MSR158

Successful Use of Propensity Score Methods for HTA in Germany: A Near-Impossible Task?

Mona Bierl[1], David Niederkofler[1], Sara Hogger[1], Stefanie Wüstner[1]

[1]AMS Advanced Medical Services GmbH, Mannheim, Germany

BACKGROUND

Marketing authorization often relies on single-arm trials, especially for orphan drugs. However, German health technology assessment (HTA) bodies **G-BA and IQWiG rarely consider non-randomised evidence**.

The G-BA can request a **routine practice data collection** (anwendungsbegleitende Datenerhebung, **AbD**), to gather real-world data. AbD should provide **comparative evidence for HTA assessments**.

AbDs are non-randomised. Thus it is challenging to control for confounding factors. **Propensity score (PS) methods** are the preferred approach to effectively account for them.

OBJECTIVE

We aim to **evaluate the viability of PS methods** for AbD facing the stringent requirements of HTA bodies in Germany.

METHODS

DATA SIMULATION

We generated **AbD** data that represent **"ideal world" conditions**:

- Outcome Overall Survival (OS) fully observed (**no censoring**).
- Treatment and control arm in a **1:1 ratio**.
- **All confounders measured** (20% continuous, 80% categorical)
- Weak link between confounders and outcome
- **Confounder distributions similar between treatment arms** and no correlation between confounders

We considered **different scenarios** with two true treatment effects (HR=0.25 and 0.45), varying sample sizes (100-1000 patients), and varying number of confounders (5, 10, 15 confounders).

ANALYSIS METHODS

We assume a substantial added benefit of the treatment on OS and aim to confirm that benefit with **different PS methods**, using similar approaches to available AbD protocols:

- **1:1 nearest neighbour matching** on logit-PS with caliper 0.25 (PSM)
- Different **PS weighting methods**, primarily Standardised Mortality Ratio (SMR) weighting

If at least one confounder was not balanced between groups (standardised mean difference (SMD) ≥0.25) or PS overlap was <50%, the data was trimmed. HR for OS was estimated using a Cox model. For each scenario, we estimated the proportion of rejections of the shifted null-hypothesis H_0: HR=0.50 specified by IQWiG in previous AbDs.

How often can we show an added benefit?

RESULTS

SIMULATED DATA: IDEAL WORLD

The majority of confounders were balanced and the overlap between PS distributions is large before the data was matched or weighted (Fig 1).

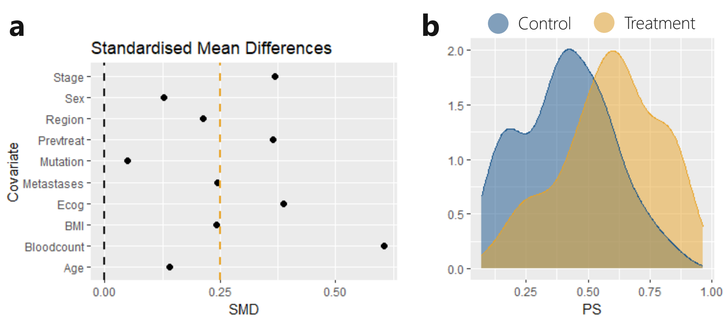

Fig 1: Simulated AbD data with ideal world conditions

ADDITIONAL SIMULATION: REAL WORLD?

Increasing the imbalance of a binary confounder (Fig 2a) led to a pronounced difference in PS distributions and reduced the overlap significantly (Fig 2b).

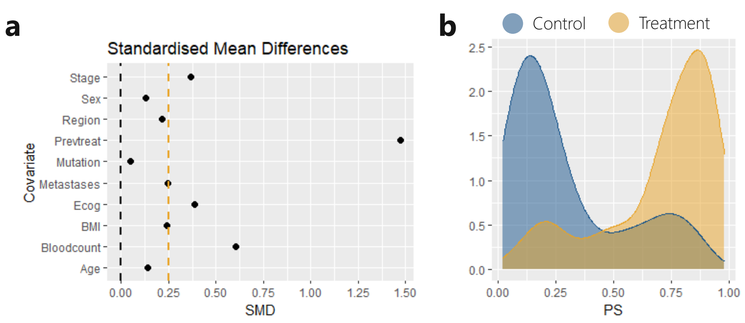

Fig 2: Simulated AbD data with an imbalanced confounder

Fig 1+2: SMD for each covariate (a) and PS distribution overlap (b) for example data (n=200 observations total, 10 confounders)

EVEN IN AN IDEAL WORLD A STRONG TREATMENT EFFECT IS DIFFICULT TO DEMONSTRATE IN ABD DATA WITH LOW SAMPLE SIZES AND MULTIPLE CONFOUNDERS

Given a true treatment effect of HR=0.25 (Fig 3, upper quadrants), the rate of rejected H_0 rises with increasing sample size and decreases with a higher number of confounders.

For typical AbD sample sizes (n≤200), the rejection rate is below 0.4 for 15 confounders, below 0.65 for 10 confounders and below 0.9 for 5 confounders using PSM. Thus it is challenging to demonstrate treatment effects, even in an ideal world setting. PSM tends to lead to higher rejection rates compared to weighting, but can also drastically reduce sample size, which complicates interpretation of results.

For a true treatment effect of HR=0.45 (Fig 3, lower quadrants), even increasing the sample size to n=1000 leads to a mean rejection rate of less than 5% for 10 or 15 confounders. Hence, it is unlikely that a treatment effect can be demonstrated with an AbD under these conditions.

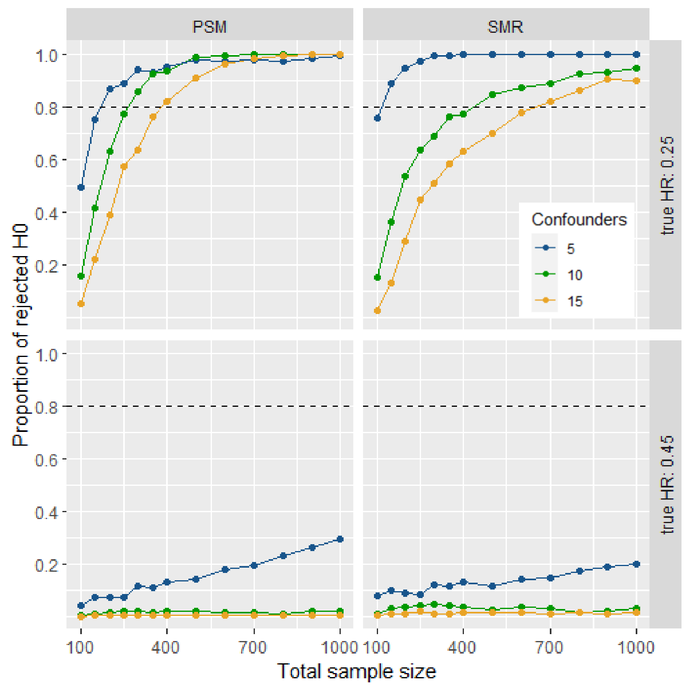

Fig 3: Proportion of rejected H_0 by true HR, PS method (PSM or SMR) for ideal world AbD data. Unbalanced covariates after use of PS method counts as failure to reject H_0.

ONE UNBALANCED CONFOUNDER CAN HEAVILY AFFECT THE CHANCE OF DEMONSTRATING A TREATMENT EFFECT

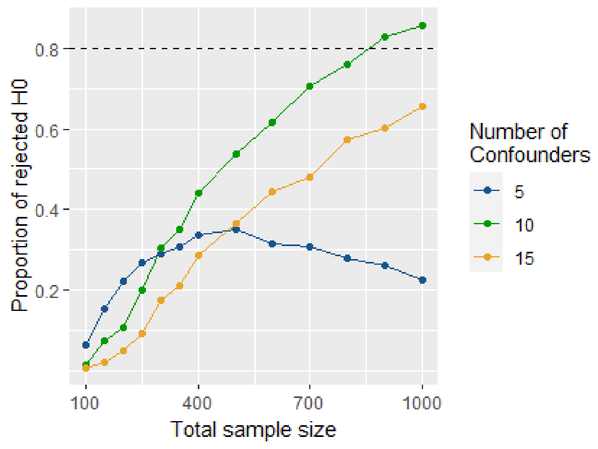

Fig 4: Proportion of rejected H_0 for data with one unbalanced confounder (see Fig 2a) and true HR 0.25 using PSM

One strongly unbalanced confounder reduces the rejection rate (Fig 4) and leads to reduction of sample size by about 60% after matching. In this scenario, the treatment effect is rarely demonstrated for small sample sizes and 10 or more confounders.

CONCLUSION

Even with ideal world data it is difficult to reject the dramatically shifted H_0 for smaller sample sizes. With a true effect close to the H_0 or more heterogenous data this becomes almost impossible.

AbD reality is far from the ideal world: small sample sizes, large number of potential confounders, possibly unobserved confounders, and unequal confounder distributions between treatment groups make requirements regarding PS balance and overlap hard to fulfil.

Data trimming and large sample size reduction to achieve balance makes the interpretation of results difficult. The population on which the effect is shown may be ill-defined.

In short: with the stringent requirements regarding PS methods, the probability to successfully demonstrate an added benefit in the AbD setting is low.

Come visit us!
www.ams-europe.com

Mona Bierl
+49 172 475 37 86
Mona.Bierl@ams-europe.com

AMS
Advanced Medical Services