BBS Webinar "RCTs meeting causal inference: principal stratum strategy and beyond", 7th September 2020

Questions from the webinar chat addressed by the members of the organizing committee and panelists

The views expressed are the personal views of the panelists and members of the organizing committee and may not be understood or quoted as being made on behalf of or reflecting the position of the Agency or the company.

1) from Valentine Jehl: Any ideas on estimands related to long term safety effect when treatment switching is happening early?

Vanessa's reply post-meeting: It depends: if you think that even brief exposure before switching can have a long-term effect, you could consider an estimand contrasting "new treatment for at X months or more" versus "always standard treatment" (this would be similar to ITT); or you could formulate a "dose-response" estimand where you would expect an increasing risk of long-term problems the longer the new drug was taken before switching – this would be the estimand in a dose-response MSM and I would say you need to be quite confident in having measured time-varying confounding affecting treatment switching so to create plausible weights for IPW. If almost everyone switches early on, then positivity will be an issue and you will not be able to identify the effect of long durations of new treatment obviously.

2) from Marc:

Thanks Jack for the nice talk. Could these methods be applicable to vaccine development? With "vaccine" being the treatment; antibody response being the "biomarker" (may not happen in all treated); and actual immunity being the final outcome...

Jack's reply in the chat: Hi Marc, absolutely I don't see any reason why not. In the tutorial paper I plan to mention this as a potential application

3) from Anh:

The fitted model based on flex dose data in the flex trial is also random. Fitting this model to estimate outcomes for patients in the (flex) placebo arm would also incur uncertainty => would this be offset by the gain in number of patients in the end so that the overall uncertainty is reduced?

Kelly's reply in the chat: Hi Anh, thank you for your question. I hope I understand your question well. By making predictions in both arms by fitting a model in flex arm, we

usually gain precision relative to just using the data in the flex arm. If the outcome model (in flex) is correct, asymptotically there is even no cost of estimating the model coefficients.

4) from Jack: Kelly slide 13, it looks like the doubly miss-specified model performs best in terms of bias and precision?

Kelly: In this table I don't show a setting where all models (selection and both outcome models) are misspecified. The proposed estimator is unbiased if at least one of these models is correctly specified, while the G-computation based estimator is only unbiased if the outcome model is correctly specified. Note that the G-computation estimator is slightly more efficient than the proposed estimator when the outcome models are correct.

5) from Jonatan: Hi Kelly, if both models in the doubly robust estimation are wrong is your experience that the bias is worse or similar to e.g. g models?

Kelly's reply in the chat: Thank you, Jonatan, for the interesting question! These doubly robust estimators can sometimes perform worse in terms of bias when both models are misspecified; for example in settings with very extreme weights. Problems can be prevented by using special, tailored estimators of the outcome and selection models. Happy to send some references if of interest.

6) from Jonatan:

Interesting talk Dominik. Could you give some more detail on how you decided on the time point for the landmark analysis? Do you think there is a risk for selection bias caused by patients needing to survive until the landmark time point?

Dominik in the chat (edited postmeeting for clarity): @Jonatan: Great question! This indeed is not trivial. Some points one need to consider: (1) it need to be early enough to allow (1.1) any clinical action to be taken if needed (1.2) and also that not too many patients are excluded because of an event prior to the landmark time point as one underlying assumption is that treatment assignment is independent of ADA, missing and covariates, so for this LM pop needs to be close to ITT. (2) you need to catch a large proportion of the overall treatment-emergent ADA-positives at the landmark in order to have meaningful information from the landmark analysis, which for our application was the case as most ADA+ were identified at the early administrations of the drug. Other considerations also encompasses planned ADA schedule actual timing of assessment in the trial.

Kaspar in the chat: @Jonatan: You raise a relevant point: the primary endpoint event might act as a competing risk to observing the intercurrent event, potentially leading to

immortal bias when utilizing naive analyses. This aspect is briefly discussed in Section 4.7 in the paper Björn mentioned: https://arxiv.org/pdf/2008.05406.pdf

7) from Jack: Hi Andrew, the baseline covariate that differentially predicts treatment adherence is necessary to test the homogeneity assumption, but is not necessary to identify the hypothetical estimand using randomization as an IV under homogeneity.

I agree that finding a baseline covariate will be challenging. The alternative is to assume you have measured all coviariates that confound biomarker response (or the intercurrent event more generally). But one thing we all agree on is that design trumps analysis, and from my perspective the IV approach sticks closest to this principle compared to methods that invoke no unmeasured confounder assumptions. This approach is really like analysing an RCT like an observational study.

Fabrizia Mealli: @Jack: we need to expand on IV, as ERs cannot be used in settings like censoring due to death or switching for example, but sometimes certain covariates may play a similar role as an IV.

Vanessa Didelez: I think in an RCT with intercurrent event we have the best chance to make "no unm. conf." plausible, because switching etc are decisions, and reasons for these decisions can be recorded, and patients are under much closer observation than in an obs. study - so by collecting the right info, you can be much more confident in that assumption than with obs. data.

Jack @Vanessa: good point I agree.

Fabrizia @Vanessa: completely agree, crucial is the collection of baseline covariates that are predictive of the intercurrent events.

Further comments:

@Fabrizia, I have the impression people are using two different definitions of princ. strat. one is the usual "cross-world" one (e.g. "compliers"), but there is also a "single-world" one (e.g. “patients who would be ADA+ under treatment”) - I wouldn't even call the latter a princ.strat. - it is like effect of treatment on the treated - what do you say?

from Fabrizia Mealli:

@Vanessa: "single-world" is also a principal stratum, it is the union of principal strata. For example in the simple noncompliance setting the treated is the union of C and AT; people something confuse the "basic" rpincipal stratification with principal stratification where a principal stratum can be any union of basic principal strata.

@Kelly and @Bjoern: I think that Kelly's transportation assumptions is a type of principal ignorability assumption, and the subgroup is indeed a principal stratum that is only observed in the flexible dose trial.

The transportability assumption we are making is the following: Y(low dose) \indep T|X with T trial indicator and X baseline covariates. This does not involve the intercurrent event "switching". In particular, this is an assumption you would always make if you transport data from one trial to another trial (or target population). I would therefore not view this a type of principal ignorability assumption, although they are both assumptions about conditional exchangeability.