Estimands for Recurrent Event data

& connection to Time-to-First Event analysis.

James Roger
London School of Hygiene & Tropical Medicine, London.

Third EFSPI Workshop on Regulatory Statistics:
Complex data types and designs in confirmatory trials.
September 2018.

Preliminaries

- Thanks to the organisers for the invitation to present.
- Thanks to all those many people and groups with whom I have discussed these concepts. But I accept responsibility for the content.

- I will assume that everyone is fluent with the **draft ICH E9 addendum** document and the working group's definition of an **estimand**.

- I have no financial interest at all to declare and am an honorary professor at the London School of Hygiene and Tropical Medicine.

Time-to-event and recurrent events: What is the big issue?

Some answers:

- Not a Normal distribution.
  - This has technical implications but is not conceptually important other than as a driver towards more robust techniques.
- When a patient is “lost” we always have some actual data for them.
  - We know the number of recurrent events before they are “lost”.
  - Unlike classic MMRM where partial data is simply used to predict.
    - I use the draft addendum example of **final visit HbA1c** as comparator.
  - This has an impact on how we handle **intercurrent events (IEs)**. The IEs compete with the event rather than simply remove data.

To be clear . . .

- When I refer to “event” I will mean the recurrent event itself.
- In the context of a clinical trial, interest is focused on a restricted fixed time interval $[0, t_0]$, usually the length of the trial.
- The term “Censoring”
  - Administrative censoring is a trial specific issue and not related to estimands.
  - Censoring as part of an estimation algorithm is also not an estimand issue.
  - Intercurrent events should not be described as “censoring”.
- The next talk will introduce concrete examples, while I address general issues about estimand definition.

Core definition of estimand (1)

A The population, that is, the patients targeted by the scientific question.

- Nothing new.

B The variable (or endpoint), to be obtained for each patient, that is required to address the scientific question.

- This includes the timing and number of the events themselves and also the timing of any relevant intercurrent events.
- This might simply be the number of events observed for a patient in the period $[0, t_0]$.

C The specification of how to account for intercurrent events to reflect the scientific question of interest.

- This is the crucially new bit. Interrelates closely with B.

Core definition of estimand (2)

D The population-level summary measure for the variable which provides, as required, a basis for a comparison between treatment conditions.

- Two concepts here.
  - Summary across patients.
  - Comparison between arms.
- The value being summarized is effectively counterfactual as it is treatment A versus B, within the same patient.
- In practice we often summarize within each arm and then take difference or ratio.
  Then the estimand can be defined for a single arm. And we compare between arms.

Core definition of estimand (2)

D The population-level summary measure for the variable which provides, as required, a basis for a comparison between treatment conditions.

- Two concepts here.
  - Summary across patients.
  - Comparison between arms.
- The value being summarized is effectively counterfactual as it is treatment A versus B, within the same patient.
- In practice we often summarize within each arm and then take difference or ratio.
  Then the estimand can be defined for a single arm. And we compare between arms.

This is the **“elephant in the room”**.

Summary statistic versus model parameter

- Are we interested in the process and trying to represent it and then draw conclusions about the process?

  Or should we simply describe what happened?

- Requirement for an **Estimand** seems to drive us back to using summary statistics.

  Hopefully this is not true and we are about to search for a compromise.

Current practice for Recurrent Event

- Time-to-first-event and Logrank test / Cox PH estimate.
  - Problems defining the estimand for time-to-event exist but are beyond this talk. But they are similar to those we see later.

Current practice for Recurrent Event

- Time-to-first-event and Logrank test / Cox PH estimate.
  - Problems defining the estimand for time-to-event exist but are beyond this talk. But they are similar to those we see later.
- Negative Binomial. Log-linear Poisson model with Gamma frailty using log(Exposure) as offset.
  - Ratio of rates is difference on log-rate scale.
  - Conditions on each patient's exposure through the offset (assumptions required).
  - Assume rate constant across period for each patient.

Current practice for Recurrent Event

- Time-to-first-event and Logrank test / Cox PH estimate.
  - Problems defining the estimand for time-to-event exist but are beyond this talk. But they are similar to those we see later.
- Negative Binomial. Log-linear Poisson model with Gamma frailty using log(Exposure) as offset.
  - Ratio of rates is difference on log-rate scale.
  - Conditions on each patient's exposure through the offset (assumptions required).
  - Assume rate constant across period for each patient.
- Anderson-Gill extension to Cox model. LWYY robust variance estimator.
  - Provides estimate of rate-ratio assumed constant across period. Nelson-Aalen can supply rate estimate for a single arm.
  - More complex frailty based models available.
  - Semi-parametric.

No intercurrent events

Trial with fixed observation time t_0 and no administrative censoring (complete data for all patients).

- Obvious estimand is the ratio of the mean observed event rate for each counterfactual A or B in period $[0, t_0]$.

$$\frac{\sum_i [N_i(A)]}{\sum_i [N_i(B)]}$$

- All current estimation methods are consistent for the estimand without further assumptions as long as there is no missing data.
  - The estimated rates from a sample in each arm are unbiased.
  - But for administrative censoring model assumptions are required for consistency.

Non-terminal intercurrent event

If the intercurrent event does not stop the patient being observable, then a series of approaches are possible

- **Treatment policy**
  - Effectively ignore the intercurrent event.
  - Just the same as last slide.
- **While on treatment**
  - Ignore subsequent recurrent events, and proceed as if event were terminal (see later).
  - For instance use of rescue or treatment termination.
    - Withdrawal of consent is similar to HbA1c example.
  - We have complete data for on treatment period.
    - Similar to using AUC over treated period in HbA1c example, rather than final visit (revised **variable**).
  - Worry is whether one is comparing like with like.

Terminal intercurrent event (IE)

For terminal IE we have two options

- **Hypothetical**
  Where a “scenario is envisaged in which the intercurrent event would not occur”
  - Assume some modified event rate after IE.
  - Add events before IE to those hypothesised after the IE.
  - Here we impute count in remaining period and add to the observed, rather than use correlation to impute final HbA1c.
  - But not suitable for death.
- Integrate the IE into the estimand.
  - **Composite** approach. e.g. score based system.
  - Specialized summary measure as follows.

Handling death as terminal event

- Any estimand needs to summarize across patients **and** across the period $[0, t_0]$ as well as compare between arms.

Handling death as terminal event

- Any estimand needs to summarize across patients **and** across the period $[0, t_0]$ as well as compare between arms.
- If the two event processes are independent then life is easy and we can focus on the recurrent event process conditional upon the time of the terminal event.
  But otherwise we need to consider the joint distribution.

Handling death as terminal event

- Any estimand needs to summarize across patients **and** across the period $[0, t_0]$ as well as compare between arms.
- If the two event processes are independent then life is easy and we can focus on the recurrent event process conditional upon the time of the terminal event.
  But otherwise we need to consider the joint distribution.
- As time progresses the marginal rate across patients may change because of
  - changing rates within patients,
  - selection of patients due to terminal event process.

Handling death as terminal event

- Any estimand needs to summarize across patients **and** across the period $[0, t_0]$ as well as compare between arms.
- If the two event processes are independent then life is easy and we can focus on the recurrent event process conditional upon the time of the terminal event.
  But otherwise we need to consider the joint distribution.
- As time progresses the marginal rate across patients may change because of
  - changing rates within patients,
  - selection of patients due to terminal event process.
- The recurrent event is likely to be related to the termination process.
  - If treatment impacts on the terminal event rate then what should the estimand represent?

Summary measure part of estimand
Two possible event rates

There are two suggested ways to summarize the event rate across patients who have differing times alive.

- **Equal-weighted Event Rate**: Calculate a rate for each subject and take a simple average of these across the patients.

$$\mathrm{E}_i\left[\frac{N_i}{T_i}\right] = \mathrm{E}_i\left[R_i\right]$$

- **Exposure-weighted Event Rate**: Count the total number of events and divide by the total exposure time.

$$\frac{\mathrm{E}_i\left[N_i\right]}{\mathrm{E}_i\left[T_i\right]} = \mathrm{E}_i\left[\left\{\frac{T_i}{\mathrm{E}_j[T_j]}\right\}R_i\right] = \mathrm{E}_i\left[W_i R_i\right]$$

N_i is the count in $[0, t_0]$ and T_i is smaller of t_0 and the terminating event time.
The symbol E_i indicates the mean over the population indexed by i, and $R_i = N_i / T_i$ while $W_i = T_i / \mathrm{E}_j[T_j]$ acts like a weight across subjects.

Comparison estimands

Taking the ratio of the event rate between treatments defines the estimand.

Equal-weighted Event Rate Ratio

$$\frac{\mathrm{E}_i\left[N_i(A)/T_i(A)\right]}{\mathrm{E}_i\left[N_i(B)/T_i(B)\right]} = \frac{\mathrm{E}_i\left[R_i(A)\right]}{\mathrm{E}_i\left[R_i(B)\right]}$$

Exposure-weighted Event Rate Ratio

$$\frac{\mathrm{E}_i\left[N_i(A)\right]/\mathrm{E}_i\left[T_i(A)\right]}{\mathrm{E}_i\left[N_i(B)\right]/\mathrm{E}_i\left[T_i(B)\right]} = \frac{\mathrm{E}_i\left[W_i(A)R_i(A)\right]}{\mathrm{E}_i\left[W_i(B)R_i(B)\right]}$$

where $W_i(A) = T_i(A)/\mathrm{E}_j[T_j(A)]$ etc.

The first focuses on patients while the second focuses on events.

Interpretation

Assume that the period of interest $[0, t_0]$ is one year.
The patient's interpretation compares between treatment options based on;

- **Equal-weighted Event Rate Ratio** :
  Ratio of the average rate of events a patient can expect while alive in the next year.

- **Exposure-weighted Event Rate Ratio**:
  Ratio of the average number of events over the next year or until death, whatever comes first, adjusted for any difference in mortality.

Plug-in estimators

Based on a randomized trial we can simply apply these formulae to the sample (indexed by $k \in$ A or B).

Equal-weighted Event Rate Ratio plug-in estimator

$$\frac{\sum_{k \in A} [N_k / T_k]}{\sum_{k \in B} [N_k / T_k]}$$

Exposure-weighted Event Rate Ratio plug-in estimator

$$\frac{\sum_{k \in A} [N_k] / \sum_{k \in A} [T_k]}{\sum_{k \in B} [N_k] / \sum_{k \in B} [T_k]}$$

- The former, while robust, may suffer from outliers caused by subjects terminating very early after a recurrent event. Especially important when the recurrent event includes instances of the terminal event.

Other estimators for Equal-weighted Event Rate Ratio

- Off-the-shelf estimators do not exist.
- Fit some Bayesian shared frailty model and sample sets of parameters.
  Then for each set derive estimand value using either integration or simulation.
  - One can do this for any summary statistic, so could apply to exposure-weighted as well.

Other estimators for Exposure-weighted Event Rate Ratio

- The Poisson log-linear model is a consistent estimator for this estimand in all circumstances.
  - Could use bootstrap to get correct standard errors.
- Classic Negative Binomial is not consistent for this estimand due to frailty aspect.
- Lin-Wei-Yan-Ying (LWYY) estimator seems to behave well under its reasonably wide assumptions.
  - Next talk demonstrates this using simulation.

Properties of summary statistic estimands in terms of models

- For any model for the population one can evaluate a summary statistic based estimand in terms of the parameters of the model.
  - But this often involves complex integration, so may need simulation to evaluate.
- By doing this for a series of parameter values one can plot the relationship to the estimand value.

Models used to define summary measures

- Also possible to define the estimand's summary measure as some function of the model parameters (ML estimates say) when applied to the whole population.
  - For instance, one could base an estimand on treatment effect in a log-linear Negative Binomial model.

- But will the E9 addendum concept of estimand encompass this? The industry needs a lead from the Working party on how far the "summary measure" part of the estimand principle can be adapted.

Conclusion

- Choosing an estimand makes you have to think very clearly.
- Estimands based on summary statistics are blunt instruments. You may need to model and interpret parameters when you want to understand the process.
- Very similar issues about summary statistic versus model parameter based summary measure hold for Time-to-Event.
- Does the E9 addendum need another round of public review?