

GUEST EDITORIAL

Applications of the propensity score weighting method in psychogeriatric research: correcting selection bias and adjusting for confounders

The propensity score (PS) weighting method is an analytic technique that has been applied in multiple fields for a number of purposes. Here, we discuss two common applications, which are (1) to correct for selection bias and (2) to adjust for confounding variables when estimating the effect of an exposure variable on the outcome of interest.

In observational psychogeriatric research, investigators often need to address the issue of selection biases that determine who ends up participating in a given study (Ganguli *et al.*, 2015; Guo *et al.*, 2015). For example, when implementing an MRI study within an ongoing prospective community-based MRI study, we would ideally want to approach every age-appropriate person in the community. In actual fact, the participants will be limited to those who are available for MR imaging at the time (alive, in the area at the time, physically able to come to the radiology department), interested in undergoing MRI, and eligible (without contraindications) for MRI.

Study participants who died or dropped out of the parent study before the MRI study was implemented introduce what we called “attrition bias” (or survival bias) into the surviving sample. Individuals expressing interest in such an MRI study tend to be younger, of better physical, mental, and cognitive health, and with higher education; we have previously designated these attributes as representing “volunteer bias,” which further renders the available sample less random or “representative” of the target population. Furthermore, perhaps due to cost considerations, investigators may select a subsample of participants for MRI based on the hypothesis being tested, e.g. according to their cognitive status, rather than include all available participants or a random subsample. They must also exclude available and interested participants ineligible for MRI scanning because of contraindications such as an implanted cardiac pacemaker. As a result, the external validity (generalizability) of the study findings are potentially limited by attrition bias, volunteer bias, and selection bias, unless a way can be found to quantify and correct these biases.

In a research area, where the randomized trial design is impractical, investigators usually use an observational study design to investigate the effect of an exposure variable on an outcome of interest (Teno *et al.*, 2012; Xu and Kane, 2013). In the absence of randomization, confounding issues can become quite challenging. For example, in studying the effect of physical activity on cognitive decline among individuals 65 years or older, the researchers might find that participants with the exposure variable of being physically active tend to be younger, healthier, with less comorbid conditions, and were less likely to be depressed, as compared with those who were physically inactive. Since the same attributes are also associated with the outcome variable of better cognitive functioning, we have a potential confounding effect. The researchers need to first address confounding to estimate accurately the effect of the non-randomized exposure variable, i.e. level of physical activities; only then the results of the study can truly emulate data from a randomized trial design.

In the two examples mentioned above, the main issues in data analysis can be resolved by postulating a propensity score model (PSM), deriving weights from the PSM, and then incorporating weights into the analysis of the main outcome model. For the MRI study, we can use the weights obtained from three appropriate PSMs to create an equivalent sample that is comparable to the target population. For studying physical activity on cognitive decline, we can use the weight obtained from an appropriate PSM to create an equivalent sample that emulates a trial which randomly assigned individuals to the physically active group or the physically inactive group.

Logistic regression is the most frequently used model for building a PSM. When a PSM is fitted and estimated from the corresponding observational data, the PS of each individual can be obtained from the predicted probability of the fitted logistic regression model.

In the example of the MRI study, three logistic regression PSMs are needed to model selection bias, volunteer bias, and attrition bias.

The first PSM is for those who are eligible and available to participate in the study. The outcome variable is dichotomous, indicating whether or not an individual is selected for the final MRI study. Covariates must include all available variables related to the selection process and the outcome, including sociodemographic variables, history of physical and mental health, physical and cognitive activities, other lifestyle factors, and the neuropsychological tests results. From the first PSM, weights can be constructed for the actual participants of the study. The weight is the inverse of the PS, which is the predicted probability, from the regression model, of selection for the MRI study.

It is important to note that the quality of a PSM is affected by the existence of unmeasured (or residual) confounders, which are covariates impacting selection (or main exposure) but excluded from the observational study. Unmeasured confounders could result in biased estimation and the inference of the exposure effect. We should perform a sensitivity analysis to determine the impact of unmeasured confounding, if it exists, on estimation (Liu *et al.*, 2013). The commonly used methods include Rosenbaum and Rubin (1983) and VanderWeele and Arah (2011). To reduce the impact of unmeasured confounders in a PSM, we can include the observed covariates that are theoretically correlated to the unmeasured confounders, include interaction terms, and apply the doubly robust estimators (Robins *et al.*, 1994).

The second PSM for the MRI study is constructed among those who expressed interest in the MRI study. The outcome variable is dichotomous, indicating whether or not an individual is interested in participation. Covariates must include all available variables potentially related to the individual's interest in MRI study participation and the outcome, such as higher education or lack of depression. From the second PSM, weights can be constructed for those who were interested in participation. The weight is the inverse of the PS, which is the predicted probability of interest obtained from the regression model.

The third PSM is for all individuals or a random sample of the original study population. The outcome variable is dichotomous, indicating whether or not an individual could be approached for his/her interest in the study. The covariates must include all variables related to the availability of a candidate for interview, i.e. for not having died or dropped out already. From the third PSM, weights can be constructed among those who were approached. The weight is the inverse of the predicted probability of a candidate being available for interview.

The final weight will be constructed for the actual participants of the MRI study by multiplying the three individual weights described above. The final weight is used to correct biases from selection, volunteer, and attrition of the participants so that the results of the study can be generalized to the target population.

In the example of studying the effect of physical activities on cognitive decline, one logistic regression PSM is needed to adjust for the confounding issue. The outcome variable of the PSM is dichotomous, indicating whether an individual is physically active or not. The covariates need to include all available confounding variables, which may be not equally distributed between the physically active group and the inactive group. From this PSM, the weight for each individual can be constructed as the predicted probability of being physically active.

For both examples, the weights are constructed using the PS-based inverse probability weighting (IPW) method. In the main analysis model, instead of giving individuals equal weights, we give those individuals higher weights if they have lower chances of being selected into the MRI study, or if they have lower chances of being physically active in the physical activity study. With these weights, the results from the MRI main model can be generalized to the target population; and the effect of physical activity on cognitive decline in the physical activity study will be similar to that obtained from a hypothetically randomized trial.

Some points are worth mentioning in using the PS weighting method. First, the standard errors of the estimated exposure effect should be corrected. Bootstrap or sampling standard errors can be used in this correction. Second, the distribution of the weights needs to be examined before being implemented into the main outcome model. If a subject has a very large weight (greater than 10), then the estimated exposure effects could be inaccurate or unstable. For the estimated results to be stable, we can either exclude individuals having large weights, yet keep in mind that the exclusion could affect the interpretation and the generalizability of the results, or use the stabilized weights (Robins *et al.*, 2000).

For the PSM, careful thought should be given to choosing variables for the model, as opposed to including as many "adjustment" variables as possible. Researchers should include variables that are theoretically both associated with the selection (or main exposure) and the outcome or variables only associated with the outcome. Conversely, variables that are theoretically associated with the selection (or the main exposure variable) but not associated with the outcome are not true

confounders and therefore not recommended for the PSM (Rubin and Thomas, 1996; Brookhart *et al.*, 2006; Stuart *et al.*, 2013). To avoid multicollinearity, we should not include variables that are highly correlated with either. It is important to point out that the decision of variable selection should not be based on the p values because the goal here is not to build a parsimonious PSM. Moreover, the commonly used strategy for variable selection (e.g. stepwise, forward, backward) is not recommended here because model overfitting is not an issue.

The purpose of fitting a PSM is to accurately predict the selection (in the MRI study) or group assignment (physically active or inactive group in the physical activity study). Therefore, the performance of a PSM cannot be judged by using the usual modeling diagnostic tools such as the area under the receiver operating characteristic curve (AUC under ROC), c-statistic, or other model-fitting statistics. Those measures used to quantify the quality of a PSM also can measure the reduction in selection bias, or the reduction in confounding, which can be done through examining the distribution of characteristics between those who were selected and who were not, and between the individuals in different groups. To adjust for confounders, it is highly recommended that researchers evaluate the overlap of the empirical PS distributions (common support) and the covariate balance between the exposure and non-exposure groups. The recommended tests include graphical and numerical diagnostic methods; for examples, standardized differences, quantile–quantile plot, density plot, love plot, and ratio of variances (Rubin, 2001; Stuart, 2010).

To adjust for confounding, one needs to be aware of the difference between multivariable modeling and the PS method. Although multivariable modeling is the most commonly used method for confounding adjustment, the PS method has the following advantages. First, the PS method summarizes a set of confounders into a single measure; therefore, it is easier to achieve balance among covariates. Although distributions of PS are similar between groups when using the PS method, covariate balance between groups still needs to be examined because two individuals with the same PS for two individuals may not have the same characteristics. Second, for smaller sample sizes, model overfitting has a smaller impact on the PS method than multivariable modeling. Third, unlike regression-based confounding adjustment methods, the PS method does not assume linearity in covariates; it compares covariates between groups within the multidimensional covariate space and does not allow for extrapolation.

In summary, in research where an observational design is used to examine the causal effect of an exposure variable on the outcome of interest, the PS weighting method is becoming a standard way to correct selection bias or adjust for confounders. By applying this methodology in making estimations and inferences, researchers can enhance the validity of their analyses. This will help clinicians and policy-makers make informed decisions and develop appropriate prevention and intervention care for the elderly population.

Acknowledgments

The author would like to thank the Editor-in-Chief Dr. Nicola T. Lautenschlager for her idea of guest editorial for statistical methodology. The author is also grateful to Drs. Hiroko Dodge, Mary Ganguli, and Nicola T. Lautenschlager for their insightful comments and suggestions on the earlier drafts of this article.

CHUNG-CHOU H. CHANG

Department of Medicine, University of Pittsburgh,
Pittsburgh, PA, USA

Department of Biostatistics, Graduate School of Public
Health, University of Pittsburgh, PA, USA

Email: changj@pitt.edu

References

- Brookhart, M. A., Schneeweiss, S., Rothman, K. J., Glynn, R. J., Avorn, J. and Stürmer, T. (2006). Variable selection for propensity score models. *American Journal of Epidemiology*, 163, 1149–1156.
- Ganguli, M. *et al.* (2015). Who wants a free brain scan? assessing and correcting for recruitment biases in a population-based sMRI pilot study. *Brain Imaging and Behavior*, 9, 204–212.
- Guo, K. *et al.* (2015). Propensity score weighting for addressing under-reporting in mortality surveillance: a proof-of-concept study using the nationally representative mortality data in China. *Population Health Metrics*, 13, 16. DOI: [10.1186/s12963-015-0051-3](https://doi.org/10.1186/s12963-015-0051-3).
- Liu, W., Kuramoto, S. J. and Stuart, E. A. (2013). An introduction to sensitivity analysis for unobserved confounding in nonexperimental prevention research. *Prevention Science*, 14, 570–580.
- Robins, J. M., Hernán, M. and Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology*, 11, 550–560.
- Robins, J. M., Rotnitzky, A. and Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, 89, 846–866.
- Rosenbaum, P. R. and Rubin, D. B. (1983). Assessing sensitivity to an unobserved binary covariate in an

observational study with binary outcome. *Journal of the Royal Statistical Society Series B*, 45, 212–218.

- Rubin, D. B.** (2001). Using propensity scores to help design observational studies: application to the tobacco litigation. *Health Services & Outcomes Research Methodology*, 2, 169–188.
- Rubin, D. B. and Thomas, N.** (1996). Matching using estimated propensity score: relating theory to practice. *Biometrics*, 52, 249–264.
- Stuart, E. A.** (2010). Matching methods for causal inference: a review and a look forward. *Statistical Science*, 25, 1–21.
- Stuart, E. A., DuGoff, E., Abrams, M., Salkever, D. and Steinwachs, D.** (2013). Estimating causal effects in observational studies using electronic health data: challenges and (some) solutions. *eGEMs (Generating Evidence & Methods to Improve Patient Outcomes)*, 1, Article 4. DOI: <http://dx.doi.org/10.13063/2327-9214.1038>.
- Teno, J. M. et al.** (2012). Does feeling tube insertion and its timing improve survival?. *Journal of the American Geriatrics Society*, 60, 1918–1921.
- VanderWeele, T. J. and Arah, O. A.** (2011). Bias formulas for sensitivity analysis of unmeasured confounding for general outcomes, treatments, and confounders. *Epidemiology*, 22, 42–52.
- Xu, D. and Kane, R. L.** (2013). Effect of urinary incontinence on older nursing home residents' self-reported quality of life. *Journal of the American Geriatrics Society*, 61, 1473–1481.