Paper AS09

Estimation of the Hypothetical Estimands using Longitudinal version of the TMLE

Shreya Wahi, Cognizant, Mumbai, India

ABSTRACT

Recent years have seen growing use of doubly robust (DR) estimators like the Target Maximum Likelihood Estimators (TMLE), in estimating causal estimands. Causal estimands are statistical quantities of interest, which are defined using the language of potential outcomes, for example, the Average Treatment Effect (ATE). The popularity of such DR estimators is partly due to the fact that the method is consistent (that is, asymptotically unbiased) when the treatment and the outcome generating mechanism are incorrectly specified. Longitudinal version of the TMLE, where we have repeated treatment, outcome and covariate (possibly time-varying), has also been developed. In this talk we illustrate how LTMLE can be used to estimate the causal estimands associated with the hypothetical strategy. Hypothetical strategy is used to handle intercurrent events when we are interested in the treatment effect had the ICE not occurred. In this case, the treatment effect can be easily written as a contrast of causal estimand defined at multiple time points which can be estimated using the methods of LTMLE.

INTRODUCTION

While conducting randomized clinical trials, there might be events that occur post-randomization that can alter the outcome of interest or the corresponding interpretation, leading to biased treatment effects estimates. These events are called Intercurrent events (ICEs) that include, intake of rescue medication, treatment switch, occurrence of some adverse event, death, etc. It is important to not treat these as missing data, rather control their effect using an appropriate statistical methodology. According to ICH E9(R1), it is necessary to define an appropriate estimand to handle these ICE, taking into account five factors: treatment conditions, targeted patients, endpoint of interest(variable), population-level summary of the variable and the intercurrent events.
It is important to specify how the intercurrent events are supposed to be handled while describing an estimand for the clinical question of interest. One of the strategies defined in the addendum is Hypothetical Strategy, in which we estimate the treatment effect in a hypothetical scenario in which ICE does not occur at all.

CAUSAL INFERENCE

Causal Inference helps us to identify and quantify the causal effect of one variable over the other, while controlling the confounding factors. The unmeasured variables that have an influence over the relationship of the variables under study (treatment and outcome), that cause potential bias in the estimation of treatment effect are called ***confounding factors***. Causal methods help to extract the pure treatment effect over the outcome of interest by balancing/controlling the external effects of the confounding factors. Causal methods can be used in observational studies as well as RCTs (randomized clinical trials) where there are occurrences of ICEs.

ESTIMATORS

Our parameter of interest is the mean difference of the outcomes of when the group is given a treatment and when it is not (average treatment effect – ATE). In reality, we can only observe one of the two outcomes for one subject, hence the counterfactual outcome has to be estimated for every subject.

For estimation of ATE, we have single robust methods, under which we have two approaches, namely g-computation and inverse probability of treatment estimator. The former works on modelling the outcome over the treatment and covariates, predicting the counterfactual outcomes for all the subjects and computing the mean difference of the outcomes over the treatment group. The latter works on modelling the probability of treatment, computing the weighted mean of the outcomes using the inverse of treatment probabilities as weights and the difference of the weighted mean of the outcomes of the treatment groups is our estimate of parameter of interest. However, the limitation with the single robust estimators is that the estimates are biased and unreliable if the model is not correctly specified and extreme treatment probabilities lead to high variance, making the estimate highly sensitive.

Double robust methods use both the outcome and treatment probability modelling, which makes the estimate unbiased and consistent if either of the models is correctly specified. The estimate is also efficient if both the models are correctly specified.

TARGETED MAXIMUM LIKELIHOOD ESTIMATOR

Targeted maximum likelihood estimator (TMLE) is one of the doubly robust methods of estimation for treatment effect in causal inference, that incorporates both outcome modelling and propensity score weighting to get unbiased, consistent and efficient estimates of treatment effects.

It is a semiparametric method that incorporates the use of machine learning for model estimation, that minimizes the requirement of data assumptions, leading to lesser chances of model misspecification.

TMLE IN LONGITUDNAL DATA

The methods of TMLE can be applied in longitudinal studies, where the causal effect of interest is the effect of a treatment regime given over time (dynamic/static), with time varying confounders, long follow-up time and complex associational relationships. The LTMLE estimator is introduced by Van der Laan and Gruber, where it has been successfully used to compute the treatment effect in complex settings.

LTMLE is a double robust method of estimation for treatment effect, that requires estimation of outcome, censoring and treatment probability at every time point. The estimator is consistent if either of the iterated treatment and censoring model or the outcome model is correctly specified. It also incorporates the use of machine learning algorithms for modelling that reduce the chances of misspecification and also requires lesser data distribution assumptions. Furthermore, the consistent estimation of treatment, censoring and outcome models improves the efficiency of the estimator, reduces finite sample bias and provide basis for valid statistical inference.

ASSUMPTIONS

LTMLE requires the same assumptions required in the implementation of TMLE.

1. *Consistency:* For each individual, the observed outcome under the treatment history they actually followed equals their counterfactual outcome under that same treatment history.
2. *Positivity:* There is a non-zero positive probability of receiving the treatment regime of interest, given all the levels of covariates.
3. *Conditional Exchangeability:* At each time point, given the observed past covariates and treatment history, the treatment assignment is independent of the potential outcomes.

METHODOLY

For LTMLE, it is required to estimate the TMLE at every time point, back-iterating till time point t=1 (t = T,…2,1). Let's say our causal quantity of interest is $\psi_T = E[\ Y_T^{\overline{dt}}\ ]$, which is the expected value of outcome at time point T under the intervention rule $\overline{dt}$, which assigns treatment as a function of the time-varying covariates $\bar{L}t$ and sets censoring determin istically to 1. The target quantity is first estimated by the g-sequential method, which is then updated during a *targeting step* with the help of propensity scores computed by using the treatment and censoring mechanisms that enables the double robust property of LTMLE.

Starting from the final time point t = T, the steps are as follows:

- **Step 1:** Defining the new data with the treatment intervention.
  We first define a new data set, by plugging in the treatment intervention of interest, keeping the outcomes and the time-varying covariates/confounders the same as the original dataset.

- **Step 2:** Initial outcome estimates.
  Secondly, the initial estimates of the outcome are found from the original data using a suitable regression model.
  $$E[Y_t|\bar{A}_{t-1}, \bar{L}_{t-1}]$$
  The outcomes of time point t are regressed over all the nodes of treatment and time varying covariates that occur before that time point. Then, using this model, predict the outcomes over the new data and store these values as $\hat{Y}_t$.

- **Step 3:** Propensity Scores and Clever Covariates
  Next, the models are defined to compute the propensity scores (i.e. the probability of treatment assignment) for at every time point.
  $$P\left[A_t = \bar{a}_t \middle| \bar{L}_t = \bar{l}_t, \bar{A}_{t-1} = \bar{a}_{t-1}\right] \ , for\ t = 1,2, \ldots . T$$
  Using this model the probabilities are predicted over the original dataset that gives the propensity scores for every time point. Similarly, the probability of censoring is also computed using the suitable model.
  $$P\left[C_t = 1 \middle| \bar{L}_t = \bar{l}_t, \bar{A}_t = \bar{a}_t, \bar{C}_{t-1} = 1\right] \ , for\ t = 1,2, \ldots . T$$
  Using the above calculated probabilities, the clever covariate is computed as:

$$\bar{H}(\bar{A}, \bar{C}, \bar{L})_t = \prod_{s=1}^{t} \frac{I(\bar{A}_s = \bar{a}_s)\ x\ I(\bar{C}_s = 1)}{P\left[A_s = \bar{a}_s \middle| \bar{L}_s = \bar{l}_s, \bar{A}_{s-1} = \bar{a}_{s-1}, \bar{C}_{s-1} = 1\right] x\ P\left[C_s = 1 \middle| \bar{L}_s = \bar{l}_s, \bar{A}_s = \bar{a}_s, \bar{C}_{s-1} = 1\right]}$$, for t = 1,2 ..T.

- **Step 4:** Update the initial estimates
  Now, the observed outcome Y_t of time t is regressed over the initial estimates and the re-weighted clever covariates, such that this model gives the estimates of the fluctuation parameter ϵ. The model is defined as below:

$$E\left[\hat{Y}_t \middle| \bar{A}_t = \bar{a}_t, \bar{L}_t = \bar{l}_t, \bar{C}_t = 1\right] = \frac{1}{1 + \exp\left(-\log\left(\frac{\hat{Y}_t}{1-\hat{Y}_t}\right) + \varepsilon \bar{H}(\bar{A}, \bar{C}, \bar{L})_t\right)}$$

  Using the above model, the updated estimates of outcome are predicted over the original dataset and can be stored as $\hat{Y}_{t.Q}$.
  Now, these updated estimates are plugged into the observed outcomes of the previous time point (t-1). Then, using these outcomes, steps 2 to 4 are repeated. The whole process is back iterated till time point (t = 1).
  The average of the updated estimated outcomes at time point (t = 1) i.e. $\hat{Y}_{1.Q}$ over all the patients, would give the estimate of the target quantity i.e. $\psi_T = E[Y_T^{\bar{d}_t}]$.

- Standard error and CI The standard error for the LTMLE estimate can be obtained by the estimated influence curve of ψ_T , which is given as:

$$\widehat{IC}_{\psi T} = \left[\sum_{s=1}^{T} \hat{H}(\bar{A}, \bar{C}, \bar{L})_{s-1}\left(\hat{Y}_s^{\bar{d}_t} - \hat{Y}_{s-1}^{\bar{d}_t}\right)\right] + \hat{Y}_0^{\bar{d}_t} - \hat{\psi}_{T,ltmle}$$

  The standard error and the 95percent confidence interval are given as:

$$\sigma_{ltmle} = \sqrt{\frac{\hat{V}ar\left(\widehat{IC}\right)}{n}}$$

$$CI = \hat{\psi}_{T,ltmle} \pm 1.96\, \sigma_{ltmle}$$

  *With help of Rpackage "ltmle()", the LTMLE estimate can be easily computed along with the standard errors.

HYPOTHETICAL STRATERGY TO HANDLE ICE

Under the hypothetical strategy for estimand to handle the ICE, a hypothetical world is envisaged in which the ICE is considered to have not occurred at all. Our outcome of interest in this case is the outcome that would have been if the ICE had not occurred. These outcomes would not be observed for the patients with ICE and these need to be predicted or imputed. The causal estimand under this strategy is the effect of treatment on an outcome when the entire population is given the treatment and when it is not, provided there was no occurrence of ICE.

CAUSAL INFERENCE IN HYPOTHETICAL STRATERGY

The idea is to apply the causal inference concepts to estimate the treatment effects when there are occurrences of intercurrent events post randomized treatment arm allotment, using the hypothetical strategy, under which we envisage such a hypothetical world where the ICE had not been occurred. In accordance with the potential outcome framework, the patients for whom ICE does not occur post randomized treatment allotment, their outcome corresponds to the outcome of interest. However, for those patients, for whom ICE did occur, their potential outcome of interest (i.e., the outcome when no ICE occurs) is missing.

Using causal literature, these potential outcomes can be estimated under the causal assumptions and the unbiased estimate of treatment effect with no ICE can be computed. For our experiment, we consider the causal estimand to be the difference in an outcome when the entire population is on the treatment arm and when it is not, provided there is no occurrence of the ICE. The potential outcomes in this case are $Y^{a=1,e=0}$ and $Y^{a=0,e=0}$, respectively. The causal estimand is given as:

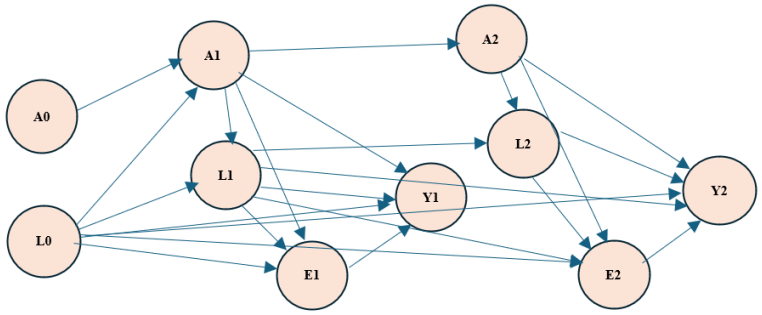
$$Treatment\ Effect = E[Y^{a=1,e=0} - Y^{a=0,e=0}]$$

LTMLE FOR HYPOTHETICAL STRATERGY

Different causal methods have been considered to estimate the effect of treatment with occurrence of intercurrent events under hypothetical strategy; referred to “Hypothetical Estimands in Clinical Trials: A Unification of Causal Inference and Missing Data Methods” (2023) by Camila Olarte Parra, Rhian M. Danielb, and Jonathan W. Bartletta.

To estimate our causal estimand of interest, we have considered double robust method of Targeted Maximum Likelihood Estimator in Longitudinal Setup (***LTMLE***). As discussed, this method uses the outcome models along with treatment and censoring models to predict the missing potential outcomes, with the step of updating the initial estimates.
We estimate the treatment effect of interest by plugging in ICEs as zero (0) and plugging the treatment regime of interest in the initial step of estimation.

Let’s consider a scenario with timepoints 0,1,2. With static treatment given at every timepoint which was randomized at baseline 0. Covariates L0, L1 and L2 and measured at every timepoint, along with outcome Y1 and Y2 after treatment administration. The intercurrent events are recorded at time point E1 and E2, after treatment administration. The DAG for this setup is given below.



The Average Treatment Effect of interest is:

$$E[Y^{a0=1,a1=1,a2=1,e1=0,e2=0} - Y^{a0=0,a1=0,a2=0,e1=0,e2=0}].$$

Firstly, data is simulated for implementation of LTMLE.

```
library(ggplot2)
library(simcausal)

######data######
DAG = DAG.empty() +
  node("A0", distr = 'rbern', prob = 0.5) +
  node("L0", distr = 'rnorm', mean = 10, sd = 0.5) +
  node("A1", distr = 'rbern', prob = ifelse(A0 == 1, 1, 0)) +
  node("L1", distr = 'rnorm', mean = 10 + 0.2*L0 + 0.1*(A1 == 1), sd = 0.5) +
  node("E1", distr = 'rbern', prob = ifelse(A1 == 0, 1/exp(0.02*L0*L1),
1/exp(0.07*L0*L1))) +
  node("Y1", distr = 'rnorm', mean = 15 + 0.2*(A1 == 1) + 0.1*(E1 == 1) +
0.02*L1*L0, sd = 1) +
  node("A2", distr = 'rbern', prob = ifelse(A1 == 1, 1, 0)) +
  node("L2", distr = 'rnorm', mean = 10 + 0.2*L1 + 0.1*(A2 == 1), sd = 0.5) +
  node("E2", distr = 'rbern', prob = ifelse(A2 == 0, 1/exp(0.001*L0*L1*L2),
1/exp(0.002*L0*L1*L2))) +
  node("Y2", distr = 'rnorm', mean = 15 + 0.2*(A2 == 1) + 0.1*(E2 == 1) +
0.006*L1*L0*L2, sd = 1)

DAG = set.DAG(DAG)
plotDAG(DAG)

data = sim(DAG, n = 1000)
```

The steps for LTMLE along with R-codes are as follows.

- **Step-1: Creating the new data incorporating treatment regime of interest and no ICE**
  From the data we have, we plug in the treatment regime of interest and no ICE, to get new dataset to predict our models over.

```
#newdata
newdata_0 = newdata_1 = data

#for treatment group (treatment given throughout)
newdata_0$E1 = newdata_0$E2 = newdata_0$A0 = newdata_0$A1 = newdata_0$A2 = 0

#for no treatment group (treatment not given throughout)
newdata_1$E1 = newdata_1$E2 = 0
newdata_1$A0 = newdata_1$A1 = newdata_1$A2 = 1
```

- **Step-2: Computing propensity scores and clever covariates.**
  Model the probability of ICEs over the past treatment, covariate and past ICE, using a binomial model, at every time point (t = 1,2). $P(E_t = 1|\bar{L}_t, \bar{A}_{t-1}, \bar{E}_{t-1})$
  Further, using these models, predict the probability of ICE and assign weights as the inverse of the product of probabilities for t,t-1,…1. for time point t.
  These weights are assigned at every time point, for the subset of the subjects having no ICE, for the other subjects the weights are zero. $W_t = \frac{1}{\prod_1^t P(E_t=1|\bar{L}_t, \bar{A}_{t-1}, \bar{E}_{t-1})}$.

```
#PS and CC : based on the ICE variable

m1_e = glm(E1 ~ A0 + L0 + L1, data = data, family = 'binomial')
m2_e = glm(E2 ~ A0 + L0 + L1 + L2 + E1, data = data, family = 'binomial')

p1_e =  gbounds((data$E1 == 1)*predict(m1_e, type = 'response', newdata = data) +
                  (data$E1 == 0)*(1 - predict(m1_e, type = 'response', newdata =
data)))
p2_e =  gbounds((data$E2 == 1)*predict(m2_e, type = 'response', newdata = data) +
                  (data$E2 == 0)*(1 - predict(m2_e, type = 'response', newdata =
data)))

#subset of patients with no ICE
subset1 <- (data$E1 == 0)
subset2 <- (data$E1 == 0) & (data$E2 == 0)

data$w1 = data$w2 = 0
data$w1[subset1] = (1/p1_e)[subset1]
data$w2[subset2] = (1/gbounds(p1_e*p2_e))[subset2]
```

- **Step-3: Outcome modelling**
  Starting from the final timepoint, we model the outcome (normalised by min-max method if continuous) over the treatment history, covariate history, ICE history and past outcomes. We then use this model to predict the outcomes by plugging the treatment of interest and no ICE.

$$E(Y_T|\bar{A}_T, \bar{L}_T, \bar{E}_T, \bar{Y}_{T-1})$$

```
# t = 2
m_y2 = glm(I((Y2 - min(Y2))/(max(Y2) - min(Y2))) ~ A0 + L0 + L1 + L2 + E1 + E2 +
I((Y1 - min(Y1))/(max(Y1) - min(Y1))) , data = data, family = 'binomial')
data$y2cap = predict(m_y2, newdata = newdata_0, type = 'response')
```

- **Step-4: Updating step**
  Now, we model the outcome from the final timepoint over the offset and logit of the predicted outcome value from step2 and weights computed in step1. Then, using this model, predict the outcome for the treatment regime of interest and no ICE.

```
m2_upd = glm(I((Y2 - min(Y2))/(max(Y2) - min(Y2))) ~ -1 + offset(qlogis(y2cap))
+ w2, data = data, family = 'binomial')
data$y2.Q =  ((predict(m2_upd, type = 'response', data =  newdata_0))*(max(data$Y2)
- min(data$Y2))) + min(data$Y2)
```

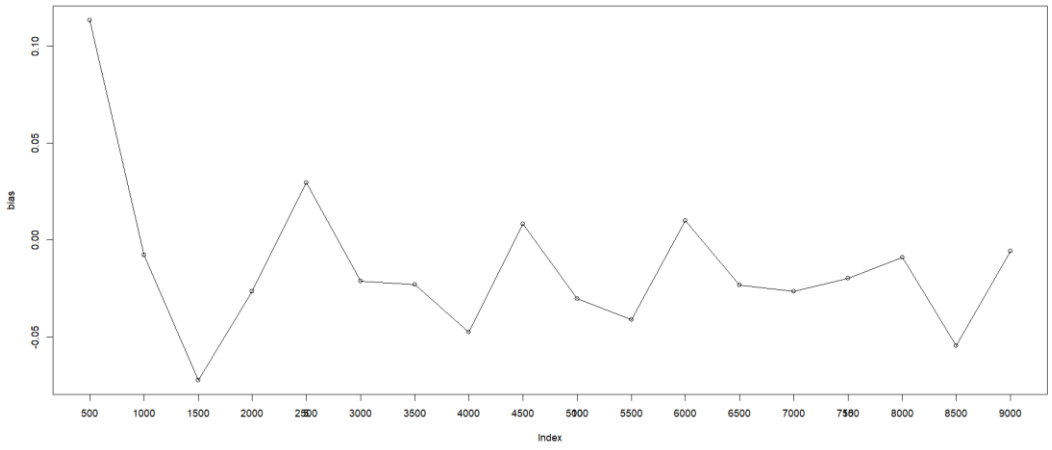
- **Step-5: Repeat steps 2 and 3 for other timepoints t…,2,1.**
  Next, we use the updated outcomes from final timepoint in time point T-1, and repeat the steps 2-3, using the corresponding weights from step1.
  This back-iteration is repeated until we get updated outcomes at time t=1.

```
# t = 1
m_y1 = glm(I((y2.Q - min(y2.Q))/(max(y2.Q) - min(y2.Q))) ~ A0 + L0 + L1 + E1 , data
= data, family = 'binomial')
data$y1cap = predict(m_y1, newdata = newdata_0, type = 'response')

m1_upd = glm(I((y2.Q - min(y2.Q))/(max(y2.Q) - min(y2.Q))) ~ -1 +
offset(qlogis(y1cap)) + w1, data = data, family = 'binomial')
data$y1.Q =  ((predict(m1_upd, type = 'response', data = newdata_0))*(max(data$Y1)
- min(data$Y1))) + min(data$Y1)

data$A0 = data$y1.Q
```

- **Step-6: ATE**
  All of the above steps are repeated for the “treatment” new data to get the final values A1.
  The final Average treatment effect is computed as the mean difference of the updated outcome A0 and A1 at time t=1, over the treatment given or not.

```
#ATE
Estimand_ATE = data$A1 - data$A0
```

The ATE estimates are further computed for different sample sizes and bias with respect to the true value from simulations was computed. From the graph below it can be seen that the bias is less than 0.1



CONCLUSION AND DISCUSSIONS

From the above simulations, the true ATE being 0.5, we can see that the use of LTLME gives an estimate near to 0.44, with bias reducing further as sample size increases. The procedure that LTMLE uses incorporates information both from the outcome model as well as treatment model, also not dropping any observations as “missing” due to occurrence of intercurrent events (ICEs). LTMLE being a semi-parametric method can also use machine-learning methods, reducing data assumptions and model misspecification, making the estimate consistent and efficient than the other methods.

It is important to note that this is a general idea of using causal methods like LTMLE and further research on different scenarios like longer follow-ups, decreasing sample sizes and different intercurrent events should be explored.

REFERENCES

- Edmonds A, Yotebieng M, Lusiama J, et al. (2011) The effect of highly active an tiretroviral therapy on the survival of HIV-infected children in a resource-deprived setting: a cohort study. PLoS Medicine.
- Guo S, Fraser MW. (2015) Propensity Score Analysis: Statistical Methods and Applications. Second ed. SAGE: Los Angeles
- Heejung Bang, James M. Robins (2005), Doubly Robust Estimation in Missing Data and Causal Inference Models, Biometrics, Volume 61, Issue 4
- Herńan MA, Robins JM (2020). Causal Inference: What If. Boca Raton: Chapman & Hall/CRC.
- Luque-Fernandez MA, Schomaker M, Rachet B, Schnitzer ME. (2018) Targeted max imum likelihood estimation for a binary treatment: A tutorial. Stat Med.
- Megan S. Schuler, Sherri Rose, (2017) Targeted Maximum Likelihood Estimation for Causal Inference in Observational Studies, American Journal of) Epidemiology, Volume 185, Issue 1 35
- Neyman, Jerzy.(1923) Sur les applications de la theorie des probabilites aux experi ences agricoles: Essai des principes. Master’s Thesis . Excerpts reprinted in English, Statistical Science, Vol. 5
- Olarte Parra, C., Daniel, R. M., & Bartlett, J. W. (2023). Hypothetical Estimands in Clinical Trials: A Unification of Causal Inference and Missing Data Methods. Statistics in Biopharmaceutical Research, 15(2), 421–432. https://doi.org/10.1080/19466315.2022.2081599
- Porter KE, Gruber S, van der Laan MJ, Sekhon JS. (2011) The relative performance of targeted maximum likelihood estimators. Int J Biostat.
- Rubin, D. B. (1979). Using Multivariate Matched Sampling and Regression Adjustment to Control Bias in Observational Studies. Journal of the American Statistical Association.
- Rubin, Donald (2005). Causal Inference Using Potential Outcomes. J. Amer. Statist. Assoc.
- Sekhon, Jasjeet (2007). The Neyman–Rubin Model of Causal Inference and Estimation via Matching Methods (PDF). The Oxford Handbook of Political Methodology. Archived from the original.
- Schomaker M, Luque-Fernandez MA, Leroy V, Davies MA. (2019) Using longitudinal targeted maximum likelihood estimation in complex settings with dynamic interventions. Stat Med. 30
- Snowden JM, Rose S, Mortimer KM. (2011) Implementation of G-computation on a simulated data set: demonstration of a causal inference technique. Am J Epidemiol.
- van der Laan MJ, Gruber S. (2012) Targeted minimum loss based estimation of causal effects of multiple time point interventions. Int J Biostat. ;8(1). doi: 10.1515/1557-4679.1370. PMID: 22611591.

ACKNOWLEDGMENTS

I would like to thank colleagues and collaborators for helpful discussions and constructive feedback during the development of this work. I also acknowledge the anonymous reviewers for their insightful comments, which significantly improved the clarity and quality of the manuscript.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Author Name: Shreya Wahi
Company: Cognizant Technology Solutions
Work Phone: 8860988595
Email: wahishreya@gmail.com