Assessing confounder balance in outcome regressions.

(Running head: Outcome regression & balance checking)

Original Research Article - Methodology

Authors

Frank Popham (corresponding author)

MRC/CSO Social and Public Health Sciences Unit, University of Glasgow, 200 Renfield Street, Glasgow G2 3QB, Scotland. frank.popham@glasgow.ac.uk, Tel: +44 0141 353 7535.

Alastair H Leyland

MRC/CSO Social and Public Health Sciences Unit, University of Glasgow, 200 Renfield Street, Glasgow G2 3QB, Scotland.

Contribution

FP had the original idea inspired by reading, Abadie, A., Diamond, A., & Hainmueller, J. (2015). Comparative Politics and the Synthetic Control Method. *American Journal of Political Science*, 59, 495-510. FP and AL developed the idea. AL derived the matrix algebra / proofs. FP wrote the first draft which was extensively edited by AL. FP wrote the Stata (StataCorp, College Station) and R code (R foundation, Vienna). FP and AL have approved the manuscript.

Conflicts of interest

None

Funding

FP and AL are funded by the Medical Research Council (MC_UU_12017/13) and Scottish Government Chief Scientist Office (SPHSU13)

Reproducibility

The example data set comes from Hernan and Robins' causal inference book ¹ and is obtainable from the book's website (<https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/>). Stata (StataCorp, College Station) and R (R foundation, Vienna) code are provided (as supplemental materials) to replicate the results described.

Acknowledgements

FP would like to thank colleagues who patiently listened to early iterations of this idea and gave feedback. Although not referenced in the paper, FP learnt a great deal about matrix transformation from the following blog post. Gould, W. (2011). Understanding matrices intuitively, part 1. *The Stata Blog: Not elsewhere classified*. <http://blog.stata.com/2011/03/03/understanding-matrices-intuitively-part-1/>

Copyright - <https://creativecommons.org/licenses/by/4.0/>

Abstract

An outcome regression controlling for observed confounders remains a popular way to assess the causal effect of an exposure in epidemiology, despite more modern causal techniques for adjusting for observed confounders, such as inverse probability weighting. A feature of inverse probability weighting is that checking balance of confounders in the control and exposure groups after confounder adjustment is simple. However, researchers using outcome regressions commonly do not check confounder balance after controlling for confounders. Although outcome regressions will balance any confounder specified in the model, the confounder value the model balances at is not transparent. We show that a matrix representation of an outcome regression reveals that an outcome regression includes a weight similar to an inverse probability weight. We also show that outcome regressions may not be balancing at the sample mean of the confounders particularly if interactions are not included with the exposure, which is typically the case in outcome regressions. Finally, we show that the coefficient of the exposure in an outcome regression is simply the difference between two weighted counterfactuals. Thus, there is an important connection between traditional outcome regression and modern causal techniques.

Introduction

Estimating causal effects is a key aim for epidemiology. A major task in doing so is achieving exchangeability of exposed and control groups, so that the treatment effect would be the same if the groups' exposure were switched. This relies on study design, and / or statistical techniques to adjust for observed and unobserved confounders¹⁻³. In observational epidemiology statistical adjustment for observed confounders remains a popular approach, one that often uses so called outcome regression¹. This is when the outcome is modelled in a single regression model as the effect of the exposure with adjustment for observed confounders. However, the potential outcomes framework for causality, from which exchangeability comes, has led to alternative methods for achieving exchangeability through confounder control¹. For example, inverse probability weighting (IPW) derives weights from a treatment model (exposure as a function of confounders) and these weights are used to adjust for confounding by weighting a model that usually includes the outcome and exposure but no confounders¹. An attractive feature of IPW is that it is easy to check the balance of confounders between exposure and control groups and thus exchangeability on observed variables after weighting⁴. An outcome regression will balance the confounders adjusted for, but it is not standard to check where this balance is achieved. Perhaps not widely appreciated in epidemiology is that an outcome regression also includes a weight^{5,6} which we will show is similar to an inverse probability weight. This is a potentially important insight; as it makes a connection between the traditional outcome regression approach and more recent causal methods; and also suggests a way in which confounder balance in outcome regressions can be checked.

A simple example

Inverse probability weight

Let us illustrate with a simple example, a continuous outcome (Y_i) and a binary exposure (X_i) coded 0 for control and 1 for exposure. While it may seem odd to start with an example without a confounder, it helps illustrate clearly the connection between IPW and outcome regression. There are 1242 people in the control group and 437 in the exposed, a total sample size of 1679. IPW works by creating pseudo- populations from the control and exposed groups where the probability of assignment is independent of the adjusted for confounders.¹ Here we use non-stabilized weights and make the exposure and control groups each have a weighted sample size of 1679. This is done by creating a weight of $1679 / 1242 = 1.35$ for the control and a weight of $1679 / 437 = 3.84$ for the exposed. We can then calculate the average weighted outcome for the exposure group and for the control group. The average effect (2.36) is the difference between the average weighted outcome for the controls (2.13) compared to the exposed (4.50). Of course with no confounder this is just the same as taking the difference between unweighted mean outcomes. The difference between these means can be obtained from the estimate for X_i from an inverse probability weighted regression of the outcome by exposure.

Outcome regression

Turning to outcome regression, we start by not including a constant but instead having two uncorrelated binary variables - X_{0i} and X_{1i} - representing, respectively, membership of the control group (coded 1 if in the control group and 0 if not) and the exposed group (coded 1 if in the exposed group and 0 if not). The regression equation

$$Y_i = \beta_0 X_{0i} + \beta_1 X_{1i} + e_i \quad (1)$$

where e_i is the (Gaussian) error term, can be written in matrix notation as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} \quad (2)$$

The weight in an outcome regression can be seen in the ordinary least squares solution for $\boldsymbol{\beta}$, the vector of parameter estimates, given by

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} \quad (3)$$

The matrix $\mathbf{X}$ contains all of the variables included in the regression (in this example just X_{0i} and X_{1i} , but in general this would include the constant, the exposure and confounders). Here it is a matrix with 1679 rows and 2 columns. $\mathbf{Y}$ is the matrix of the outcome (Y_i) with 1679 rows and 1 column.

We can break equation (3) down into its component parts. In an example containing two variables X_{0i} and X_{1i} in the regression equation the matrix of cross-products $\mathbf{X}^T \mathbf{X}$ is given algebraically by

$$\mathbf{X}^T \mathbf{X} = \begin{bmatrix} \sum x_{0i}^2 & \sum x_{0i} x_{1i} \\ \sum x_{0i} x_{1i} & \sum x_{1i}^2 \end{bmatrix}$$

In this particular case either x_{0i} or x_{1i} is equal to 1 but not both, so $\sum x_{0i} x_{1i} = 0$. The $\sum x_{0i}^2$ and $\sum x_{1i}^2$ here are just the numbers in the control and exposed group respectively so

$$\mathbf{X}^T \mathbf{X} = \begin{bmatrix} 1242 & 0 \\ 0 & 437 \end{bmatrix}$$

The inverse of this matrix is given by

$$(\mathbf{X}^T \mathbf{X})^{-1} = \begin{bmatrix} 1/1242 & 0 \\ 0 & 1/437 \end{bmatrix}$$

The diagonal elements in this matrix are equivalent in value to the inverse probability weights divided by the sample size (1679). So unlike IPW, where the weighted sample size for each group (control and exposed) is the original total sample size, these matrix weights sum to 1 for each group, a simple matter of a difference of scale.

$(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$ is similar to an IPW in that it is applied to the observed outcome $\mathbf{Y}$ to provide the parameter estimates $\hat{\boldsymbol{\beta}}$. In this particular parameterisation, these weights give the mean of the y_i for the control and exposed groups.

If we now revert to a more common regression parameterisation by specifying x_{0i} as a constant (=1) such that

$$y_i = \beta_0 + \beta_1 x_{1i} + e_i \quad (4)$$

The coefficient β_1 represents the mean difference between the exposed and control groups. Now we have

$$\mathbf{X}^T \mathbf{X} = \begin{bmatrix} 1679 & 437 \\ 437 & 437 \end{bmatrix}$$

with these elements being the total sample size and the number of exposed cases. In general terms the inverse matrix $(\mathbf{X}^T \mathbf{X})^{-1}$ is given by

$$(\mathbf{X}^T \mathbf{X})^{-1} = \frac{1}{\Delta} \begin{bmatrix} \sum x_{1i}^2 & -\sum x_{0i} x_{1i} \\ -\sum x_{0i} x_{1i} & \sum x_{0i}^2 \end{bmatrix}$$

where

$$\Delta = \sum x_{0i}^2 \sum x_{1i}^2 - (\sum x_{0i} x_{1i})^2$$

so in this example

$$(\mathbf{X}^T \mathbf{X})^{-1} = \frac{1}{542754} \begin{bmatrix} 437 & -437 \\ -437 & 1679 \end{bmatrix}$$

The matrix weight is then

$$\begin{aligned} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T &= \begin{bmatrix} \{w_{\beta_{0i}}\} & \{w_{\beta_{1i}}\} \end{bmatrix} \\ &= \frac{1}{\Delta} \begin{bmatrix} \{x_{0i} \sum x_{1i}^2 - x_{1i} \sum x_{0i} x_{1i}\} & \{x_{1i} \sum x_{0i}^2 - x_{0i} \sum x_{0i} x_{1i}\} \end{bmatrix}^T \end{aligned}$$

which in our example gives weights for β_0 of 437/542754 and -437/542754 for control ($x_{1i} = 0$) and exposed ($x_{1i} = 1$) respectively, and weights for β_1 of -437/542754 and 1242/542754.

Concentrating on the weights $w_{\beta_{1i}}$, the estimate $\hat{\beta}_1$ is given by

$$\hat{\beta}_1 = \sum w_{\beta_{1i}} y_i$$

which in this example comes to 2.36. This can be broken down into the two components for controls $\sum w_{\beta_{1i}} y_i |_{x_{1i}=0} = -2.13$ and exposed $\sum w_{\beta_{1i}} y_i |_{x_{1i}=1} = 4.50$. These are the negative of the mean of the y_i for the controls and the mean of the y_i for the exposed. The coefficient is the difference between the two given the negative sign for the control group.

Adding a confounder

Inverse probability weighting

We now add a binary confounder (Z_i). The observed relationship between Z_i and X_i is given in the table 1 below.

Table 1. Relationship between exposure and confounder

		Z_i		
		0	1	Total
X_i	0	586	656	1,242
	1	239	198	437
	Total	825	854	1,679

We now calculate the IPW for each level of Z_i . For $Z_i = 0$ and $X_i = 0$, the IPW is $825/586 = 1.41$ and for $Z_i = 0$, $X_i = 1$ the IPW is $825/239 = 3.45$. The IPW for $Z_i = 1$, $X_i = 0$ is $854/656 = 1.30$ and for $Z_i = 1$, $X_i = 1$ is $854/198 = 4.31$. The weights for $X_i = 0$ and $X_i = 1$ sum over values of Z_i to give a sample size of 1679 (for example, when $X_i = 0$ we have $586 \cdot 825/586 + 656 \cdot 854/656$). We can also check the balance of the confounder across the values of X_i in the weighted sample to confirm they are now balanced. Table 2 shows how the IPW balances Z_i over X_i at the population mean of Z_i (i.e. $Z_i = 1 = 854/1679 = 51\%$)

Table 2. IPW balances the confounder over exposure at the population mean of the confounder

		Z_i		
		0	1	Total
X_i	0	$586 \cdot 1.41 = 825$	$656 \cdot 1.30 = 854$	1,679
	1	$239 \cdot 3.45 = 825$	$198 \cdot 4.31 = 854$	1,679

The probability of exposure in the pseudo-populations is now independent of Z_i . Applying the weights to values of y_i produces a mean y_i for $X_i = 1$ of 4.45 and $X_i = 0$ of 2.13, giving an effect (the difference between the two) of 2.32.

Outcome regression

Let us start by acknowledging that there is no expectation that the matrix derived weights from an outcome regression will produce the same results as the IPW when we have a simple binary confounder. The reason is that the IPW above effectively allows for the interaction between X_i and Z_i . Standardly when reporting the average effect of an exposure from an outcome regression (the coefficient associated with X_i) researchers do not include an interaction between X_i and Z_i as it is not possible to read the average effect of X_i directly from the regression output. If an interaction is included an average effect can be obtained, for example, by applying post regression weights to calculate the average effect of X_i (as in standardisation)¹.

In the regression equation

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 Z_i + e_i \quad (5)$$

the matrix of cross-products $\mathbf{X}^T \mathbf{X}$ is now given algebraically by

$$\mathbf{X}^T \mathbf{X} = \begin{bmatrix} n & \sum x_i & \sum z_i \\ \sum x_i & \sum x_i^2 & \sum x_i z_i \\ \sum z_i & \sum x_i z_i & \sum z_i^2 \end{bmatrix}$$

and

$$(\mathbf{X}^T \mathbf{X})^{-1} = \frac{1}{\Delta} \begin{bmatrix} \sum x_i^2 \sum z_i^2 - (\sum x_i z_i)^2 & \sum z_i \sum x_i z_i - \sum x_i \sum z_i^2 & \sum x_i \sum x_i z_i - \sum z_i \sum x_i^2 \\ \sum z_i \sum x_i z_i - \sum x_i \sum z_i^2 & n \sum z_i^2 - (\sum z_i)^2 & \sum x_i \sum z_i - n \sum x_i z_i \\ \sum x_i \sum x_i z_i - \sum z_i \sum x_i^2 & \sum x_i \sum z_i - n \sum x_i z_i & n \sum x_i^2 - (\sum x_i)^2 \end{bmatrix}$$

$$\Delta = n \sum x_i^2 \sum z_i^2 + 2 \sum x_i \sum z_i \sum x_i z_i - n (\sum x_i z_i)^2 - \sum x_i^2 (\sum z_i)^2 - \sum z_i^2 (\sum x_i)^2$$

The matrix weight associated with β_1 , the coefficient of the X_i , is then given by the elements of the second row of $(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$

$$\frac{1}{\Delta} \left(\sum z_i \sum x_i z_i - \sum x_i \sum z_i^2 + n x_i \sum z_i^2 - x_i (\sum z_i)^2 + z_i \sum x_i \sum z_i - n z_i \sum x_i z_i \right)$$

In our example for combinations of X_i and Z_i this comes to the weights in Table 3 (for comparison with IPWs in Table 2 the outcome weights have been multiplied by the sample size and their absolute value shown)

Table 3 Weights for the exposure after adjustment for confounding in an outcome regression multiplied by total sample size with the absolute value shown.

		Z_i	
		0	1
X_i	0	1.51	1.21
	1	3.71	4.01

These weights are not the same as the IPWs. They also balance (see Table 4) but not at the mean of Z_i in the population as the IPW did.

Table 4 Outcome regression weights balance the confounder over exposure but not at the population mean of the confounder.

		Z_i		
		0	1	Total
X_i	0	586*1.51 = 886	656*1.21 = 793	1,679
	1	239*3.71 = 886	198*4.01 = 793	1,679

$Z_i = 1$ is balanced at 47% $((793.41 / 1679) * 100)$ rather than the population mean of 51% which is where the IPW balanced it. Now applying these weights to Y_i by X_i , we get a mean of Y_i for $X_i = 1 = 4.48$ and for $X_i = 0 = -2.13$, when summed in the matrix $(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$, a coefficient for X_i of 2.35 is returned. This is close but different to the result of the IPW but it is, of course, the result we obtain for the coefficient of X_i from running an outcome regression in statistical software (i.e. without going through matrix steps).

Now let us include an interaction between X_i and Z_i in the matrix. We can rewrite the regression equation

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 Z_i + \beta_3 X_i Z_i + e_i \quad (6)$$

as

$$y_i = \beta_{00}(1 - x_i)(1 - z_i) + \beta_{01}(1 - x_i)z_i + \beta_{10}x_i(1 - z_i) + \beta_{11}x_i z_i + e_i \quad (7)$$

The two are equivalent since there is a direct mapping from one set of parameters to the other:

$$\begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & 1 & 0 & 0 \\ 1 & -1 & -1 & 1 \end{bmatrix} \begin{bmatrix} \beta_{00} \\ \beta_{01} \\ \beta_{10} \\ \beta_{11} \end{bmatrix}$$

We can therefore write

$$\mathbf{X} = \left[\{(1 - x_i)(1 - z_i)\} \quad \{(1 - x_i)z_i\} \quad \{x_i(1 - z_i)\} \quad \{x_i z_i\} \right]$$

These variables are therefore uncorrelated and the matrix $\mathbf{X}^T \mathbf{X}$ is given by

$$\mathbf{X}^T \mathbf{X} = \begin{bmatrix} n_{00} & 0 & 0 & 0 \\ 0 & n_{01} & 0 & 0 \\ 0 & 0 & n_{10} & 0 \\ 0 & 0 & 0 & n_{11} \end{bmatrix}$$

where n_{xz} is the number of observations for which $x_i = X$ and $z_i = Z$. In our example,

$$\mathbf{X}^T \mathbf{X} = \begin{bmatrix} 586 & 0 & 0 & 0 \\ 0 & 656 & 0 & 0 \\ 0 & 0 & 239 & 0 \\ 0 & 0 & 0 & 198 \end{bmatrix}$$

The rows of $(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$ then contain the weights associated with the four regression parameters

β_{00} , β_{01} , β_{10} and β_{11} :

$$(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T = \left[\{(1 - x_i)(1 - z_i)/n_{00}\} \quad \{(1 - x_i)z_i/n_{01}\} \quad \{x_i(1 - z_i)/n_{10}\} \quad \{x_i z_i/n_{11}\} \right]^T$$

Because we have no constant, the four coefficients returned by $(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$ are the means of

y_i by the four combinations of x_i and z_i . The four denominators for the outcome regression weights contained in the matrix $\mathbf{X}^T \mathbf{X}$ above are the same as the denominators for the IPW but the numerator for when we invert the matrix is 1 in all cases for the outcome regression but the numerator for the IPW is not a constant but reflects the different size of $z_i = 0$ (825) and $z_i = 1$ (854). In this example we have four matrix weights from the outcome regression that cover the exposure but as each is only non-zero for its unique combination of x_i and z_i , we can combine into a single weight. This weight balances z_i at 0.5 for both $z_i = 0$ and $z_i = 1$. This is because it gives

equal weight to each strata of Z_i rather than reflecting their different size as the IPW does. The outcome regression is conditional, there are two “experiments”, one in each strata of Z_i reflected in the fact weights sum to 4 overall, 2 in each strata of Z_i . As stated already, in reality researchers do not usually include an interaction between the exposure and confounders when estimating a main or average effect using an outcome regression.

Multiple confounders

Let us add a further binary confounder, Q_i , so there are now two confounders (Z_i and Q_i). Table 5 shows their individual and joint distribution in the dataset (column labelled no weight). The column IPW is their distribution when weighted by an IPW derived from a treatment model of X_i that included the main effects of Z_i and Q_i while the next column shows the results when also including the interaction of Z_i and Q_i as well as their main effects. The fourth and fifth columns are their distribution using the weight from an outcome regression of X_i , Z_i and Q_i on Y_i with no interaction between Z_i and Q_i and the interaction between Z_i and Q_i . Both IPW models reweight the distribution of Z_i , Q_i and their interaction close to the population average (total) for both values of X_i . Without the interaction between Z_i and Q_i , Z_i and Q_i and their interaction are not quite exactly matched over X_i but with the interaction they are. The outcome regression without the interaction reweights Z_i and Q_i to be the same (but not to the population average) but does not balance the interaction in their joint distribution across X_i . Including the interaction between Z_i and Q_i achieves balance on Z_i , Q_i and their interaction but again not at the average level for the population.

Table 5 Confounder balance for two binary confounders and their interaction using IPW and outcome regression.

					IPW (interaction between z and q)				Outcome regression (interaction between z and q)			
	No weight		IPW									
	Total	X=0	X=1	X=0	X=1	X=0	X=1	X=0	X=1	X=0	X=1	X=1
Z=0	49.1	47.2	54.7	49.1	49.1	49.1	49.1	52.8	52.8	52.7	52.7	
Z=1	50.9	52.8	45.3	50.9	50.9	50.9	50.9	47.2	47.2	47.3	47.3	
Q=0	86.7	85.3	90.4	86.7	86.6	86.7	86.7	89.4	89.4	89.4	89.4	
Q=1	13.3	14.7	9.6	13.3	13.4	13.3	13.3	10.6	10.6	10.6	10.6	
Z=0, Q=0	43.2	40.8	49.9	43.2	43.2	43.2	43.2	47.4	47.6	47.6	47.6	
Z=0, Q=1	6	6.4	4.8	6	5.9	6	6	5.4	5.1	5.2	5.2	
Z=1, Q=0	43.5	44.5	40.5	43.5	43.4	43.5	43.5	42.0	41.8	41.8	41.8	
Z=1, Q=1	7.4	8.3	4.8	7.4	7.5	7.4	7.4	5.2	5.5	5.4	5.4	

Conclusion

As far as we are aware it is not usual in epidemiology to check the balance of confounders over exposure levels after using an outcome regression. Of course this is commonly done to show the imbalance beforehand. In contrast it is standard practice to check balance when causal methods such as IPW have been applied. This paper has shown that checking post regression balance is possible using weights derived from a matrix representation of an outcome regression. Even though an outcome regression will balance the confounders adjusted for, the counterfactuals constructed may not be at the population mean as one might expect. Moreover, any interactions between confounding variables may not be balanced if not directly modelled nor might the balance be at the population mean if interactions with the exposure are not accounted for. This is the practical aspect of this paper. It draws on the insight of Abadie et al. that outcome regressions can be seen as weighting estimators when expressed in matrix form⁵. We extend this by drawing comparisons with inverse probability weighting and illustrating how the balance in standard regression methods can be checked. Of course the properties of outcome regressions and their matrix representation are well known, we are not showing anything new statistically. For example, the projection matrix or hat matrix $(\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T)$ has long been used to assess how influential particular data points are in mapping observed values of the outcome to predicted values (leverage)⁷. Our paper also highlights the connection between outcome regression and a more “modern” causal estimator (IPW) in that they involve similar weights and that effect estimates arise from the comparison of counterfactuals. This is an important insight given recent debates that have included criticism of the potential outcomes approach’s influence on causal methods in epidemiology⁸. Standard outcome regression approaches fall within a potential outcomes framework as well. Moreover, an outcome regression could be run in two stages with $(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ run first and then the resulting weight applied to the outcome. This two stage outcome regression would share the two stages of the IPW, one additional advantage of which is that in the first stage the analyst can be blinded to the outcome reducing the chance that confounder selection is influenced by seeing its effect on the outcome⁹. Further work is needed to extend our approach to continuous exposures and non-linear models but we believe that it can be generalised, most naturally perhaps in a generalised linear model framework. An additional area for future work is to explore what the outcome regression weights tell us about positivity (extrapolation). In the binary case as the control group’s weights are usually negative and the exposure group’s weights are positive, if either includes a weight with the opposite sign this would likely suggest extrapolation we think.

In summary, matrix derived weights allow researchers to check where their outcome regressions are balancing confounders. They may be surprised that it is often not at the mean of their population.

References

1. Hernán M, Robins J. *Causal Inference Book*. Boca Raton, FL: Chapman Hall/CRC. (forthcoming), <https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/>
2. Imbens GW, Rubin DB. *Causal inference in statistics, social, and biomedical sciences* Cambridge University Press, 2015.
3. Angrist JD, Pischke J-S. *Mostly harmless econometrics: An empiricist's companion* Princeton University Press, 2008.
4. Austin PC, Stuart EA. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Statistics in Medicine* 2015;**34**(28):3661-3679.
5. Abadie, A., Diamond, A., & Hainmueller, J. (2015). Comparative Politics and the Synthetic Control Method. *American Journal of Political Science*, 59, 495-510
6. Brown JD. *Linear Models in Matrix Form: A Hands-On Approach for the Behavioral Sciences*. London: Springer, 2014.
7. Hoaglin DC, Welsch RE. The hat matrix in regression and ANOVA. *The American Statistician* 1978;**32**(1):17-22.
8. Krieger N, Davey Smith G. The tale wagged by the DAG: broadening the scope of causal inference and explanation for epidemiology. *International Journal of Epidemiology* 2016;**45**(6):1787-1808.
9. Rubin DB. For objective causal inference, design trumps analysis. *Annals of Applied Statistics* 2008;**2**(3):808-840.