Estimation of Principal Stratum Effects, an Overview and Potential Applications in Oncology

Björn Bornkamp

on behalf of the causal subteam

Oncology Estimand Working Group

DAGStat Conference 2019, München

18 March 2018

Estimands in Oncology WG

- initiated and led by Evgeny Degtyarev (Novartis) and Kaspar Rufibach (Roche), first TC Feb 2018
- main purpose: ensure common understanding and consistent definitions for key estimands in Oncology across industry
- 31 members (14 from Europe and 17 from US) representing 19 companies
- established as EFSPI SIG for Estimands in Oncology in Nov 2018
- close collaboration with regulators from EMA, FDA, China, Taiwan and Canada

Estimands in Oncology WG

Communication plan for 2019

- whitepaper(s) and presentations at statistical and clinical conferences
- plans to further engage with Clinical community beyond ASCO

DAGStat (Munich)

Session with 4 WG talks

LiDS (Pittsburgh)

Session with 3 WG talks + EMA discussant

PSI (London)

2 WG talks

ISCB (Leuven)

2 abstracts submitted

ASA Biopharm Section Regulatory-Indus Statistics Workshop (Washington)

Session with 2 WG talks + FDA talk

MAR | APR | MAY | JUN | JUL | AUG | SEP

ASCO (Chicago)

3 abstracts submitted in collaboration with KOLs and industry clinicians

DIA (San Diego)

1 WG talk

JSM (Denver)

Session with 4 WG talks + FDA discussant

ASCO: American Society of Clinical Oncology
LiDS: Lifetime Data Science (ASA Section)

Causal-Subteam

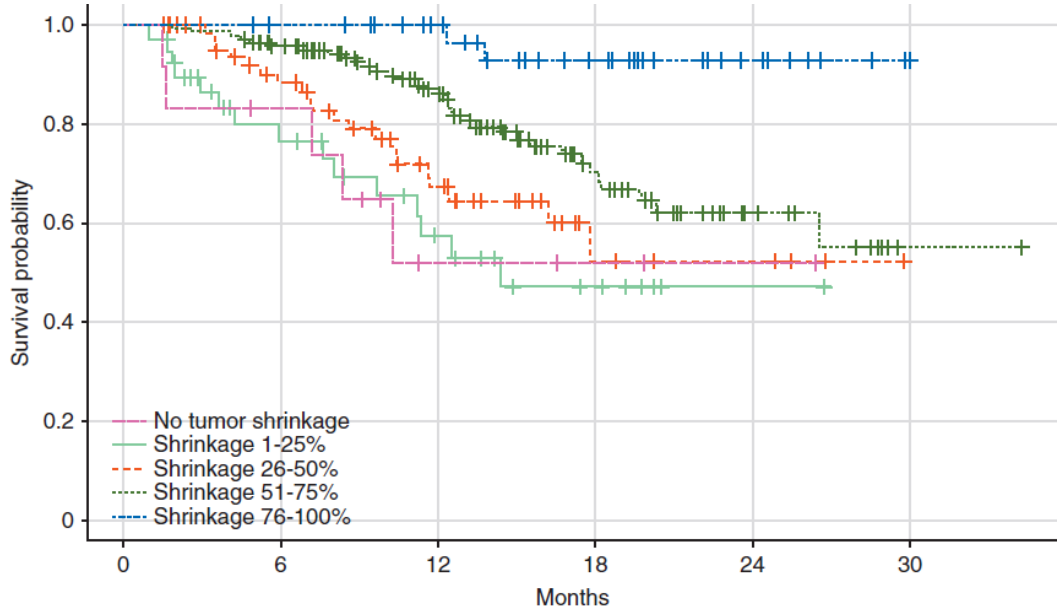
- Kaspar Rufibach (Roche), lead
- Vera Beckers (Abbvie)
- Björn Bornkamp (Novartis)
- Audrey Boruvka (Roche)
- Andreas Brandt (BfARM)
- Marie-Laure Casadebaig (Celgene)
- Feng Liu (AstraZeneca)
- Yi Liu (Nektar)
- Juliane Manitz (EMD Serono)
- Emily Martin (EMD Serono)
- Devan Mehrotra (Merck)
- Alan Phillips (ICON)
- Satrajit Roychoudhury (Pfizer)
- Anja Schiel (NoMA)
- An Vandebosch (Janssen)

Overview

- Clinical questions
- Estimation of principal stratum effects
- Criticisms
- Summary

Clinical questions

- Short term tumour shrinkage versus overall survival
  - Short term tumour shrinkage can be a good predictor of overall survival
- What is the treatment effect versus control (on overall survival) in patients that have a tumour shrinkage < X % at Y weeks if on the investigational treatment?



Clinical questions

- Biologic treatments and antidrug antibodies (ADA)
  - For biologic treatments (e.g. cancer immunotherapies) ADAs might form and may (or may not) have a neutralizing effect on the treatment
- What is the treatment effect versus control (e.g. on overall survival) in patients that develop ADAs if on the investigational treatment?
  - NB: The control might be a non-biologic drug (i.e. ADAs will not form)

Clinical questions

- Cancer prevention trial
  - Do patients that develop cancer (if on investigational treatment and if on control treatment) have a different cancer severity than if given the control treatment?

- Treatment switching
  - What is the treatment effect in patients that do not switch (if on investigational and if on control treatment)?

Clinical questions

- Commonality
  - Evaluate the treatment effect in the subgroup of patients where a specific post-randomization event would (or would not) occur
  - Challenge: Post-randomization event itself may be affected by treatment
    - Randomization cannot be relied upon to ensure comparable groups on investigational treatment and control → Selection bias

Clinical questions

- Commonality
  - Evaluate the treatment effect in the subgroup of patients where a specific post-randomization event would (or would not) occur
  - Challenge: Post-randomization event itself may be affected by treatment
    - Randomization cannot be relied upon to ensure comparable groups on investigational treatment and control → Selection bias
- Class of questions is quite frequent in oncology
  - See Section 7.6.5 of the EMA anticancer guidance on „Analyses based on a grouping of patients on an outcome of treatment“
    - Highlights problematic nature of naive analyses
    - Encourages search for „unexpected findings“ based on such exploratory analyses (by each treatment arm; not formally comparing arms due to non-randomized nature)

Principal Stratification Estimands

- Concept introduced in Frangakis & Rubin (2002)
  - Conceptual idea: Potential outcomes S(0) and S(1) are unaffected by treatment
  - Determine treatment effect in subset(s) (principal strata) of population defined by S(0) and S(1)

S – occurence of postbaseline event
S(0) – potential outcome control
S(1) – potential outcome treatment

- Can classify every patient in one of these four cells (= principal strata)
  - E.g. S is tumour shrinkage < X % at Y weeks or presence of ADAs

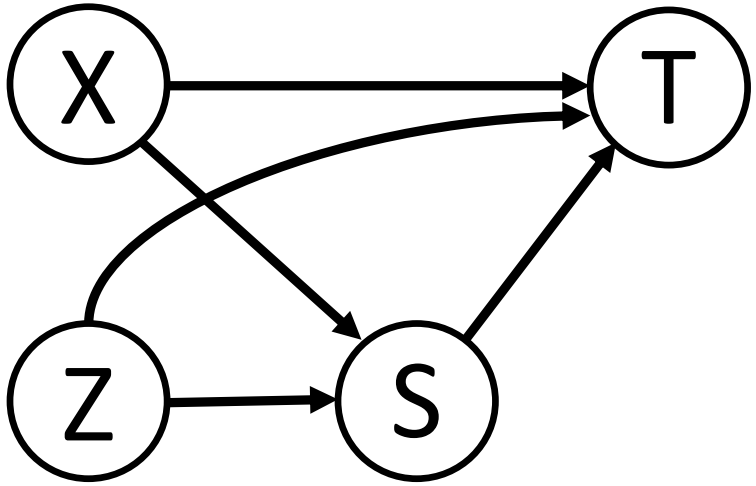
	S(0) = 0	**S(0) = 1**
S(1) = 0		
S(1) = 1		

- Note: One of S(0) or S(1) is observed for an individual patient the other unobserved

Principal Stratification Estimands

- By itself the principal stratum formulation does not provide a solution
  - Just a way of framing a particular problem
  - But: Provides a clear inferential target (treatment effect in principal strata)
  - Easier to discuss assumptions etc if inferential target is clear
- Determination of treatment effects in strata requires assumptions!
  - E.g. Principal stratum membership is not observed
- Let‘s illustrate with the ADA example in more detail

Consider ADA example in more detail



X – Baseline Characteristics
Z – Treatment
Z=0 control, Z=1 treatment
S – Development of ADAs
T – Survival time

- Quantity of interest?

- Survival time distributions for patients under treatment and control that develop ADAs, if taking treatment (S(1) = 1)
- In potential outcome notation: Compare

$$T(1) \mid S(1) = 1 \text{ versus } T(0) \mid S(1) = 1$$

Potential outcomes
T(z) – Potential survival time
S(z) – ADA presence post-baseline

ADA example

- In potential outcome notation: Compare

$$T(1) \mid S(1) = 1 \text{ versus } T(0) \mid S(1) = 1$$

  - e.g. estimate survival functions $P(T(1) > t \mid S(1) = 1)$ and $P(T(0) > t \mid S(1) = 1)$ and derive a difference based on those

ADA example

- In potential outcome notation: Compare

$$T(1) \mid S(1) = 1 \text{ versus } T(0) \mid S(1) = 1$$

  - e.g. estimate survival functions $P(T(1) > t \mid S(1) = 1)$ and $P(T(0) > t \mid S(1) = 1)$ and derive a difference based on those
- Rather easy to derive an estimate for $P(T(1) > t \mid S(1) = 1)$: This was observed on the treatment arm

ADA example

- In potential outcome notation: Compare

$$T(1) \mid S(1) = 1 \text{ versus } T(0) \mid S(1) = 1$$

  - e.g. estimate survival functions $P(T(1) > t \mid S(1) = 1)$ and $P(T(0) > t \mid S(1) = 1)$ and derive a difference based on those
- Rather easy to derive an estimate for $P(T(1) > t \mid S(1) = 1)$: This was observed on the treatment arm
- How to derive estimate of $P(T(0) > t \mid S(1) = 1)$?
  - Unclear whether patients on the control arm would have developed ADAs if given treatment
  - Even worse: No patient on control will develop ADAs (i.e. $S(0) = 0$ for all)
  - No one-size-fits-all solution in the Frangakis and Rubin (2002) paper

ADA example: Full Bayesian estimation

- We know that

$$p(T(0)) = \pi\, p(T(0) \mid S(1) = 1) + (1 - \pi)\, p(T(0) \mid S(1) = 0)$$

where $\pi = P(S(1) = 1)$ can be estimated from the treatment arm

ADA example: Full Bayesian estimation

- We know that

$$p(T(0)) = \pi\, p(T(0) \mid S(1) = 1) + (1 - \pi)\, p(T(0) \mid S(1) = 0)$$

where $\pi = P(S(1) = 1)$ can be estimated from the treatment arm

- Densities (or parameters describing the densities)
  - $p(T(0) \mid S(1) = 1)$ and $p(T(0) \mid S(1) = 0)$ are not identified based on the data

  → even for „infinite“ sample size, likelihood will not contract to a single point

ADA example: Full Bayesian estimation

- We know that

$$p(T(0)) = \pi\, p(T(0) \mid S(1) = 1) + (1 - \pi)\, p(T(0) \mid S(1) = 0)$$

where $\pi = P(S(1) = 1)$ can be estimated from the treatment arm

- Densities (or parameters describing the densities)
  - $p(T(0) \mid S(1) = 1)$ and $p(T(0) \mid S(1) = 0)$ are not identified based on the data

  → even for „infinite“ sample size, likelihood will not contract to a single point
- For a proper prior also the posterior will be proper
  - For some parameters more information might be available for others less
    → Need to evaluate impact of „weakly-informative“ priors carefully
  - See Magnussen et al. (2018) for a related approach/application

ADA example: Utilizing covariates

- Assume one can find all covariates X such that
  - Conditional on covariates X, T(0) and S(1) are independent: $T(0) \perp S(1) \mid X$
  - Principal ignorability, see Ding et al. 2017, Feller et al. 2017
  - If this is true the conditional distribution $p(T(0) \mid S(1), X) = p(T(0) \mid X)$

ADA example: Utilizing covariates

- Assume one can find all covariates X such that
  - Conditional on covariates X, T(0) and S(1) are independent: $T(0) \perp S(1) \mid X$
  - Principal ignorability, see Ding et al. 2017, Feller et al. 2017
  - If this is true the conditional distribution $p(T(0) \mid S(1), X) = p(T(0) \mid X)$
- Estimation
  - $p(T(0) \mid X)$ can be estimated on the control group, and averaging with respect to $p(X \mid S(1) = 1)$ provides an estimate of $p(T(0) \mid S(1) = 1)$ (standardization)
  - Alternative estimation strategies
    - Build a model for S(1) = 1 on the treatment arm (depending on X), and multiply impute S(1) for the control arm → Combine estimate with Rubin‘s rules
    - Matching on X and „standard“ analysis

ADA example: Utilizing covariates

- Case-specific whether one would be willing to make this assumption
  - Principal ignorability is an untestable assumption (independence assumption „across worlds“); sensitivity analyses possible, see Ding et al. (2017)
  - If S(0) would be predictive of S(1) further analyses/assumptions would be possible → in this case as S(0) = 0 for all patients

Criticisms

- Complication: Benefit-risk analyses for principal strata
  - Typical analysis strategies do not clearly identify the population of patients in the principal stratum. How to perform safety analyses?
- Hernán & Scharfstein (2018)
  - “... subgroup that cannot be clinically identified ...”
- Scharfstein (2018)
  - „... Principal stratification is scientifically interesting but just too assumption-laden to be primary ...“
  - “... Lowers the level of evidence. ...”
- Also controversially discussed in the causal inference community
  - See Pearl (2011) → Principal stratification overused

Summary

- Clinically relevant questions
  - Sometimes assumptions too strong to answer based on the data at hand
  - But: Incorrect (& potentially mis-leading) analyses are already performed for these questions → utilizing causal inference techniques will raise the level of discussion on the questions and possible assumptions
- Due to assumptions required for identification, the principal stratum strategy might not often be part of the primary estimand
- Could still be important to contribute to an “overall” picture of the drug’s properties

References

- Ding, P. and Lu, J. (2017) Principal stratification analysis using principal scores." Journal of the Royal Statistical Society: Series B 79: 757-777.
- Feller, A., Mealli, F. and Miratrix, L. (2017) Principal score methods: Assumptions, extensions, and practical considerations. Journal of Educational and Behavioral Statistics 42: 726-758.
- Frangakis, C. E. and Rubin, D. (2002) Principal stratification in causal inference. Biometrics 58: 21-29.
- Hernán, M. A. and Scharfstein, D. (2018) Cautions as Regulators Move to End Exclusive Reliance on Intention to Treat. Annals of internal medicine 168: 515-516.
- Magnusson, B., Schmidli, H., Rouyrre, N. and Scharfstein, D. (2018) Bayesian inference for a principal stratum estimand to assess the treatment effect in a subgroup characterized by post-randomization events, arXiv:1809.03741
- Pearl, J. (2011) Principal stratification -- a goal or a tool? The International Journal of Biostatistics 7: 1-13.
- Scharfstein, D. (2018) A (Constructive/Provocative) Critique of the ICH E9 Addendum, Presentation given at 11th Annual Conference on Statistical Issues in Clinical Trials Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania https://www.cceb.med.upenn.edu/sites/default/files/uploads/cci/2018_Clin_Trials/ScharfsteinD%20materials.pdf