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 5:

Events and Time to Event

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 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

2

2

Events as Primary Outcomes

Where am I going?

- The goal of a RCT is to find effective treatment indications
- The primary outcome is a crucial element of the indication
- The most important clinical outcome is often measured as an "event"

3

3

Primary Endpoint: Clinical

- Consider (in order)
  - The most relevant clinical endpoint
    - Survival, quality of life
  - The endpoint the treatment is most likely to affect
  - The endpoint that can be assessed most accurately and precisely

4

4

Additional Endpoints

- Other outcomes are then relegated to a “secondary” status
  - Supportive and confirmatory
  - Safety
- Some outcomes are considered “exploratory”
  - Subgroup effects
  - Effect modification

5

5

Primary Endpoint: Clinical

- Consider (in order of importance)
  - The phase of study: What is current burden of proof?
  - The most relevant clinical endpoint
    - Survival, quality of life
    - Proven surrogates for the above
      - But how can we be sure?
  - The endpoint the treatment is most likely to affect
    - Therapies directed toward improving survival
    - Therapies directed toward decreasing AEs
  - The endpoint that can be assessed most accurately and precisely
    - Avoid unnecessarily highly invasive measurements
    - Avoid poorly reproducible endpoints

6

6

Multiple Endpoints

- Sometimes we must consider multiple endpoints
- We then control experimentwise error
- Possible methods
  - Composite endpoint
    - AND: Individual success must satisfy all
    - OR: Individual success must only satisfy one
    - AVERAGE: Sum of individual scores
    - EARLIEST: e.g., event free survival
  - Co-primary endpoints
    - Must show improvement in treatment group on all endpoints
      - No guarantee that the same subjects are experiencing the improvement

7

7

Event: A Binary Measurement

- Only two possible values, which can be either
  - E.g., “Alive” or “Dead”
- Time of measurement
  - Fixed point in time
    - E.g., neurologically intact survival 6 months after cardiac arrest
  - Longitudinally measured up to a fixed point in time
    - E.g., time until death
  - Recurrent events
    - E.g., hospitalization
      - May have to distinguish extended events vs recurrent events

8

Scientific Basis

- A clinical trial is planned to detect the effect of a treatment on some outcome
- Statement of the outcome is a fundamental part of the scientific hypothesis

9

9

Ethical Basis

- Generally, subjects participating in a clinical trial are hoping that they will benefit in some way from the trial
- Clinical endpoints are therefore of more interest than purely biological endpoints

10

10

Regulatory Basis

- FDA: Evidence that treatment has a positive impact on how a patient: "***feels, functions, or survives"***
- How a patient feels or functions in daily life: Clinical Outcome Assessments (COAs):
  - Patient-Reported Outcomes (PROs)
  - Clinician-Reported Outcomes (ClinROs)
  - Observer-Reported Outcomes (ObsROs)
  - Performance Outcomes (PerfOs)
- Clinical Benefit / Survival
- In all cases, need to "anchor" observed (statistically significant) differences to clinical importance
  - Population level: e.g., mean change
  - Individual level: proportion showing an important difference

11

11

Overall Survival

- Should overall survival be the primary endpoint?
  - It depends – need to consider how patient "feels, functions"
    - Also whether logistically feasible
  - (Mortality is <u>always</u> an important safety endpoint)
- Rapidly fatal vs ultimately fatal vs nonfatal diseases
  - E.g., Advanced cancer vs cardiovascular disease vs pain relief
- Significant degradation in quality of life prior to death
  - E.g, Advanced cancer, Alzheimer's disease, hemorrhagic stroke
- Treatments prone to Grade 5 adverse events
  - E.g., Cancer chemotx, conditioning regimens for stem cell transplant
- Competing causes of death potentially unaffected by treatment
  - E.g., Screening trials to prevent lung cancer deaths

12

12

Multifactorial Measures of Disease Progression

- Most diseases have multiple deleterious signs and symptoms
  - Some may be measured on a continuous scale
- Great individual variability among diseased patients in incidence and penetrance
- The various potential measures are only partially ordered
  - When comparing two patients, a patient who is worse on every possible measure is clearly worse
  - But it is not always clear how to judge when first patient is worse with respect to sign/symptom A, and second patient is worse with respect to sign/symptom B
- It is thus common to define "disease progression" by prespecifying thresholds for each sign/symptom and using earliest

13

13

Example: Cardiovascular Disease

- Usually adjudicated by blinded Endpoint Adjudication Committee
- Cardiovascular death (incl sudden death and unknown cause)
- Composite outcomes (many other variants)
  - Major adverse cardiovascular events (MACE)
    - Nonfatal myocardial infarction
    - Nonfatal stroke
    - Cardiovascular death
  - Heart failure composite
    - Hospitalization or urgent treatment for heart failure
    - Cardiovascular death
  - Major adverse limb events (MALE)
    - Hospitalization for acute limb ischemia
    - Hospitalization for chronic limb ischemia

14

14

Example: Cancer

- Overall survival
- Progression by Blinded Independent Central Review (BICR) vs investigator
  - Response Evaluation Criteria in Solid Tumors (RECIST)
  - Tumor growth, new metastases
  - Complete response, partial response, stable disease, progression
  - Objective response rate, Clinical benefit rate at specified time
- Composite outcomes
  - Progression free survival
  - Recurrence free survival
  - Event free survival
  - Disease free survival

15

15

Example: Renal Disease

- Composite outcome
  - The onset of a persistent ≥50% reduction in eGFR (CKD-EPI 2009) compared with baseline
  - The onset of persistent eGFR (CKD-EPI) < 15 mL/min/1.73 m_2
  - The initiation of chronic renal replacement therapy
    - dialysis or kidney transplantation
  - Renal death
  - CV death

16

16

Example: Liver Disease

- Overall survival
- Composite
  - Transplant free survival
  - Clinical decompensation
    - Ascites
    - Hepatic encephalopathy
    - Variceal bleeding
    - Transplant
    - Death (all causes)

17

17

Example: Pulmonary Diseases

- Cystic Fibrosis
  - Exacerbations
- Asthma
  - Rate of asthma exacerbations
- Chronic obstructive pulmonary disease (COPD)
  - Moderate / severe exacerbations
    - Rate (recurrences)
    - Time to first
  - All cause mortality
  - St. George's Respiratory Questionaire-COPD responder rate

18

18

Example: Reproductive Drugs

- Female infertility
  - Chemical pregnancy
  - Pregnancy
  - Live birth
- Female sterilization
  - Pregnancy
- Hypoactive Sexual Desire Disorder
  - Number of Sexually Satisfying Episodes

19

19

Competing Risks

- Occurrence of some "nuisance" event precludes observation of the event of greatest interest, because
  - Further observation impossible
    - E.g., death from CVD in cancer study
  - Further observation irrelevant
    - E.g., patient advances to other therapy (transplant)
- Methods
  - Event free survival: time to earliest event
  - Time to progression: censor competing risks
  - "U statistics": define ranking based on both events

20

20

Competing Risks Caveats

- Competing risks produce missing data on the event of greatest interest
- As with all missing data problems, there is nothing in your data that can tell you whether your actions are appropriate
  - Are subjects with competing risk more or less likely to have event of interest?
  - (the term "competing risk" has become shorthand for a setting in which your results are in doubt)
- Even when competing risks are "noninformative", special analysis methods may be necessary to accurately judge the proportion of subjects who will experience the event

21

21

Statistical Issues

- Relevant clinical outcomes are often relatively rare events that occur after a significant delay
  - Believe that earlier interventions have greater chance of benefit
- Difficulty in measuring clinical outcome
  - Quality of life needs to be assessed over a sufficiently long period of time

22

22

Impact on Clinical Trial Design

- Large sample size required to assess treatment effect on rare events
- Long period of follow-up needed to assess endpoints
- Possible solutions
  - Analyses prior to full observation on all subjects
    - The major topic of this course
  - Surrogate outcomes
    - Often used in screening trials (discussed later)
    - Validation needed before use in confirmatory trials (not covered here)

23

23

Statistics and Game Theory

- Multiple comparison issues
  - Type I error for each endpoint
    - In absence of treatment effect, will still decide a benefit exists with probability, say, .025
- Multiple endpoints increase the chance of deciding an ineffective treatment should be adopted
  - This problem exists with either frequentist or Bayesian criteria for evidence
    - But the consideration of the impact of multiple comparisons is primarily frequentist
  - The actual inflation of the type I error depends
    - the number of multiple comparisons, and
    - the correlation between the endpoints

24

24

Ex: Level 0.05 per Decision

- Experiment-wise Error Rate

Number Compared	Worst Case	Correlation 0.00	0.30	0.50	0.75	0.90
1	.050	.050	.050	.050	.050	.050
2	.100	.098	.095	.090	.081	.070
3	.150	.143	.137	.126	.104	.084
5	.250	.226	.208	.184	.138	.101
10	.500	.401	.353	.284	.193	.127
20	1.000	.642	.540	.420	.258	.154
50	1.000	.923	.806	.624	.353	.193

25

25

Events over Time

Where am I going?

- Judging the occurrence of events over time is often of great scientific importance
- When scientifically appropriate, we can gain statistical precision by considering the time course

26

26

Events Occurring Over Time: Options

- Clinically, there are multiple options that might be considered
- Prevalence of event at a fixed point in time
- Cumulative incidence of an event over a time interval
- Time to occurrence of first event over a time interval
- Rate of recurrent events over a time interval

27

27

Prevalence at a Fixed Point in Time

- Ex: in rheumatologic disease, off steroids two years after accrual
- Prespecified, clinically relevant time point
- Allows for variability in time to achieving the event
  - But depending on the event and the time point, time in a "good" state may be more clinically important
- Truly this can be addressed even when not all patients are followed to the prespecified time point
  - Providing:
    - no time trends in patient accrual, and
    - time of follow-up for each patient is prespecified equally by arm
  - Then incomplete data is MAR
    - Impute final data using subjects followed for full period

28

28

Cumulative Incidence at a Fixed Point in Time

- Ex: five year survival probability in childhood cancer
- Prespecified, clinically relevant time point
- Especially indicated when
  - Early differences in incidence rate are clinically less important,
  - Crossing survival functions are plausible, and / or
  - Crossing hazard functions might affect statistical precision
- Truly this can be addressed even when not all patients are followed to the prespecified time point
  - Providing:
    - no time trends in patient accrual, and
    - time of follow-up for each patient is prespecified equally by arm
  - Then incomplete data is MAR
    - Impute final data using subjects followed for full period

29

29

Time to First Event over a Time Interval

- Ex: time to death from any cause in stage 4 lung cancer
- Prespecified criteria for time of performing analysis
- Especially indicated when
  - earlier occurrence of events is indicative of disease severity,
  - crossing hazards and crossing survival curves are unlikely, and
  - there is a possibility that patient treatment will change after first event
- Truly this can be addressed even when not all patients are followed to the prespecified time point
  - Providing:
    - no time trends in patient accrual, and
    - time of follow-up for each patient is prespecified equally by arm
  - Then incomplete data is MAR
    - Impute final data using subjects followed for full period

30

30

Rate of Recurrent Events over a Time Interval

- Ex: number / rate of hospitalizations for heart failure
- Prespecified criteria for time of performing analysis
- Especially indicated when
  - number of events is indicative of worse quality of life,
  - treatment might only lead to early improvements, and/or
  - treatment effect might be delayed
- Truly this can be addressed even when not all patients are followed to the prespecified time point
  - Providing:
    - no time trends in patient accrual, and
    - time of follow-up for each patient is prespecified equally by arm
  - Then incomplete data is MAR
    - Impute final data using subjects followed for full period

31

31

Example: Time to "Biochemical" Diabetes

- Obesity patients receiving semaglutide or placebo
  - Time to HbA1c > 6.5% in all patients (C) or baseline > 5.7% (D)

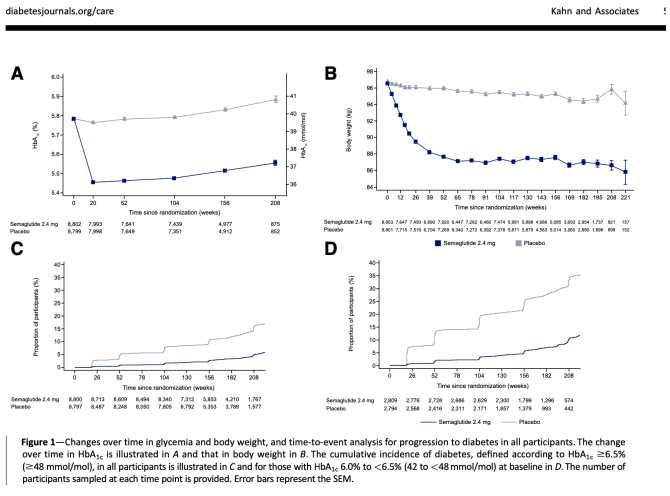


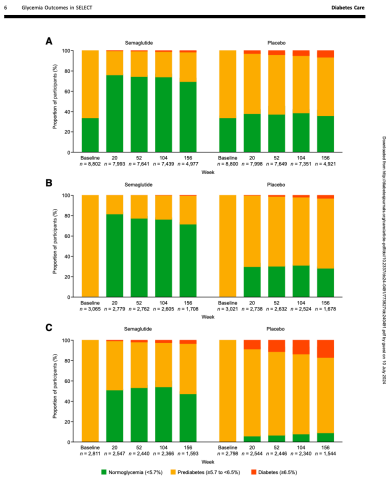
Figure 1—Changes over time in glycemia and body weight, and time-to-event analysis for progression to diabetes in all participants. The change over time in HbA_{1c} is illustrated in *A* and that in body weight in *B*. The cumulative incidence of diabetes, defined according to HbA_{1c} ≥6.5% (≥48 mmol/mol), in all participants is illustrated in *C* and for those with HbA_{1c} 6.0% to <6.5% (42 to <48 mmol/mol) at baseline in *D*. The number of participants sampled at each time point is provided. Error bars represent the SEM.

32

32

Example: Time to "Biochemical" Diabetes

- Obesity patients receiving semaglutide or placebo
  - HbA1c categories over time: avoid cumulative incidence for regression
  - All subjects
  - Baseline 5.7% - < 6.0%
  - Baseline 6.0% - < 6.5%

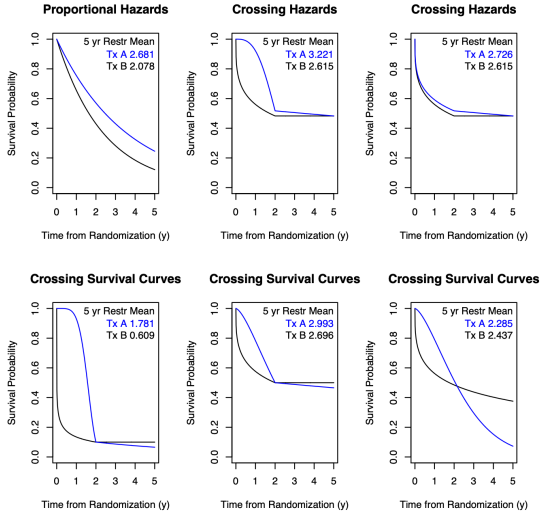


33

33

Hypothetical Survival Curves

- Settings where alternative strategies of events over time might be indicated



34

34

Variable Follow-up on Subjects

Where am I going?

- In all methods of dealing with event distributions over time, I remarked that time of follow-up could vary between individuals
- It is useful to revisit the various reasons that lead to different follow-up times

35

35

Informative vs Noninformative Censoring

- Variable follow-up on individuals does not affect our inference about event rates, providing the length of follow-up is not informative about impending events
- We thus need to consider the mechanisms whereby length of follow-up will vary
  - Administrative censoring
  - Intercurrent events
    - Competing risks
    - Protopathic events
  - Withdrawal of consent
  - Loss to follow-up

36

36

Administrative Censoring

- The preponderance of variable follow-up in time to event studies relates to study design
  - Patients are accrued to the study over some time interval
  - Analysis of data occurs at a prespecified point in calendar time
    - Most often "prespecified" by the number of events required
- This "administrative censoring" is typically regarded as MAR
  - Presuming that randomization is blocked and that there is no major prognostic factor that varies by calendar time
- Issues might arise if
  - early vs late accruals differ by prevalent vs incident disease
  - geographic variation as clinical sites are opened
  - But in either of these cases, blocked randomization will mitigate issues

37

37

Intercurrent Events

- Competing risks
  - If noninformative about event of primary interest, may still bias estimates of cumulative incidence
  - If informative on primary event, then differential bias may arise
    - E.g., cardiovascular death as component of MACE and HF composites
- Protopathic bias
  - Sometimes the earliest sign/symptom of an impending event is vague symptomatology leading to study drug d/c or study withdrawal
  - Censoring such patients may introduce systematic bias

38

38

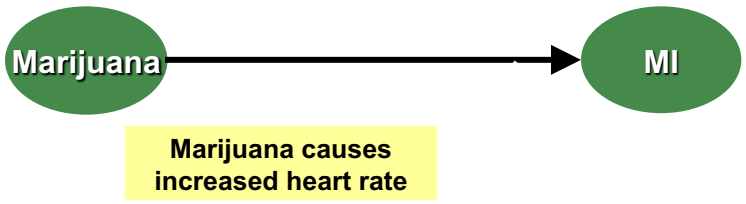
Issues: Informative Censoring

- Possibility that impending event causes informative censoring (confounding?)
- Types of variables
  - Extrinsic: Unaffected by individual decisions
    - As a rule, time-varying extrinsic variables will not cause informative censoring
    - E.g., Air pollution on a given day in an asthma study
      - (providing it does not affect relocation)
  - Intrinsic: Potentially affected by impending event
    - E.g., Marijuana use

39

Causation versus Association

- Example: Scientific interest in causal pathways between marijuana use and heart attacks (MI)
  - Pictorial representation of hypothetical causal effect of marijuana on MI that might be of scientific interest

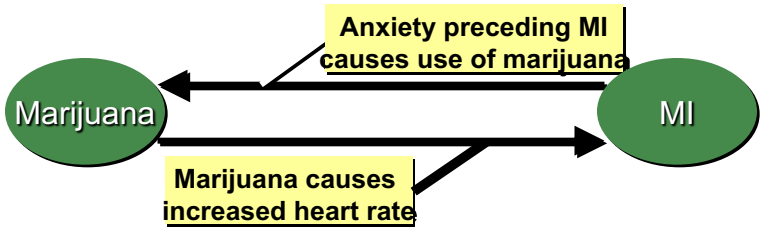


Marijuana causes increased heart rate

40

Causation versus Association

- In an observational study, we cannot thus be sure which causative mechanism an association might represent
  - Either of these mechanisms will result in an association between marijuana use and MI



41

Withdrawal and Loss to Follow-up

- The underlying mechanism behind patient withdrawal or loss to follow-up is rarely known
  - Going home to die surrounded by family vs trekking in Nepal
- We might speculate based on
  - prior clinical course having excessive AEs
  - prior clinical course suggestive of disease progression
- In some cases (and some countries) vital status records can be abstracted from public data sources
- In some extreme cases, private investigators have been used to try to find LTFU patients
- But ultimately, we just have to do sensitivity analyses

42

42

Next Lecture

- 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**

43

43