Life is not perfect: Noncompliance Analysis in practice

Els Goetghebeur,Krista Fischer and Stijn Vansteelandt
Ghent University, Applied Mathematics and computer science
Krijgslaan 281-S9
B-9000 Ghent, Belgium
els.goetghebeur@rug.ac.be

1. Introduction

Assigning exposures in a randomized fashion is the most promising avenue for establishing their causal effects. Human beings however rarely follow a given assignment perfectly. Partial compliance is a reality and tends to be associated with natural disease progression and/or perceived effect of treatment. It begs the question how the different achieved exposures relate to potential treatment free outcome, and what their effect has been - or could be in a more general population. To answer questions of effect, a growing range of exciting statistical methods has been proposed over the last decade. Nevertheless in practice as well as in theory, one must confront challenges which the simple 'Intent-to-treat analysis' manages to avoid. In this talk we discuss structural mean and distribution models for the effect of actual exposure in a randomized setting. We are guided by three clinical trials to illustrate several challenges and present some solutions. Specific questions concern the selection of appropriate structural models, the optimization of precision, and also diagnostics, sensitivity and robustness. First we give a brief description of the trials.

2. Three clinical trials

1. **Blood pressure reduction**
   In a randomized placebo-controlled trial of blood pressure reduction patients are assigned one pill per day of nebivolol, atenolol or placebo, after a 4 week run-in on placebo. Compliance is electronically monitored over the active period yielding an exact record of the times at which individual pill containers have been opened. One is interested in the effect of 'percentage of active treatment actually received' on diastolic blood pressure reduction, measured continuously or as a binary indicator of treatment success.

2. **Prevention of HIV transmission**
   A recent multi-center placebo-controlled trial conducted in Africa and Asia studied a vaginal nonoxynol-9 (N9)gel for the prevention of male to female transmission of HIV. Unexpectedly, a significant *increase* in HIV incidence was observed when comparing the treatment arm with the placebo gel. Because the N9-gel is worldwide on the market as a

spermicidal gel in condoms, one hopes that the increased risk is dose-related and confined to the highly active study population of prostitutes. To support this hypothesis-or not- a measure of treatment effect in function of observed treatment exposure is needed. Several (imperfect) measures of exposure have been recorded: diaries of sexual acts, counts of returned tubes of gel (used as well as unused ones) and interviews at clinic visits about recent exposures. HIV testing happens 4-monthly leaving the primary outcome interval-censored. Incidence of STDs (Sexually Transmitted Diseases) forms a secondary end point of interest. In practice, there is a large variation in actual timings of the visits and a serious drop-out rate.

3. **Anti-depressants**
   One is comparing two anti-depressant drugs in a randomized trial after a run-in period during which all patients receive placebo. Each drug is to be taken once a day, but both drugs have very different pharmacokinetic half lives. With psychiatric disorders one expects a high drop-out rate from treatment as well as varying compliance while persistent with the treatment. There is great potential for compliance being dependent on outcome (the patients mood). Compliance is electronically monitored from the start of the study, and the outcome regularly assessed through validated scores. To predict drop-out of treatment as well as compliance (while on treatment) a baseline questionnaire has been designed.

3. Different challenges

1. The first challenge lies in formalizing the causal parameters to be estimated. Typically this is expressed in 'what if' terms following the potential outcomes formulation of Rubin (1978) and Holland (1986). Suppose one agrees that a particular contrast (e.g. difference in means) between a potential treatment-free and treated outcome is of interest after controlling for a set of baseline covariates. Does one estimate the effect of treatment specifically for the subgroup who received a given treatment level, or does one condition additionally on the treatment-free outcome that would have been achieved and/or the compliance with placebo (the other) assignment that would have been established? The blood pressure data suggests for instance that untreated low blood pressures can hardly benefit from treatment whereas the high blood pressures seen in the control arm are absent from the treatment arm. Generally, finer conditioning yields more detailed information but requires stronger modeling assumptions. Moving in the opposite direction, one sometimes extrapolates from those (subsets) who received a given dose to estimate effects for all who could possibly receive it. Variables and parameters that capture this effect are labeled 'counterfactual'. G-estimation and marginal models have been developed by Jamie Robins to make this possible. We do not consider this further in this talk.

2. Having established the study group to which the effect measures will pertain, one must identify summaries of exposure as relevant drivers of causal effects. Ideally clinicians, pharmacologists and statisticians collaborate on this. In doing so, one should keep parsimony in mind. Indeed, one is typically dealing with high dimensional measures of dose-timing. Furthermore, the potential outcomes formulation involves latent variables. Hence relatively little information can be abstracted and precision is a big issue.

3. The efficiency of structural estimators depends on the ability to predict exposure (and treatment free outcome) from baseline. Designing studies that measure such predictors is important. In the anti-depressant trial, it is however not obvious how to best incorporate the baseline questionnaire in the structural equations. Patterns of compliance are complex, with some patients dropping out of treatment altogether at non-scheduled time points, and others showing varying degrees of dose-timing depending on how they feel. When the exposure measure shows high variability over the study population and is well predicted from baseline data, its effect will be most accurately estimated from structural models. In addition, the potentially strong feedback from input to output calls here for more detailed dynamic causal modeling. Random effects structural models may help provide an answer and have now been developed. Structural nested and marginal models are alternatives.

4. The interpretation and estimation of causal parameters in a chosen *linear* structural mean model for a (placebo) controlled study is becoming well understood (Robins, 1994, Goetghebeur and Lapp, 1997). Loglinear models and structural accelerated failure time models have also been documented in the literature. Recently, we have proposed estimators for structural generalized (linear) mean models and for structural proportional hazards models for survival. We will illustrate how structural generalized linear models can be applied to find the effect of observed dose on treatment success in the blood pressure study. In contrast to the structural linear models, the generalized linear estimators require a model for the selection effect. We show how a special choice in a class of weighted estimating equations yields structural estimators for the causal odds ratio which is robust, under the null hypothesis of no treatment effect, to misspecification of the selection model. These new additions open up a wider choice of modeling and model selection techniques are called for.

5. **Diagnostics, sensitivity and robustness**
   In the context of structural linear mean models adjusting for treatment free outcome we propose to estimate structural variance components to help guide the choice between richer and poorer models.

   In general, because (semi-parametric) structural modeling involves latent variables, model

checking based on observable dimensions is far from obvious. Graphical model checking tools exist. More formal tools which focus on model components at risk of driving conclusions would be very useful. Robustness and sensitivity analysis become important points to consider. Some recent work on this will be presented in this session.

Compliance measures are never perfect and potentially quite error prone. Particularly in the N-9 study a substantial amount of error in the exposure measure(s) is anticipated. There is however opportunity to evaluate the quality of the exposure measure and hence incorporate its uncertainty in the structural estimator.

4. In conclusion

Structural mean and distribution models for estimating the effect of observed exposure in randomized settings are outgrowing puberty. Ever more flexible models are being developed as well as many of the tools available for more classical inference. Besides an enormous potential, structural modeling also faces its limitations in this setting. The practical problems remain immensely challenging and many more exciting theoretical and practical developments are to be expected in the future.

REFERENCES

Goetghebeur, E. and Lapp, K. (1997). The effect of treatment compliance in a placebo-controlled trial: regression with unpaired data. *Applied Statistics*, **46**, 351–364.

Holland, P.W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, **81**, 945–970.

Keiding, N., Filiberti, M., Esbjerg, S., Robins, J.M. and Jacobsen, N. (1999). The graft versus leukemia effect after bone marrow transplantation: A case study using structural nested failure time models. *Biometrics*, **55**, 23-28.

Robins, J.M. (1994). Correcting for non-compliance in randomized trials using structural nested mean models. *Communications in Statistics- Theory and methods*, **23:** 2379–2412.

Rubin D.B. (1978). Bayesian Inference for causal effects: the role of randomization. *Annals of Statistics* **7:**34–58.

RESUME

On considère l'analyse structurelle (en moyenne et en distribution) pour l'estimation des effets d'expositions observées dans les essais cliniques randomisées. A travers trois essais cliniques on parcourt un nombres d'obstacles pour arriver a des mesures d'effet qui sont bien définies, sensées et estimables avec suffisamment de précision.