Estimands: Statistical Perspectives.

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

ISCB Mini-symposium: Choosing Estimands in a Clinical Trial
July 2017.

Outline

Is this just a better way to express things? Or will designs and analyses need to change.
Some issues:

- Is the estimand defined in terms of observables or of parameters of distributions, such as mean?
- Impact of intercurrent event (IE) rate on value of estimand. Variation across regions.
- Attentuation. Impact of apparently reduced treatment effect.
- Missing data. Using partial data collected after intercurrent event (IE).
- Indirect estimation. Trial not matched to secondary estimands.
- Different implications for different disease areas. For some little may change.
  (e.g. Cuffe et al, 2015. Missing CD4+ cell response in randomized clinical trials of maraviroc and dolutegrair. Example where I am an author and we may need to think again.)

So what has changed?

The past:

- A model describes the distribution of outcome measure (perhaps multivariate) in terms of treatment and covariates. [ML or Bayesian]
- Interpetation is in terms of estimated value of a parameter in the model related to treatment difference.
- No desire to collect data after any intercurrent event (IE).

In between:

- The impact of intercurrent events (IE) handled as missing data. Often seen as if these were **hypothetical** estimands.
- Alternative estimands along **treatment policy** lines such as Jump-to-Reference (J2R) appear as “sensitivity analyses”, although in fact using a different estimand.
  Becomes clear that estimated values
  - are generally attenuated in value.
  - inherently depend in value on the rate of intercurrent events (IE).

So what has changed?

The new approach:

- A conceptual population where patients can take repeated values of the outcome interspersed with IEs.
- The estimand defines a single value for each subject dependent upon their outcome(s) and any associated IEs.
- We summarise these values across the nominal members of the population, using an average or median perhaps.

That defines the estimand.

- We estimate its value based on data from a sample of patients from a selected population, by evaluating each sampled patient's score using the the estimand rules and then summarize across all patients.

What could be easier?
Is this just a re-badging of the old ideas?

Core definition (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 encompasses things like AUC etc. to give one potential value for each subject.
- But are these observables, such as count of events? Or are they parameters, such as the expected number of events?
- I supect they need to be parameters. But most people seem to be thinking in terms of observables.

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 (2)

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

- Two concepts here. Summary is across patients but also involves treatment comparison.
- The value being summarized is effectively counterfactual as it is treatment A versus B, within the same patient. Either a mean for the patient or his observable.
- As we get to comparing rates between arms this gets more complicated.

Estimand is not the mean for a specific treatment regime, but a comparison such as difference or perhaps ratio.

How we estimate

- We define a specific population.
- We sample in some way from this population. In a quasi-random way. Classic issue here of whether trial matches real practice.
- Evaluate for each sampled patient allowing for IEs. The new bit.
- Summarise across our sampled patients, differencing by treatment arm.
- Estimate standard error (SE).
- Perhaps do this accounting for baseline covariate values.

But this assumes that the “variable” in B is an observable rather than a parameter, such as expected count.

- Using the patients mean for definition of estimand is easy.
  But estimand is defined in terms of counterfactual difference within patient. Sampled data comparisons have to cope with patient frailty.

Types of outcome

- Continuous (Normal) data summarises easily across patients. Few issues about margining.
- Binary and ordered categorical, include frailty issues often seen as overdispersion.
- Time to event is more complex (below).
- Recurrent event (below).

Time to event data

The intercurrent event (often death) becomes the outcome. What could be simpler?

- Classically comparison is described using **hazard ratio** based on proportional hazards (Cox) model.
- The summarization stage (D) implies we need some summary such as **risk** or **risk ratio**, which is a different thing to hazard ratio.
- If we try averaging hazard ratio across time how would it be weighted. e.g. Cox versus Wilcoxon.
  - Often proportional hazards assumption does not hold in practice.
  - Presumably want to give equal weight however many have been “censored”, to encompass the estimand philosphy.
- Restricted mean time survival time (RMST), or risk ratio for a specific, period will become more evident.
- Frailty issues presumably removed in estimand definition by final differencing within a patient before summarising.

Recurrent event data

- This is easier as we observe number of events within a subject over a fixed period.
- We can simply average total counts across patients.
- Again issues about frailty / over-dispersion.
- Early withdrawal creates issues about rate varying across time (see below).

Missing data

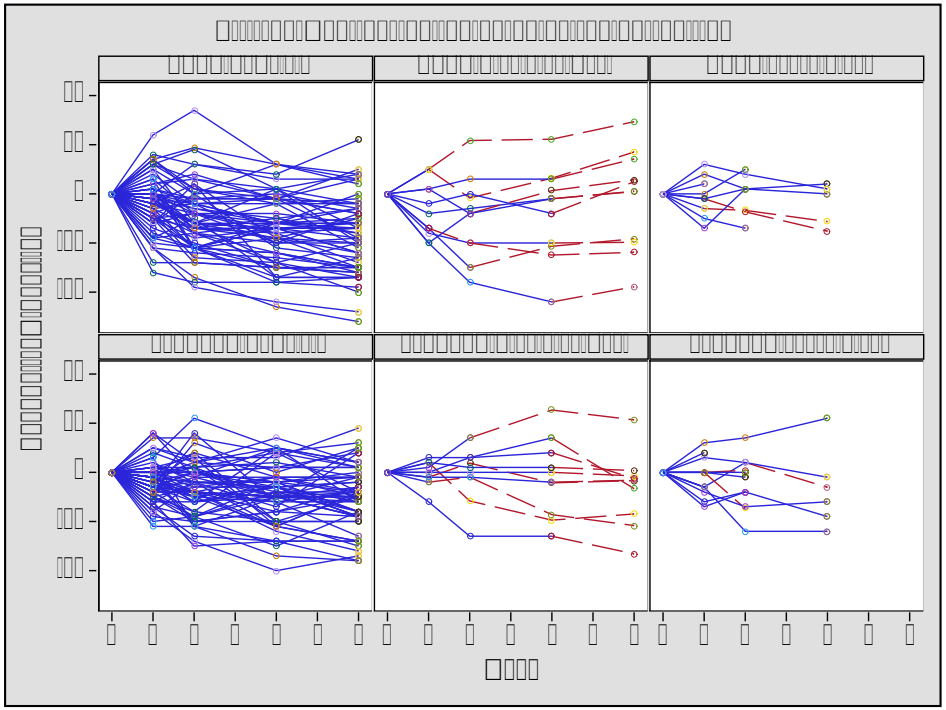
- These are values required for the within-patient summary that are not available for reasons other than IEs.
- If not related to patient abandoning the trial completely, then these will be rare, and some simple MAR imputation will often suffice.
- For patients who abandon trial (withdrawal), multiple imputation of their remaining data is an obvious route.
  - Imputation of remaining patient history may be need to allow for IEs after withdrawal.
  - Then apply routine estimand rules within this patient for each imputation.
- For some scenarios we can shortcut this process. Depends on handling of IEs.
- All imputation will be done completely within treatment arm, like classic MI rather than direct-likelihood approaches to MAR.
- Need to be clear about assumptions (e.g. MAR).

Missing data: Some strategies.

- Treatment policy strategy
  - Attempt to impute from **similar** patients.
  - May require modelling of IE process.
- Composite strategy
  - Dichotomizing may also simplify missing data issues as well as handle the IE.
    Beware loss of power.
- While on treatment.
  - Solves itself.
- Hypothetical.
  - MMRM and other direct-likelihood approaches seem useful.
    many tools already available.

Treament policy raises interesting statistical modelling issues.

Missing data: Repeated measures Normal.

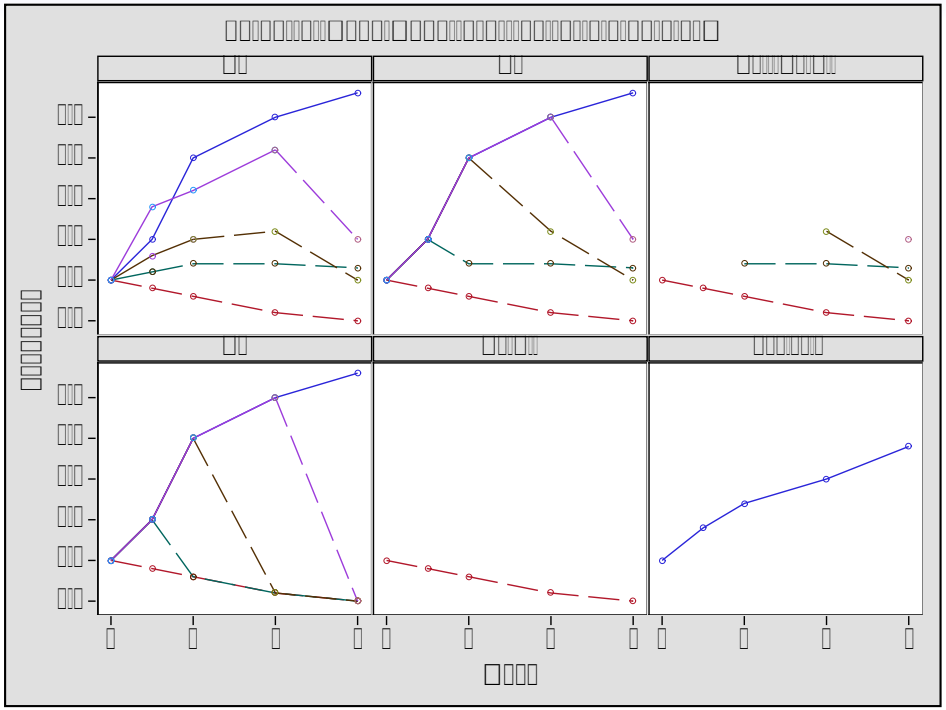


Special case:

- Treatment policy strategy.
- Patient can only have a single IE.
- Can only withdraw at some stage **after** an IE.

A randomly extended version of the DIA missing data example data set.

Some possible models.



- Model M4 is most complex with separate post-IE distributions for each visit where IE occurs.
- All except M2 can use the stepwise regression in proc MI.
- Model 2 requires regressing on previous residuals.
  Use the %MISTEP macro (from DIA working group web site at, www.missingdata.org.uk).

Missing data: Recurrent event data.

- Treatment policy estimand.
- Impute data based on the negative quadranomial to model counts in the three regions.
  - Before IE.
  - After IE up until withdrawal.
  - After withdrawal.
- Or use a Bayesian model with random subject effect (frailty) to impute.

Missing data: Other issues.

- Type 1 error and use of Rubin's rules to summarise across impuatations.
- Type 1 error is defined in terms of re-running the same study, with a potentially different number of IEs.
  - Non-Normality of the "variable" across sampled patients. We must not stratify analysis by IE.
- Is their a role for Bootstrapping?

Secondary estimands.

- Often these will not be matched to the trial design (chosen for primary estimand).
- These may involve the inclusion of information (rates of events etc.) from outside the study.
- May require use of hypothetical estimands to handle IEs. Potential role for methods like **reference-based imputation**.

Estimand: Other issues.

- Is the primary estimand the route for testing?

  Permutt suggested a **simple** robust test, followed by in depth estimation.

Attenuation

- Driven by desire for handling of IE to be “conservatiave” in some sense.
- But in some cases sensible estimand will increase treatment difference compared to naive ITT.
- Potential increase in sample size as SEDs not equally shrunk.
- Clinical interpretation. How do you interpet an average treatment effect that allows for patients not adhering?
- Implication for meta-analyses

Conclusion

- Lots of interesting questions for Statisticians to think about.
- Hopefully this is the end of “per-protocol” analyses.
- We can expect larger differences in treatment effect between regions due to differences in intercurrent event rates.