Propensity-Score Based Methods: An Application to Data on Cognitive Function in the Elderly

Schröder H[1], Caputo A[1], Debling, D[2], Stürmer T[3]
[1]Institut für Medizinische Biometrie und Medizinische Informatik, Universitätsklinikum Freiburg, Deutschland
[2]Psychologisches Institut, Universität Heidelberg, Deutschland
[3]Division of Pharmacoepidemiology and Pharmacoeconomics, Brigham and Women's Hospital, Harvard Medical School, Boston, MA. USA
anca@imbi.uni-freiburg.de

Background and Objectives   Methods used to cope with confounding in observational studies include traditional regression techniques and propensity score based approaches. An unadjusted analysis will typically lead to a biased estimate of the treatment or exposure effect since differences may be caused by treatment itself or by differences in measured or unmeasured confounders. The number of publications describing the application of propensity score based methods has grown rapidly in the last few years, reflecting the interest in alternative and potentially more powerful approaches to dealing with confounding [1,2]. To address this problem, the propensity score explores the relation between treatment assignment and potential confounders. This is a feature very different in spirit to the approach used in conventional regression. Here it is common to select confounders for inclusion in the model as a result of an observed relation with the outcome rather than a relation with treatment assignment.

The propensity score is the conditional probability of receiving a specific treatment or exposure, given a set of observed covariates. Rosenbaum and Rubin [3] have shown that within subgroups with identical propensity score, treatment groups are comparable in the sense that the probability distribution of covariates incorporated into the true score is identical. Based on this property, an unbiased estimator of a linear treatment effect has been proposed based on stratification for the propensity score [3]. Since the true propensity score is unknown, estimated scores, typically derived by logistic regression, are used in practice, and stratification is based on grouped values of the estimated scores (e.g. using quantiles). Alternatively, matching procedures can be used [4,5].

We performed both linear regression and a propensity score analysis in a population of elderly people with the objective to investigate whether the use of pain medication has an impact on the cognitive function and to find out which method is more appropriate and should be preferred. It was hypothesized that the intake of NSAIDS (non-steroidal anti-inflammatory drugs) would have no effect on the cognitive function in the elderly [6]. Here the term "cognitive function" is used to describe cognitive abilities like memory, thinking, verbal fluency and knowledge about one's personality. The study population was part of the ongoing Heidelberg long-term population-based cohort study (HeiDE).

Material and Methods   A number of standardized tests to explore the cognitive function were performed in HeiDE participants ≥ 70 years, including the East Boston Memory Test, a verbal fluency test and a Telephone Interview on Cognitive Status (TICS). TICS is summarized in a score taking values between 0 and 41 which is made up of several items testing orientation, attentiveness, short-term verbal memory, concentration, long-term memory, association and practical abilities. In the HeiDE study, out of a cohort of 729 eligible elderly 473 (64.9%) persons gave their written informed consent and participated in the telephone interview.

To study the association between NSAIDS and TICS scores, we used linear regression and propensity score analysis accounting for a series of covariates which were regarded as potential confounders. These were age, sex, educational level, body mass index, degree of physical activity, alcohol consumption, smoking status, and comorbidities (myocardial infarction, stroke, cancer, hypertension, diabetes, depressive symptoms). Use of NSAIDS was formalized as a binary treatment variable (NSAID = 1 if NSAIDs or aspirin were used, ignoring dose; NSAID = 0 if no NSAIDs and no aspirin were used).

The analysis strategy was defined prospectively as follows. First a linear regression was fitted in which TICS was modeled as a function of NSAIDs use and covariates, based on a combination of backward and forward selection steps. Covariates were selected from the full model including NSAID and all covariates without interaction terms, by backward selection with significance level 0.10. Interaction terms of the remaining variables were included into an intermediate model (forward step). The final model was determined by reapplication of backward selection. Secondly, an analysis approach based on the propensity score was performed. In our context the propensity score was the conditional probability of receiving NSAIDs, given the covariates. Logistic regression was used to generate estimated propensity scores for each person in the sample. The same selection procedure as above was applied correspondingly, modelling NSAID via logistic regression as a function of covariates (backward selection in the full model, forward selection of interaction terms, reapplication of backward selection). Next, the sample was stratified into the five quintile groups of the estimated propensity scores. The effect of NSAID on TICS was estimated by the weighted mean of unadjusted within-stratum differences between the two treatment groups, with weights proportional to the inverse variance of the stratum-specific estimates. Additionally, more refined methods for propensity score analysis were also planned (to be reported elsewhere). Standard errors and confidence intervals were calculated and compared for both methods.

Results   A mean TICS of 33.5 and a median TICS of 34.0 were observed in the 473 study participants. Information on treatment exposure was complete, with 54 (11.4%) persons reporting NSAID use. Due to missing values in covariates, the analysis sample was restricted to 404 participants with 49 (12.1%) exposed to NSAIDs.

In the linear regression model of TICS as a function of NSAIDs and covariates, application of the selection strategy left sex, educational level, physical activity, stroke, diabetes and depressive symptoms as covariate predictors of TICS in the final model. No interaction terms were included. NSAID had no influence on TICS, with an estimated mean improvement of 0.29 and 95% confidence interval [–0.52; 1.11]. In the propensity score analysis, covariates found to be relevant for the probability to receive NSAIDs in the final logistic model were educational level, body mass index and stroke. Again, no interaction terms were included. Due to the low number of exposed persons, quintile stratification was not possible since it would have produced empty treatment groups within strata. Thus the sample was divided into two strata based on the median estimated propensity score. The comparability of treatment groups within strata was inspected graphically and found to have improved for some covariates and strata and deteriorated for others. Within stratum estimates for the effect of NSAID were –0.40 and +0.36 (standard deviations 1.14 and 0.61). The combined estimate using weights proportional to the inverse variances is 0.19, with 95% confidence interval [–0.86; 1.25].

Discussion   Linear regression and analysis using the propensity score unequivocally showed no effect of NSAIDs on cognitive function as measured by TICS scores in a subpopulation of the HeiDE study. The estimated treatment effects were similar, with a slightly larger confidence interval found in the propensity score approach. This is in line with the findings of systematic reviews investigating publications in which results based on regression and propensity score analyses were compared [1,2]. Work in progress includes more elaborate model building [7] and applications of the propensity score methodology, like restriction to a subpopulation where the estimated propensity scores overlap between treatment groups, and matching based on estimated propensity scores [8]. While the linear regression could be calculated as planned in the cohort of complete cases, we could not perform the propensity score analysis as planned since there was not enough overlap in the upper quintile of estimated scores between treated and untreated persons. This could either be due to inclusion of non-confounders associated with exposure in the propensity score model based on the automated variable selection that would lead to unnecessary non-overlap of propensity score distributions in exposed and unexposed [7] or due to non-overlap of distributions of real confounders between exposed and unexposed leading to extrapolations in the outcome model.

Thus, the two approaches can be compared to some degree, based on the experiences made in a particular data set. However, in order to investigate other interesting properties like the degree of bias correction and efficiency for the two methods, the true treatment effect has to be known. This will be explored in a simulation study with a resampling design, based on the data described above. The vector of covariates will be divided into two subvectors. The subvectors will be randomly reallocated to the subjects in order to achieve independence between that set of covariates and NSAID or response. The internal association structure of the subvectors will remain unaffected. This design will allow to study extreme situations, for instance independence between subvectors and response, and thus allow to draw further conclusions on the properties of the two approaches in a realistic setting.

References

[1]   Sha BR, Laupacis A, Hux JE, Austin PC. Propensity score methods gave similar results to traditional regression modeling in observational studies: a systematic review. J Clin Epidemiol 2005; 58: 550-59.

[2]   Stürmer T, Joshi M, Glynn RJ, Avorn J, Rothman KJ, Schneeweiss S. A review of the application of propensity score methods yielded increasing use, advantages in specific settings, but not substantially different estimates compared with conventional multivariable methods. J Clin Epidemiol (in press online)

[3]   Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika 1983; 70: 41-55.

[4]   Rosenbaum PR, Rubin DB. Reducing bias in observational studies using subclassificationon the propensity score. J Am Stat Assoc 1984; 79: 516-524.

[5]   Weitzen S, Lapane KL, Toledano AY, Hume AL, Mor V. Principles for modeling propensity scores in medical research: a systematic literature review. Pharmacoepidemiol Drug Safe 2004; 13: 841-53.

[6]   Stürmer T, Glynn RJ, Field TS, Taylor JO, Hennekens CH. Aspirin use and cognitive function in the elderly. Am J Epidemiol 1996; 143: 683-691

[7]   Brookhart MA, Schneeweiss S, Rothman KJ,.Glynn RJ, Avorn J, Stürmer T. Variable selection in propensity score models: some insights from a simulation study. Am J Epidemiol (in press)

[8]   Stürmer T, Schneeweiss S, Brookhart MA, Rothman KJ, Avorn J, Glynn RJ. Analytic strategies to adjust confounding using exposure propensity scores and disease risk scores: nonsteroidal anti-inflammatory drugs and short-term mortality in the elderly. Am J Epidemiol 2005; 161: 891-98.