2024 Summer Institute In Statistics for Clinical & Epidemiological Research

Module 3:

Design, Conduct, and Analysis of Randomized Clinical Trials with Time to Event Primary Endpoints

Lecture 4:

ICH E9 (R1) strategies for intercurrent events

Scott S. Emerson, M.D., Ph.D.
Professor Emeritus of Biostatistics
University of Washington

1

1

RCT Settings Using Time to Event Outcomes

- Overall goal: Drug discovery
- **Estimands**
  - Clinical
  - RCT
  - **ICH E9 (R1) strategies for intercurrent events**
- Why an “event”? Why ”time to event”?
- Why incomplete observations: Informative vs noninformative?
  - Administrative censoring
  - Competing risks
  - Intercurrent events and protopathic events
  - Loss to follow-up
- How to define “tends to be”?
  - Choice of summary measure

2

2

Estimands

ICH E9 (R1) Strategies for Intercurrent Events

Where am I going?

The International Conference on Harmonization issued an E9 amendment to address the estimand framework.

The amendment was adopted by both EMA and FDA.

The implementation of this amendment in clinical trials by sponsors is coming under some criticism.

3

3

ICH E9 R(1)

- **Intercurrent Events:**
  - Events occurring after treatment initiation that affect either the interpretation or the existence of the measurements associated with the clinical question of interest. It is necessary to address intercurrent events when describing the clinical question of interest in order to precisely define the treatment effect that is to be estimated.

- **Missing Data:**
  - Data that would be meaningful for the analysis of a given estimand but were not collected. They should be distinguished from data that do not exist or data that are not considered meaningful because of an intercurrent event.

4

4

Intercurrent Events

- Estimand: Clinical effect of treatment
  - Disease, population, treatment regimen, outcome
- Intercurrent-event
  - Post-randomization events that might complicate initial definition of outcome
- ICH E-9 Addendum tries to draw a distinction between missing data and intercurrent events
  - I beg to differ
  - If you cannot estimate your estimand based on ITT, then you must use methods appropriate for missing data

5

ICH E9 (R1) Strategies

- Examples of intercurrent events
  - Competing risks
    - Make measurement impossible: Death from irrelevant causes
    - Make measurement irrelevant: Organ transplant
  - Advancing to other treatments
    - "Rescue treatments" that we try to avoid: opioids, steroids
    - Treatments with progression that confuses safety, short term efficacy measurements
  - Patient withdrawal of consent
  - Loss to follow-up
  - Discontinuation of treatment
- ICH E9 (R1) envisions that a "strategy" of handling each such intercurrent event will be described in protocol for analysis of primary outcome

6

6

ICH E9 (R1) Strategies

- Treatment Policy (per randomization, ITT)
  - Intercurrent events are generally irrelevant
- Hypothetical strategies
  - Imagine intercurrent event would not occur
- Composite endpoint strategies
  - Incorporate intercurrent event as part of outcome
- While on treatment strategies
  - Only incorporate experience prior to intercurrent event
- Principal stratum strategies

7

7

ICH E9 R(1)

- **Principal Stratification:**
  - Classification of subjects according to the potential occurrence of an intercurrent event on all treatments. With two treatments, there are four principal strata with respect to a given intercurrent event: subjects who would not experience the event on either treatment, subjects who would experience the event on treatment A but not B, subjects who would experience the event on treatment B but not A, and subjects who would experience the event on both treatments. In this document a **principal stratum** refers to any of the strata (or combination of strata) defined by principal stratification.

8

8

ICH E-9: Modification of Estimands

- Treatment policy strategy (ITT)
  - Per randomization with similar follow-up on all individuals
  - "cannot be implemented for intercurrent events that are terminal"
- In the spirit of the treatment policy strategy, terminal intercurrent events could instead be handled by
  - Hypothetical:
    - Impute data as if the terminal event did not occur
      - MAR vs MNAR
  - Composite:
    - Incorporate the death as one of the events in the endpoint, or
    - Assign "worst" score to the patients with a terminal event
      - QoL is 0
      - Liver, kidney, lung function is 0
      - But: not SBP or HbA1c of 0 when treating HTN or DM

9

ICH E-9: Modification of Estimands

- Hypothetical strategies
  - "scenario … in which the intercurrent event would not occur"
    - Availability of alternative treatments vs restricted subpopulation
  - "clinical and regulatory interest of such hypotheticals is limited"
- I have often seen protocols invoke hypothetical strategies
  - Unwittingly: "censored in the time to event analysis"
  - Wittingly: "A hypothetical strategy will be used to impute data"
    - Invariably no discussion of the causal justification of the strategy
      - Force patients to take the treatment?
      - Never enroll subjects who could / would not take treatment?
      - We imagine advances in ancillary treatments?
- Different imputation strategies might be indicated
  - Sensitivity analyses would also need to be prominent

10

ICH E-9: Modification of Estimands

- Composite variable strategies with ITT
  - E.g., intercurrent event is assigned
    - treatment failure, or
    - worst case outcome
  - Examples: progression to opioids, steroids, transplant
- Issue: Comparability of contributing events:
  - Death vs therapy d/c
- Composite endpoints widely used
  - Progression free survival in cancer
  - Major adverse cardiovascular events in cardiovascular outcome trials

11

Composite Endpoints in Obesity CVOT

Semaglutide and cardiovascular outcomes in obesity without diabetes

A. Michael Lincoff, M.D.,[1] Kirstine Brown-Frandsen, M.D.,[2] Helen M. Colhoun, M.D.,[3] John Deanfield, M.D.,[4] Scott S. Emerson, M.D., Ph.D.,[5] Sille Esbjerg, M.Sc.,[2] Søren Hardt-Lindberg, M.D., Ph.D.,[2] G. Kees Hovingh, M.D., Ph.D.,[2,6] Steven E. Kahn, M.B., Ch.B.,[7] Robert F. Kushner, M.D.,[8] Ildiko Lingvay, M.D., M.P.H., M.S.C.S.,[9] Tugce K. Oral, M.D.,[2] Marie M. Michelsen, M.D., Ph.D.,[2] Jorge Plutzky, M.D.,[10] Christoffer W. Tornøe, M.Sc., Ph.D.,[2] Donna H. Ryan, M.D.,[11] for the SELECT Trial Investigators

[1]Department of Cardiovascular Medicine, Cleveland Clinic, Cleveland Clinic Lerner College of Medicine of Case Western Reserve University, Cleveland, OH, US
[2]Novo Nordisk A/S, Søborg, Denmark
[3]Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK
[4]National Institute for Cardiovascular Outcomes Research, University College London, London, UK
[5]Department of Biostatistics, University of Washington, Seattle, WA, US
[6]Department of Vascular Medicine, Academic Medical Center, Amsterdam, the Netherlands
[7]Department of Medicine, VA Puget Sound Health Care System and University of Washington, Seattle, WA, US
[8]Department of Medicine, Northwestern University Feinberg School of Medicine, Northwestern University, Chicago, IL, US
[9]Department of Internal Medicine/Endocrinology and Peter O'Donnel Jr School of Public Health, UT Southwestern Medical Center, Dallas, TX, US
[10]Department of Cardiovascular Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, US
[11]Pennington Biomedical Research Center, Baton Rouge, LA, US
This trial was funded by Novo Nordisk and is registered with ClinicalTrials.gov (NCT03574597). Lucy Ambrose, DPhil, and Casey McKeown, RVN, FdSc, of Apollo, OPEN Health Communications, provided administrative and editorial support, including development of the figures (funded by Novo Nordisk).
Lincoff AM, et al. N Engl J Med 2023;doi:10.1056/NEJMoa2307563.

12

Composite Endpoints in Obesity CVOT

SELECT trial clinical endpoints*

Primary endpoint

Time-to-first-event analysis: first occurrence of any component of a composite of:

- Death from CV causes
- Non-fatal MI
- Non-fatal stroke

Supportive secondary endpoints

Time-to-first-event analyses without multiplicity of:

- Expanded CV composite endpoints
- Individual components of CV composite endpoints
- Composite nephropathy endpoint
- Progression to diabetes or pre-diabetes

Additional endpoints

Change from randomization to week 104:

- Body weight
- Waist circumference
- PROs
- HbA_{1c}
- Heart rate
- Blood pressure
- High-sensitivity CRP and lipid levels

Confirmatory secondary endpoints

Time-to-first-event analyses, tested in hierarchical order:

1. Death from CV causes
2. Composite HF endpoint (death from CV causes or HF hospitalization or urgent HF visit)
3. Death from any cause

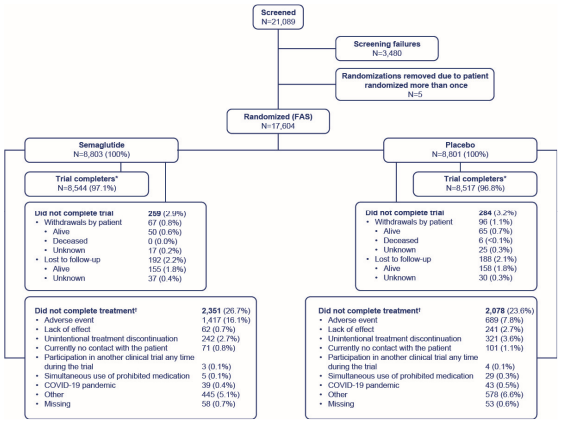
*An independent event adjudication committee, whose members were unaware of the trial-group assignments, adjudicated cause-of-death and cardiovascular endpoints and events related to kidney transplantation, initiation of chronic renal replacement therapy, and acute pancreatitis.
CRP, C-reactive protein; CV, cardiovascular; HbA_{1c}, glycated hemoglobin; HF, heart failure; MI, myocardial infarction; PRO, patient-reported outcome.

Lincoff AM, et al. N Engl J Med 2023;doi:10.1056/NEJMoa2307563.

13

Composite Endpoints in Obesity CVOT

Figure S1. Flow of Patients Through the Trial.



*Patients who attended the follow-up visit or who died during the trial.

†Primary reason for not completing treatment, according to the Dose Change form. Treatment discontinuations that occurred less than 30 days before the end-of-treatment visit were not counted.

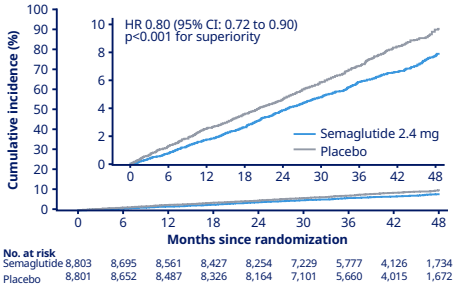
FAS, full analysis set.

14

Composite Endpoints in Obesity CVOT

Primary cardiovascular composite efficacy endpoint*

Time-to-first-event analysis of the primary endpoint



No. at risk								
Semaglutide	8,803	8,695	8,561	8,427	8,254	7,229	5,777	4,126	1,734
Placebo	8,801	8,652	8,487	8,326	8,164	7,101	5,660	4,015	1,672

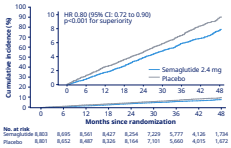
*Defined as death from cardiovascular causes, nonfatal myocardial infarction, or nonfatal stroke.
The cumulative incidences were estimated using the Aalen–Johansen method accounting for competing risk,[1] and the hazard ratios were estimated with the Cox proportional hazards regression model. The insets show the same data on an enlarged y axis. The x axis is truncated at 48 months due to the limited number of patients in the trial after 48 months.
1. Andersen PK, et al. Statistical Models Based on Counting Processes. 1 ed. New York, NY: Springer; 1993.
CI, confidence interval; HR, hazard ratio.
Adapted from Figure 1. Time-to-Event Analysis for Primary and Confirmatory Secondary Efficacy Endpoints.
Lincoff AM, et al. N Engl J Med 2023;doi:10.1056/NEJMoa2307563.

15

Composite Endpoints in Obesity CVOT

Primary cardiovascular composite efficacy endpoint*

Time-to-first-event analysis of the primary endpoint

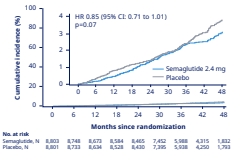


No. at risk								
Semaglutide	8,803	8,695	8,561	8,427	8,254	7,229	5,777	4,126	1,734
Placebo	8,801	8,652	8,487	8,326	8,164	7,101	5,660	4,015	1,672

*Defined as death from cardiovascular causes, nonfatal myocardial infarction, or nonfatal stroke.
The cumulative incidences were estimated using the Aalen–Johansen method accounting for competing risk,[1] and the hazard ratios were estimated with the Cox proportional hazards regression model. The insets show the same data on an enlarged y axis. The x axis is truncated at 48 months due to the limited number of patients in the trial after 48 months.
1. Andersen PK, et al. Statistical Models Based on Counting Processes. 1 ed. New York, NY: Springer; 1993.
Adapted from Figure 1. Time-to-Event Analysis for Primary and Confirmatory Secondary Efficacy Endpoints.
Lincoff AM, et al. N Engl J Med 2023;doi:10.1056/NEJMoa2307563.

Death from cardiovascular causes

Time-to-first-event analysis of confirmatory secondary endpoint

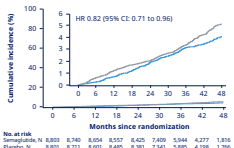


No. at risk								
Semaglutide, N	8,803	8,748	8,673	8,584	8,465	7,452	5,988	4,315	1,832
Placebo, N	8,801	8,733	8,634	8,528	8,430	7,395	5,938	4,250	1,793

The cumulative incidences were estimated using the Aalen–Johansen method accounting for competing risk,[1] and the hazard ratios were estimated with the Cox proportional hazards regression model. The insets show the same data on an enlarged y axis. The x axis is truncated at 48 months due to the limited number of patients in the trial after 48 months.
1. Andersen PK, et al. Statistical Models Based on Counting Processes. 1 ed. New York, NY: Springer; 1993.
Adapted from Figure 1. Time-to-Event Analysis for Primary and Confirmatory Secondary Efficacy Endpoints.
Lincoff AM, et al. N Engl J Med 2023;doi:10.1056/NEJMoa2307563.

Composite heart failure* endpoint

Time-to-first-event analysis of confirmatory secondary endpoint

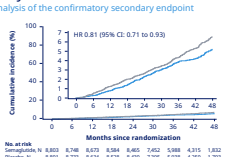


No. at risk								
Semaglutide, N	8,803	8,740	8,654	8,557	8,425	7,409	5,944	4,277	1,816
Placebo, N	8,801	8,711	8,601	8,485	8,381	7,341	5,885	4,198	1,766

*Composite heart failure endpoint of death from cardiovascular causes or hospitalization or urgent medical visit for heart failure.
The cumulative incidences were estimated using the Aalen–Johansen method accounting for competing risk,[1] and the hazard ratios were estimated with the Cox proportional hazards regression model. Because differences for death from cardiovascular causes did not meet the required P value for hierarchical testing, results for the two endpoints subsequent in the hierarchical testing are reported as point estimates and 95% confidence intervals. The widths of the confidence intervals that are presented have not been adjusted for multiplicity and cannot be used to infer definitive treatment effects for these secondary endpoints. The insets show the same data on an enlarged y axis. The x axis is truncated at 48 months due to the limited number of patients in the trial after 48 months.
1. Andersen PK, et al. Statistical Models Based on Counting Processes. 1 ed. New York, NY: Springer; 1993.
Adapted from Figure 1. Time-to-Event Analysis for Primary and Confirmatory Secondary Efficacy Endpoints.
Lincoff AM, et al. N Engl J Med 2023;doi:10.1056/NEJMoa2307563.

Death from any cause

Time-to-first-event analysis of the confirmatory secondary endpoint



No. at risk								
Semaglutide, N	8,803	8,748	8,673	8,584	8,465	7,452	5,988	4,315	1,832
Placebo, N	8,801	8,733	8,634	8,528	8,430	7,395	5,938	4,250	1,793

The cumulative incidences were estimated using the Aalen–Johansen method accounting for competing risk,[1] and the hazard ratios were estimated with the Cox proportional hazards regression model. Because differences for death from cardiovascular causes did not meet the required P value for hierarchical testing, results for the two endpoints subsequent in the hierarchical testing are reported as point estimates and 95% confidence intervals. The widths of the confidence intervals that are presented have not been adjusted for multiplicity and cannot be used to infer definitive treatment effects for these secondary endpoints. The insets show the same data on an enlarged y axis. The x axis is truncated at 48 months due to the limited number of patients in the trial after 48 months.
1. Andersen PK, et al. Statistical Models Based on Counting Processes. 1 ed. New York, NY: Springer; 1993.
Adapted from Figure 1. Time-to-Event Analysis for Primary and Confirmatory Secondary Efficacy Endpoints.
Lincoff AM, et al. N Engl J Med 2023;doi:10.1056/NEJMoa2307563.

16

Composite Endpoints in Obesity CVOT

Primary and secondary time-to-first-event efficacy endpoints (1/3)

Outcome	Semaglutide 2.4 mg (N = 8,803)	Placebo (N = 8,801)	Hazard ratio (95% CI)	P value
	Number of patients (%)			
Primary cardiovascular composite endpoint*	569 (6.5)	701 (8.0)	0.80 (0.72 to 0.90)	<0.001
Confirmatory secondary endpoints†				
Death from cardiovascular causes	223 (2.5)	262 (3.0)	0.85 (0.71 to 1.01)	0.07
Heart failure composite endpoint‡	300 (3.4)	361 (4.1)	0.82 (0.71 to 0.96)	NA†
Death from any cause	375 (4.3)	458 (5.2)	0.81 (0.71 to 0.93)	NA†

Data are for the full analysis set during the in-trial observation period (from randomization to the final follow-up visit). All endpoints were analyzed using a Cox proportional hazards model with treatment as categorical fixed factor. Data from patients without events of interest were censored at the end of their in-trial period.
**The primary efficacy endpoint was a composite of death from cardiovascular causes, nonfatal myocardial infarction, or nonfatal stroke. The hazard ratio, 95% CI, and P value were adjusted for the group sequential design using the likelihood ratio ordering, and the nominal two-sided significance level was 0.046.*
†Confirmatory secondary endpoints were analyzed under multiplicity control via a stagewise hierarchical testing scheme, for which all P values after the first nonsignificant P value are not reported. The P value (unadjusted) for the primary and confirmatory secondary endpoints were to be compared with the nominal significance level derived from the relevant alpha spending function for the endpoint; if the P value was below the nominal limit, superiority would be shown. The nominal two-sided significance level was 0.023 for death from cardiovascular causes.
‡The heart failure composite endpoint was the first occurrence of death from cardiovascular causes or hospitalization or urgent medical visit for heart failure.
CI, confidence interval; NA, not applicable.
Adapted from Table 2. Primary and Secondary Time-to-Event Efficacy Endpoints.
Lincoff AM, et al. N Engl J Med 2023;doi:10.1056/NEJMoa2307563.

17

Composite Endpoints in Obesity CVOT

Primary and secondary time-to-first-event efficacy endpoints (2/3)

Outcome	Semaglutide 2.4 mg (N = 8,803)	Placebo (N = 8,801)	Hazard ratio (95% CI)	P value
	Number of patients (%)			
Supportive secondary endpoints*				
Cardiovascular expanded composite endpoint†	873 (9.9)	1,074 (12.2)	0.80 (0.73 to 0.87)	NA*
Cardiovascular expanded composite endpoint with death from any cause‡	710 (8.1)	877 (10.0)	0.80 (0.72 to 0.88)	NA*
Non-fatal myocardial infarction	234 (2.7)	322 (3.7)	0.72 (0.61 to 0.85)	NA*
Non-fatal stroke	154 (1.7)	165 (1.9)	0.93 (0.74 to 1.15)	NA*
Hospitalization or urgent visit for heart failure	97 (1.1)	122 (1.4)	0.79 (0.60 to 1.03)	NA*

Data are for the full analysis set during the in-trial observation period (from randomization to the final follow-up visit). All endpoints were analyzed using a Cox proportional hazards model with treatment as categorical fixed factor. Data from patients without events of interest were censored at the end of their in-trial period.
**Because supportive secondary endpoints were not corrected for multiplicity, results are reported as point estimates and 95% CIs. The widths of the CIs have not been adjusted for multiplicity, so the intervals should not be used to infer definitive treatment effects for supportive secondary endpoints.*
†The cardiovascular expanded endpoint was a composite of death from cardiovascular causes, nonfatal myocardial infarction, nonfatal stroke, coronary revascularization, or unstable angina requiring hospitalization.
‡The cardiovascular endpoint with death from any cause was a composite of death from any cause, nonfatal myocardial infarction, or nonfatal stroke.
CI, confidence interval; NA, not applicable.
Adapted from Table 2. Primary and Secondary Time-to-Event Efficacy Endpoints.
Lincoff AM, et al. N Engl J Med 2023;doi:10.1056/NEJMoa2307563.

18

Composite Endpoints in Obesity CVOT

Primary and secondary time-to-first-event efficacy endpoints (3/3)

Outcome	Semaglutide 2.4 mg (N = 8,803)	Placebo (N = 8,801)	Hazard ratio (95% CI)	P value
	Number of patients (%)			
Supportive secondary endpoints*				
Coronary revascularization	473 (5.4)	608 (6.9)	0.77 (0.68 to 0.87)	NA*
Unstable angina requiring hospitalization	109 (1.2)	124 (1.4)	0.87 (0.67 to 1.13)	NA*
Glycated hemoglobin ≥6.5%†	306 (3.5)	1,059 (12.0)	0.27 (0.24 to 0.31)	NA*
Nephropathy composite endpoint‡	155 (1.8)	198 (2.2)	0.78 (0.63 to 0.96)	NA*
Glycated hemoglobin ≥5.7% among patients with baseline glycated hemoglobin <5.7%§	623 (21.3)	1,501 (50.4)	0.33 (0.30 to 0.36)	NA*

Data are for the full analysis set during the in-trial observation period (from randomization to the final follow-up visit). All endpoints were analyzed using a Cox proportional hazards model with treatment as categorical fixed factor. Data from patients without events of interest were censored at the end of their in-trial period.
**Because supportive secondary endpoints were not corrected for multiplicity, results are reported as point estimates and 95% CIs. The widths of the CIs have not been adjusted for multiplicity, so the intervals should not be used to infer definitive treatment effects for supportive secondary endpoints.*
†Patients who underwent randomization in error and had a baseline glycated hemoglobin level higher than 6.5% (48 mmol/mol) were excluded from this analysis; 8,800 patients in the semaglutide group and 8,797 patients in the placebo group were included.
‡The nephropathy endpoint was a five-component composite of death from renal causes, initiation of chronic renal replacement therapy (dialysis or transplantation), onset of persistent estimated glomerular filtration rate <15 mL/min/1.73 m², persistent 50% reduction in estimated glomerular filtration rate compared with baseline, or onset of persistent macroalbuminuria (urinary albumin-to-creatinine ratio >300 mg/g).
§A glycated hemoglobin level of 5.7% or higher was assessed in a time-to-event analysis only among patients whose glycated hemoglobin was lower than 5.7% at baseline screening; 2,925 patients in the semaglutide group and 2,980 patients in the placebo group were included.
CI, confidence interval; NA, not applicable.
Adapted from Table 2. Primary and Secondary Time-to-Event Efficacy Endpoints.
Lincoff AM, et al. N Engl J Med 2023;doi:10.1056/NEJMoa2307563.

19

Composite Endpoints in Obesity CVOT

Supportive continuous secondary endpoints*

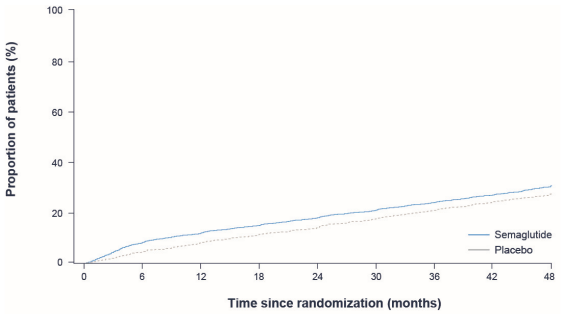
Outcome	Semaglutide 2.4 mg (N = 8,803)	Placebo (N = 8,801)	Difference† (95% CI)
	Mean change from randomization to week 104		
Bodyweight, %	−9.39 ± 0.09	−0.88 ± 0.08	−8.51 (−8.75 to −8.27)
Waist circumference, cm	−7.56 ± 0.09	−1.03 ± 0.09	−6.53 (−6.79 to −6.27)
Glycated hemoglobin level, percentage points	−0.31 ± 0.00	0.01 ± 0.00	−0.32 (−0.33 to −0.31)
Systolic blood pressure, mmHg	−3.82 ± 0.16	−0.51 ± 0.16	−3.31 (−3.75 to −2.88)
Diastolic blood pressure, mmHg	−1.02 ± 0.10	−0.47 ± 0.10	−0.55 (−0.83 to −0.27)
Heart rate, beats/min	3.79 ± 0.11	0.69 ± 0.11	3.10 (2.80 to 3.39)

Data are from the full analysis set; plus-minus values are means ± SE.
The continuous endpoints assessing change from randomization to week 104 were analyzed using ANCOVA with treatment as factor and the baseline value as covariate, using multiple imputation for missing values under a missing-at-random assumption.
**Because supportive secondary endpoints were not corrected for multiplicity, results are reported as point estimates and 95% CIs. The widths of the CIs have not been adjusted for multiplicity, so the intervals should not be used to infer definitive treatment effects for supportive secondary endpoints.*
†differences are given as the difference for the changes in continuous endpoints.
ANCOVA, analysis of covariance; CI, confidence interval; SE, standard error.
Adapted from Table 3. Supportive Binary and Continuous Secondary Endpoints.
Lincoff AM, et al. N Engl J Med 2023;doi:10.1056/NEJMoa2307563.

20

Composite Endpoints in Obesity CVOT

Figure S2. Cumulative Proportion of Patients Who Permanently Prematurely Discontinued Treatment.



Semaglutide	8,803	8,051	7,671	7,364	7,063	5,719	4,670	2,812	914
Placebo	8,801	8,342	7,979	7,663	7,379	5,957	4,862	2,953	936

Cumulative incidence estimates are based on time from randomization to permanent treatment discontinuation, with death modeled as a competing risk. Patients never exposed are censored at day 1. Treatment discontinuations that occurred less than 30 days before the end-of-treatment visit were not counted.

21

Overall Survival vs Complete Response in AML

CLINICAL TRIALS AND OBSERVATIONS

Phase 3 randomized, placebo-controlled, double-blind study of high-dose continuous infusion cytarabine alone or with laromustine (VNP40101M) in patients with acute myeloid leukemia in first relapse

Francis Giles,[1] Norbert Vey,[2] Daniel DeAngelo,[3] Karen Seiter,[4] Wendy Stock,[5] Robert Stuart,[6] Darinka Boskovic,[7] Arnaud Pigneux,[8] Martin Tallman,[9] Joseph Brandwein,[10] Jonathan Kell,[11] Tadeusz Robak,[12] Peter Staib,[13] Xavier Thomas,[14] Ann Cahill,[15] Maher Albitar,[16] and Susan O'Brien[17]

[1]Cancer Therapy & Research Center (CTRC) at The University of Texas Health Science Center, San Antonio; [2]Institut Paoli Calmettes, Marseille, France; [3]Dana-Farber Cancer Institute, Boston, MA; [4]New York Medical College, Valhalla; [5]University of Chicago, IL; [6]Medical University of South Carolina (MUSC) Hollings Cancer Center, Charleston; [7]Clinical Center of Serbia, Belgrade, Serbia; [8]Hôpital Haut Lévèque, Bordeaux, France; [9]Robert H. Lurie Comprehensive Cancer Center, Chicago, IL; [10]Princess Margaret Hospital, Toronto, ON; [11]University Hospital of Wales, Cardiff, United Kingdom; [12]Medical University of Lodz, Lodz, Poland; [13]Klinikum der Universität Köln, Köln, Germany; [14]Hôpital Edouard Herriot, Lyon, France; [15]Vion Pharmaceuticals, New Haven, CT; [16]Quest Diagnostics Nichols Institute, San Juan Capistrano, CA; and [17]University of Texas (UT) M. D. Anderson Cancer Center, Houston

Laromustine is a sulfonylhdrazine alkylator with significant antileukemia activity. An international, randomized (2:1), double-blind, placebo-controlled study was conducted to compare complete remission (CR) rates and overall survival (OS) in patients with first relapse acute myeloid leukemia (AML) treated with laromustine and high-dose cytarabine (HDAC) versus HDAC/placebo. Patients received 1.5 g/m^2 per day cytarabine continuous infusion for 3 days and laromustine 600 mg/m^2 (n = 177) or placebo (n = 86) on day 2. Patients in CR received consolidation with laromustine/HDAC or HDAC/placebo as per initial randomization. After interim analysis at 50% enrollment, the Data Safety Monitoring Board (DSMB) expressed concern that any advantage in CR would be compromised by the observed on-study mortality, and enrollment was held. The CR rate was significantly higher for the laromustine/HDAC group (35% vs 19%, *P* = .005). However, the 30-day mortality rate and median progression-free survival were significantly worse in this group compared with HDAC/placebo (11% vs 2%; *P* = .016; 54 days vs 34; *P* = .002). OS and median response durations were similar in both groups. Laromustine/HDAC induced significantly more CR than HDAC/placebo, but OS was not improved due to mortality associated with myelosuppression and its sequelae. The DSMB subsequently approved a revised protocol with laromustine dose reduction and recombinant growth factor support. The study was registered as NCT00112554 at http://www.clinicaltrials.gov. (Blood. 2009;114:4027-4033)

22

Overall Survival vs Complete Response in AML

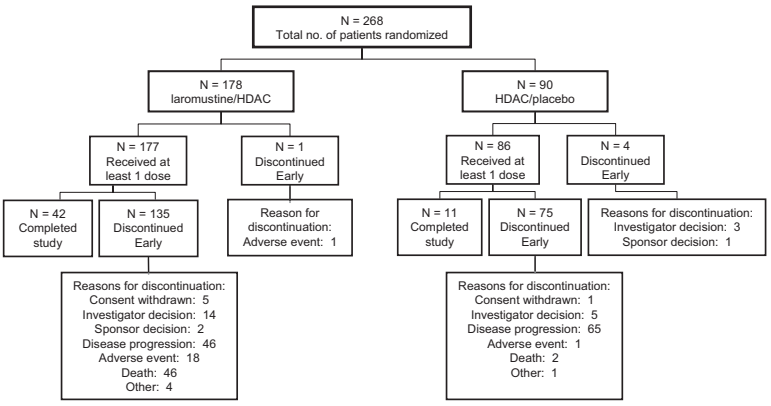


Figure 1. Patient disposition.

23

Overall Survival vs Complete Response in AML

- Total CR: 35% v 19%, p= .005

4030 GILES et al BLOOD, 5 NOVEMBER 2009 • VOLUME 114, NUMBER 19

Table 2. Response rates by patient stratum

	Complete response		Complete response with inadequate platelet recovery		Total complete response*†	
	Laromustine/HDAC, no. (%)	HDAC/placebo, no. (%)	Laromustine/HDAC, no. (%)	HDAC/placebo, no. (%)	Laromustine/HDAC, no. (%)	HDAC/placebo, no. (%)
All patients. Laromustine/HDAC, n = 179; HDAC/placebo, n = 86	36 (20)	14 (16)	26 (15)	2 (2)	62 (35)*	16 (19)
Stratum 1: Younger than 60 y, CR less than 12 mo. Laromustine/HDAC, n = 61; HDAC/placebo, n = 26	7 (11)	4 (15)	9 (15)	0	16 (26)	4 (15)
Stratum 2: Younger than 60 y, CR 12 or more mo. Laromustine/HDAC, n = 31; HDAC/placebo, n = 16	7 (22)	7 (44)	3 (10)	0	10 (32)	7 (44)
Stratum 3: 60 y or older, CR less than 12 mo. Laromustine/HDAC, n = 51; HDAC/placebo, n = 26	14 (28)	1 (4)	4 (8)	0	18 (35)	1 (4)
Stratum 4: 60 y or older, CR 12 or more mo. Laromustine/HDAC, n = 34; HDAC/placebo, n = 18	8 (24)	2 (11)	10 (29)	2 (11)	18 (53)	4 (22)

*Complete response plus complete response with inadequate platelet recovery.

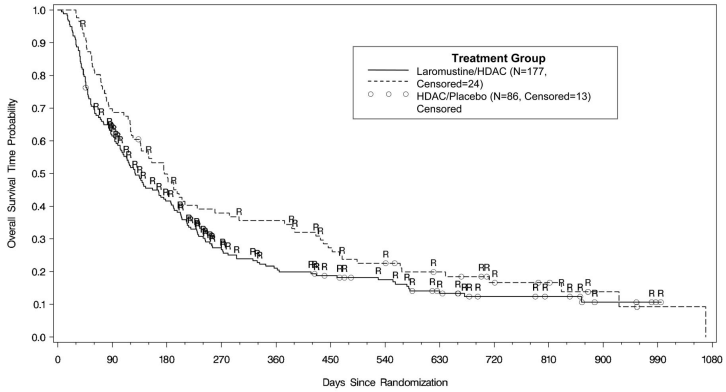
†P = .005 comparing the overall response rate between treatment arms using the Chochran-Mantel-Haenszal test stratified by age group and duration of first CR/CRp. Difference (95% confidence interval) between all strata of each treatment group (laromustine/HDAC–HDAC/placebo) was 16% (56%-27%).

24

Overall Survival vs Complete Response in AML

- OS: logrank p= .087; first 30 days post induction: p= .015

BLOOD, 5 NOVEMBER 2009 • VOLUME 114, NUMBER 19 HDAC ± LAROMUSTINE IN FIRST RELAPSE AML 4031



Note: Patients alive as of the date of last follow–up have their survival time censored at last date of known mortality status.
R = patient achieved a CR or CRp.

Figure 2. Kaplan-Meier estimate of overall survival time.

25

Comments for Later

- From Obesity CVOT:
  - Hierarchical endpoints
  - Events defined by continuous laboratory measurements
  - Composite endpoints vs individual components for MACE, HHF
  - Cumulative incidence curves “accounting” for competing risks
- From laromustine in AML:
  - Primary CR vs secondary OS endpoints
  - Early stopping for safety
  - Data-driven hypotheses in reporting
  - Principal stratification by response

26

26

ICH E-9: Modification of Estimands

- While on treatment strategies
  - "response to treatment prior to … the intercurrent event"
  - "Particular care is required if the occurrence of the intercurrent event differs between the treatments being compared"
- This is not a rigorous scientific comparison
  - We randomize for a reason
  - There are multiple reasons a subject might stop treatment
    - Protopathic and indication bias of particular concern
  - "if the occurrence of the intercurrent event differs"
    - Not just a question of how often, but in whom and why

27

ICH E-9: Modification of Estimands

- Principal stratum
  - "Target population might be taken to be the "principal stratum" in which an intercurrent event would (conversely, would not) occur."
    - E.g., population of "tolerators", "responders", "compliers"
    - E.g., population that would be coerced into continuing treatment
  - Distinguished from subsetting on observed intercurrent event
    - Hence, a hypothetical scenario based on predictive model for principal stratum membership
- Examples of principal stratum approaches include
  - Duration of response in clinical trials
  - Severity of infection in vaccine trials
- Alternative approaches based on pre-randomization variables are vastly to be preferred

28

Bottom Line

- The validity of any analysis that is not based on randomization (i.e., any non-ITT analysis) depends on untestable assumptions
  - The enemy: "Assuming no unmeasured confounding"
  - It is extremely rare (unheard of) that prognostic variables would have $R^2 = 1.0$ in predictive models
- With respect to estimands related to adherence to protocol, non-ITT is particularly problematic
  - Indication for treatment changes are rarely well documented
    - Controlling for indication bias is most often problematic
  - Protopathic bias is by its very definition unmeasured confounding
- Sensitivity analyses to judge the impact of assumptions underlying any but the per randomization analysis are key

29

RCT Settings Using Time to Event Outcomes

- Overall goal: Drug discovery
- Estimands
  - Clinical
  - RCT
  - ICH E9 (R1) strategies for intercurrent events
- **Why an "event"? Why "time to event"?**
- **Why incomplete observation: Informative vs noninformative?**
  - **Administrative censoring**
  - **Competing risks**
  - **Intercurrent events and protopathic events**
  - **Loss to follow-up**
- How to define "tends to be"?
  - Choice of summary measure

30

30