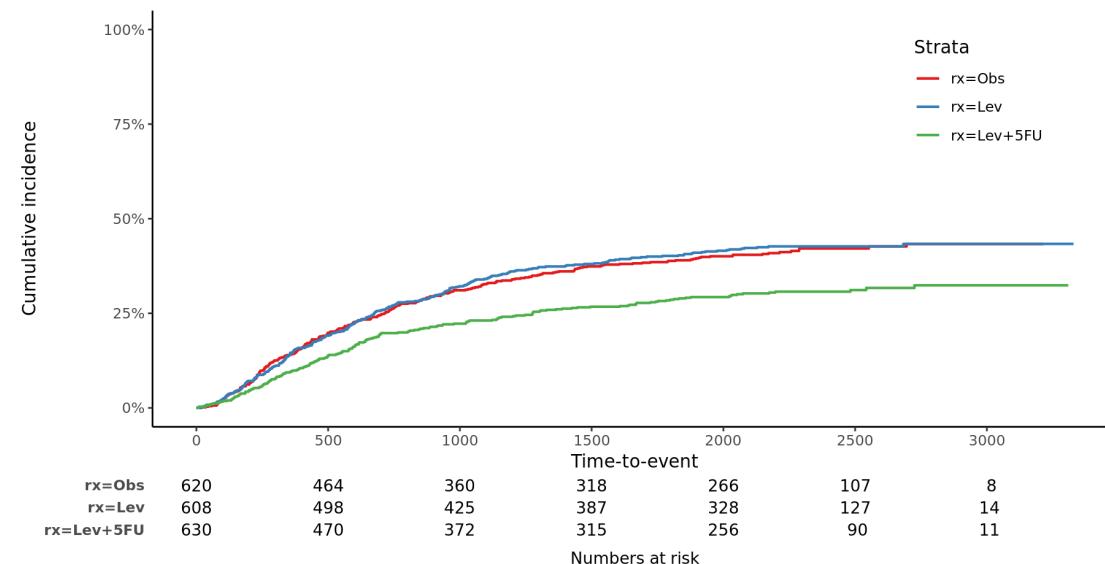
- 다기관 고려한 stratified 분석해라.
- R `coxph` 식에 **+strata(기관변수)** 추가

```
1 coxph(Surv(time, status) ~ group + strata(hospital), data = data)
```

RENOVATE-COMPLEX-PCI 리비전

3. P13 L36-39. The analyses that properly account for competing risks should be the primary analyses here; the standard K-M or Cox model estimates that ignore competing risks are incorrect. Please describe the method used to accommodate competing risks in the methods section, and present these results as the primary analysis.

-> Competing risk analysis 를 primary로 해라



```
1 library(jskm);library(survival)
2 colon$status2 <- colon$status
3 colon$status2[1:400] <- 2
4 colon$status2 <- factor(colon$status2)
5 fit2 <- survfit(Surv(time,status2)~rx, data=colon)
6 jskm(fit2, mark = F, surv.scale = "percent", table = T, status.cmp
```

RENOVATE-COMPLEX-PCI 리비전

2. P13 L32. What was the “per protocol” analysis? This is the first time this is mentioned; it should be explained in the methods section. Please also note that per protocol analyses that are based on an analysis dataset constructed by **eliminating cases based on post-randomization events (e.g., treatment crossover)** are generally biased and are not allowed. More principled methods of analyzing a trial subject to non-adherence should be used. See for instance Hernan and Robins, NEJM 2017; 377: 1391-1398.

-> Per-protocol 하면서 연구디자인 깨지는거 아니냐?

Per-protocol: Image군에서 Angio, Angio군에서 Image를 한 환자를 Protocol violation이라고 미리 정의.

PP scenario

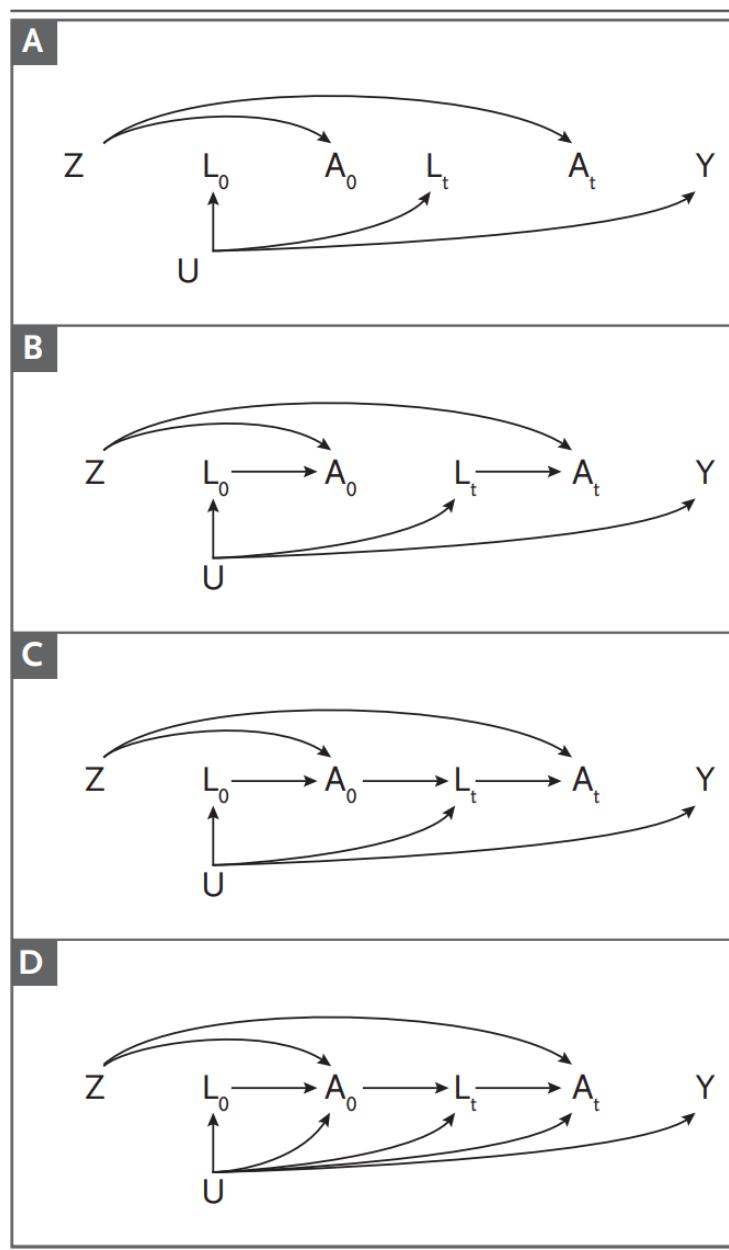


Figure 2. Four Possible Scenarios in a Randomized Trial of Sustained Treatment Strategies.

Shown are four possible scenarios in a randomized trial of sustained treatment strategies, according to the determinants of adherence to the treatment strategies. Z is an indicator for randomization group, Y is the outcome of interest, L_0 represents prognostic factors measured at baseline (time 0), L_t represents prognostic factors measured after baseline (at time t), A_0 is an indicator for adherence to the protocol at baseline, A_t is an indicator for adherence to the protocol at time t , and U represents unmeasured baseline or postbaseline factors. To avoid clutter, we show only the first two time points, assume no losses to follow-up, and assume a null comparative effect of the treatment strategies. In scenario 1 (Panel A), adherence occurs at random (no arrows from any variable to A). The per-protocol analysis does not have to adjust for any factors. In scenario 2 (Panel B), adherence depends only on the measured factors (arrows from L to A). The per-protocol analysis has to adjust for these factors. In scenario 3 (Panel C), adherence depends only on the measured factors and the measured factors affect future adherence (arrows from L to A and from A to L). The per-protocol analysis has to adjust for these factors; the adjustment for the postbaseline factors requires g-methods. In scenario 4 (Panel D), adherence depends on both measured and unmeasured factors (arrows from L and U to A). A valid per-protocol analysis should use a form of instrumental variable estimation based on g-estimation of a structural nested model, which relies on strong assumptions.

A 는 OK, $B \rightarrow C \rightarrow D$ 로 갈수록 복잡한 보정 필요.

- A 에 해당함을 주장.

감사합니다

