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 6:

Choice of Summary Measure

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

## Descriptive Statistics

- Where am I going?
  - Computing descriptive statistics on sample data is easy
  - 10th grade WASL
- Understanding what to use when and what they tell you is much harder
  - Important when it comes to inference:
    - "Parameters" are often descriptive statistics on the population
    - "Distribution free" inference generally parallels descriptives
- When data is incomplete (e.g., censored), it is more difficult to compute scientifically relevant descriptive statistics
  - The methods we use provide great insight to the behavior of statistical models used for inference

3

## Purpose of Descriptive Statistics

- Identify errors in measurement, data collection
- Characterize materials and methods
- Assess validity of assumptions needed for analysis
  - Scientific
  - Statistical
- Straightforward estimates to address scientific question
- Hypothesis generation

4

## Purpose of Descriptive Statistics

- Identify errors in measurement, data collection
- Characterize materials and methods
- Assess validity of assumptions needed for analysis
  - Scientific
  - Statistical
- **Straightforward estimates to address scientific question**
- Hypothesis generation

5

## Scientific Studies

- A well designed study
  - Discriminates between the most important, viable hypotheses
    - "Discriminates" defined by what convinces your audience
  - Is equally informative for all possible study results
    - Binary search using prior probability of being true
    - Also consider simplicity of experiments, time, cost

6

## Scientific Questions

- Ultimately, scientific questions are most often concerned with investigating cause and effect
- E.g., in biomedical settings:
  - What are the causes of disease?
  - What are the effects of interventions?

7

## Typical Inferential Setting

- In the studies considered here, we define
  - Some "primary outcome" measurement
    - A "response variable" in regression
  - Groups that are homogeneous with respect to the level of some factor(s)
    - Predictor of interest
    - Effect modifiers
    - Confounders
    - Precision variables

8

## Primary Outcome Measurement

- The primary outcome can be derived from more than one measured variable
- E.g., for repeated measurements made on the same experimental unit
  - Contrast across measurements made on different variables
  - Weighted average of measurements made on different variables
- E.g., for random process defined by longitudinal follow up of experimental units
  - Contrast across time
  - Weighted average over time
  - Time until an event

9

## Typical Scientific Hypotheses

- The specified level of some factor **will cause** outcome measurements that are

higher than,
lower than, or
about the same as

an absolute standard, or
measurements in a comparison group

10

## Causation vs Association

- Truly determining causation requires a suitable interventional study (experiment)
- Statistical analyses tell us about associations
- Associations in the presence of an appropriate experimental design allows us to infer causation
- But even then, we need to be circumspect in identifying the true mechanistic cause
  - E.g., a treatment that causes headaches, and therefore aspirin use, may result in lower heart attack rates due entirely to the use of aspirin

11

## First Statistical Refinement

- The group with the specified level of some factor **will have** outcome measurements that are

| | |
|---|---|
| higher than,<br>lower than, or<br>about the same as | an absolute standard, or<br>measurements in a comparison group |

12

## Deterministic Setting

- Conditions of scientific studies might make answering questions difficult even when study results are deterministic
- Difficulties in isolating specific causes
  - E.g., isolating REM sleep from total sleep
  - E.g., interactions between genetics and environment
- Difficulties in measuring potential effects
  - E.g., measuring time to survival
    - length of study
    - competing risks

13

## Can Statistics Help?

- Litmus Test # 1:
  - If the scientific question cannot be answered by an experiment when outcomes are entirely deterministic, there is NO chance that statistics can be of any help.

14

## Variation in Response

- There is, of course, usually variation in outcome measurements across repetitions of an experiment
- Variation can be due to
  - Unmeasured (hidden) variables
    - E.g., mix of etiologies, duration of disease, comorbid conditions, genetics when studying new cancer therapies
  - Inherent randomness
    - (as dictated by quantum theory)

15

## Second Statistical Refinement

- The group with the specified level of some factor will **tend to** have outcome measurements that are

| | |
|---|---|
| higher than,<br>lower than, or<br>about the same as | an absolute standard, or<br>measurements in a comparison group |

16

## Refining Scientific Hypotheses

- In order to be able to perform analysis we must define “will tend to have”
- Probability model for response
  - Nonparametric, semiparametric, parametric
- (Looking ahead: I am a big proponent of nonparametric interpretations of statistical analyses)

17

## Ordering Probability Distributions

- In general, the space of all probability distributions is not totally ordered
- There are an infinite number of ways we can define a tendency toward a “larger” outcome
- This can be difficult to decide even when we have data on the entire population
  - Ex: Is the highest paid occupation in the US the one with
    - the higher mean?
    - the higher median?
    - the higher maximum?
    - the higher proportion making $1M per year?

18

## Can Statistics Help?

- Litmus Test # 2:
  - If the scientific researcher cannot decide on an ordering of probability distributions which would be appropriate when measurements are available on the entire population, there is NO chance that statistics can be of any help.

19

## Summary Measures

- Typically we order probability distributions on the basis of some summary measure
- Statistical hypotheses are then stated in terms of the summary measure
- Primary analysis based on detecting an effect on (most often) one summary measure
  - Avoids pitfalls of multiple comparisons
    - Especially important in a regulatory environment

20

## Purposeful Vagueness

- What I call "summary measures", others might call "parameters"
- "Parameters" suggests use of parametric and semiparametric statistical models
  - I am generally against such analysis methods
- "Functionals" is probably the best word
  - "Functional"= anything computed from a cdf
  - But too much of a feeling of "statistical jargon"

21

## Marginal Summary Measures

- Many times, statistical hypotheses are stated in terms of summary measures for univariate (marginal) distributions
  - Means (arithmetic, geometric, harmonic, …)
  - Medians (or other quantiles)
  - Proportion exceeding some threshold
  - Odds of exceeding some threshold
  - Time averaged hazard function (instantaneous risk)
  - …

22

## Comparisons Across Groups

- Comparisons across groups then use differences or ratios
  - Difference / ratio of means (arithmetic, geometric, …)
  - Difference / ratio of proportion exceeding some threshold
  - Difference / ratio of medians (or other quantiles)
  - Ratio of odds of exceeding some threshold
  - Ratio of hazard (averaged across time?)
  - …

23

## Joint Summary Measures

- Other times groups are compared using a summary measure for the joint distribution
  - Median difference / ratio of paired observations
  - Probability that a randomly chosen measurement from one population might exceed that from the other
  - …

24

## Looking Ahead: Transitivity

- The distinction between marginal versus joint summary measures impacts comparisons across studies
- Most often (always?) transitivity is not guaranteed unless comparisons can be defined using marginal distributions
- Intransitivity: Pairwise comparisons might suggest
  - A > B, and
  - B > C, but
  - C > A

25

## Can Statisticians Help?

- While I claim that the choice of the definition for "tends to be larger" is primarily a scientific issue, statisticians do usually play an important role
- Statisticians do explain how different summary measures capture key features of a probability distribution

26

## Next Lecture: A Quiz

- 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**

27

27