

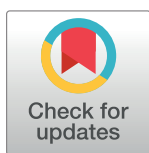
RESEARCH ARTICLE

The Targeted Maximum Likelihood estimation to estimate the causal effects of the previous tuberculosis treatment in Multidrug-resistant tuberculosis in Sudan

Adel Hussein Elduma^{1,2*}, Kourosh Holakouie-Naieni², Amir Almasi-Hashiani^{3,4}, Abbas Rahimi Foroushani², Hamdan Mustafa Hamdan Ali⁵, Muatsim Ahmed Mohammed Adam⁶, Asma Elsony⁷, Mohammad Ali Mansournia²

1 Department of Epidemiology, National Public Health Laboratory, Khartoum, Sudan, **2** Department of Epidemiology and Biostatistics, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran, **3** Department of Epidemiology, School of Health, Arak University of Medical Sciences, Arak, Iran, **4** Traditional and Complementary Medicine Research Center (TCMRC), Arak University of Medical Sciences, Arak, Iran, **5** Multidrug resistant Tuberculosis Unit- Communicable and Non-Communicable Diseases Control Directorate, Ministry of Health Sudan, Khartoum, Sudan, **6** National Tuberculosis Reference Laboratory, National Public Health Laboratory- Ministry of Health, Khartoum, Sudan, **7** The Epidemiological Laboratory (Epi-Lab), Khartoum, Sudan

* dumanet@yahoo.com



OPEN ACCESS

Citation: Elduma AH, Holakouie-Naieni K, Almasi-Hashiani A, Rahimi Foroushani A, Mustafa Hamdan Ali H, Adam MAM, et al. (2023) The Targeted Maximum Likelihood estimation to estimate the causal effects of the previous tuberculosis treatment in Multidrug-resistant tuberculosis in Sudan. *PLoS ONE* 18(1): e0279976. <https://doi.org/10.1371/journal.pone.0279976>

Editor: Selvakumar Subbian, Rutgers Biomedical and Health Sciences, UNITED STATES

Received: May 5, 2022

Accepted: December 19, 2022

Published: January 17, 2023

Copyright: © 2023 Elduma et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its [Supporting Information](#) files.

Funding: This study was partially funded by Tehran University of Medical Sciences (Project number: 9131490001). I would to state that the funders had no role in study design, data collection and analysis, decision to publish, or preparation of the

Abstract

Introduction

This study used Targeted Maximum Likelihood Estimation (TMLE) as a double robust method to estimate the causal effect of previous tuberculosis treatment history on the occurrence of multidrug-resistant tuberculosis (MDR-TB). TMLE is a method to estimate the marginal statistical parameters in case-control study design. The aim of this study was to estimate the causal effect of the previous tuberculosis treatment on the occurrence of MDR-TB using TMLE in Sudan.

Method

A case-control study design combined with TMLE was used to estimate parameters. Cases were MDR-TB patients and controls were patients who cured from tuberculosis. The history of previous TB treatment was considered the main exposure, and MDR-TB as an outcome. A designed questionnaire was used to collect a set of covariates including age, time to reach a health facility, number of times stopping treatment, gender, education level, and contact with MDR-TB cases. TMLE method was used to estimate the causal association of parameters. Statistical analysis was carried out with *ltmle* package in R-software. Result presented in graph and tables.

Results

A total number of 430 cases and 860 controls were included in this study. The estimated risk difference of the previous tuberculosis treatment was (0.189, 95% CI; 0.161, 0.218) with SE

manuscript. Also, no authors received a salary from the fund.

Competing interests: The authors declare no conflict of interest.

0.014, and p-value (<0.001). In addition, the estimated risk ratio was (16.1, 95% CI; 12.932, 20.001) with SE = 0.014 and p-value (<0.001).

Conclusion

Our findings indicated that previous tuberculosis treatment history was determined as a risk factor for MDR-TB in Sudan. Also, TMLE method can be used to estimate the risk difference and the risk ratio in a case-control study design.

Introduction

Multidrug resistant tuberculosis (MDR-TB) is defined as the infection with both isoniazid and rifampicin. Globally, there were 132222 reported cases of MDR-TB in 2020. Studies indicated that previous tuberculosis treatment and treatment interruption are considered a main cause of MDR-TB [1, 2]. Epidemiologists define case-control study as a biased sampling design. Case-control study design focuses on parametric logistic regression to calculate adjusted Odds Ratio (OR) conditional on a set of covariates. However, in order to build a causal estimate marginal population, OR should be estimated. Epidemiologists define case control as a biased sampling design of the proportion of those who have diseases compared to the target population. Case-control study design focuses on parametric logistic regression to calculate the OR conditional on a set of covariates. To build a causal estimate we have to estimate the marginal population OR [3].

The targeted maximum likelihood estimation (TMLE) is a double robust methodology uses machine learning algorithm to minimize the risk of bias [4]. Inverse probability treatment weighting (IPTW) is a causal method uses to adjust time-varying confounders by creating similar groups that examine the effect of the treatment on the exposure. IPTW method is based on the probability of exposure given the confounder which is known as propensity score (SP) [5]. IPTW has many drawbacks in case-control studies such that estimators cannot make any claims about asymptotic efficiency and efficiency issues in finite samples. Also, IPTW has disadvantage the so-called positivity violation which happened when treatment or exposure group are very rare in certain strata defined by set of covariates [6]. Thus, case-control weighted TMLE (CCW-TMLE) method provides double robust methods to estimate unbiased parameters estimation. This method is consistent if any expected parameters from outcome model given exposure and covariates or the exposure model given covariate are correct [7]. CCW-TMLE requires knowledge of the prevalence probability of the outcome to reduce the design bias [8]. In addition, CCW-TMLE estimates various parameters such as risk ratio and risk difference which are not available in traditional analysis of case-control study. In addition, TMLE can estimate marginal causal effects, correct specification and propensity score. The TMLE estimates all parameters with assumption that the exposure status of each individual does not affect the potential outcome of any other individuals. The main causal assumption is that there are no unmeasured confounders; therefore, common causes of exposure and outcome have been measured [9].

There are two broad approaches to control confounding during analysis. The first approach is to use standard regression model and the second one is to follow a causal method. Standard regression model is not able to estimate the average causal effect of the exposure in the presence of possible confounding or interaction between exposure and covariates. The reason is that, this method assumes no interaction between exposure and confounders to estimate the pool effect. What's more, the standard regression model fails to adjust for time-varying

confounding. Also, the logistic regression model in the presence of time-varying confounders influenced by previous treatment can lead to over-adjustment and collider stratification bias [10–12].

The *ltmle* (longitudinal targeted maximum likelihood estimation) is statistical double robust method can be used to estimate the causal association of the outcome regression. The advantage of TMLE method is that it can report risk ratio and risk difference of exposure in case-control studies, providing a robust estimate of parameters. In addition, the censoring mechanism based on the influence curve enables *ltmle* to estimate the standard errors and confidence intervals. *ltmle* can incorporate machine learning algorithms to provide a double robust estimator [12]. MDR-TB has been reported in many studies in Sudan. Approximately, 19% of tuberculosis cases were classified as retreated patients while 1.8% of them developed MDR-TB [13]. The aim of this study was to estimate the causal association between previous tuberculosis treatment and the occurrence of MDR-TB using TMLE in Sudan.

Material and methods

The study population consists of Multidrug resistant tuberculosis patients and tuberculosis patients who were susceptible to anti-tuberculosis drugs. The sample size was calculated according to the two-proportion equation where the null hypothesis is that proportion one (p_1) equal proportion two (p_2). The equation for the sample size calculation was $n = ([r+1]/2) * (Z_{\alpha/2} + Z_{\beta})^2 * p(1-p)/(p_1-p_2)^2$; where $p = (p_1+p_2)/2$ (p_1 is the exposure among cases and p_2 the exposure among control groups. Alpha ($Z_{\alpha/2}$) was calculated as 0.050 (2-sided), the power (Z_{β}) equaled 0.900, and r was defined as the ratio of controls to cases. Inclusion criteria for the cases were resistant to rifampin and isoniazid and laboratory confirmation. Extrapulmonary and cases died before the commencement of the study were excluded from the study. Controls were selected from cured TB patients.

Data collection

A designed questionnaire was used to collect data from study participants who met the inclusion criteria. The questionnaire was paper-based, validated and piloted prior to data collection. Data was collected by trained interviewers. The study began in May 2017 and was completed in February 2019. Taking previous TB treatment was considered the main exposure variable ($A = 1$) versus no previous history of TB treatment ($A = 0$). The outcome ($Y = 1$) was a patient infected with MDR-TB, ($Y = 0$) for cured tuberculosis patient, and (W) for possible confounder. Six covariates (w_1 to w_6) including age, time to reach health facility, number of times stop tuberculosis treatment, gender, education level, and contact with MDR-TB were determined. Other confounders of the MDR-TB include HIV [14], malnutrition, and comorbidities [15]. We assumed that history of previous TB treatment was independent from the occurrence of MDR-TB while controlling for other confounders (w) to achieve conditional exchangeability between exposure and outcome. This means that the risk of getting MDR-TB due exposure was the same as for unexposed. Covariates included in the analysis were significantly associated with MDR-TB in a previous case control study [16].

Also, we assume that in the confounders (w) stratum, each subject had a non-zero probability for both exposure conditions. In addition, there was consistency in the observation of possible outcome corresponding to the observed exposure for each study participants. With all these assumptions, we can interpret our findings as a marginal estimation. We used case control weighted TMLE which requires knowledge of the prevalence of the outcome from an external source, here we used Sudan TB data. In the *ltmle* package, we considered the prevalence of the outcome ($Y = 1$), which is (0.029), as a weighting number added in the calculation

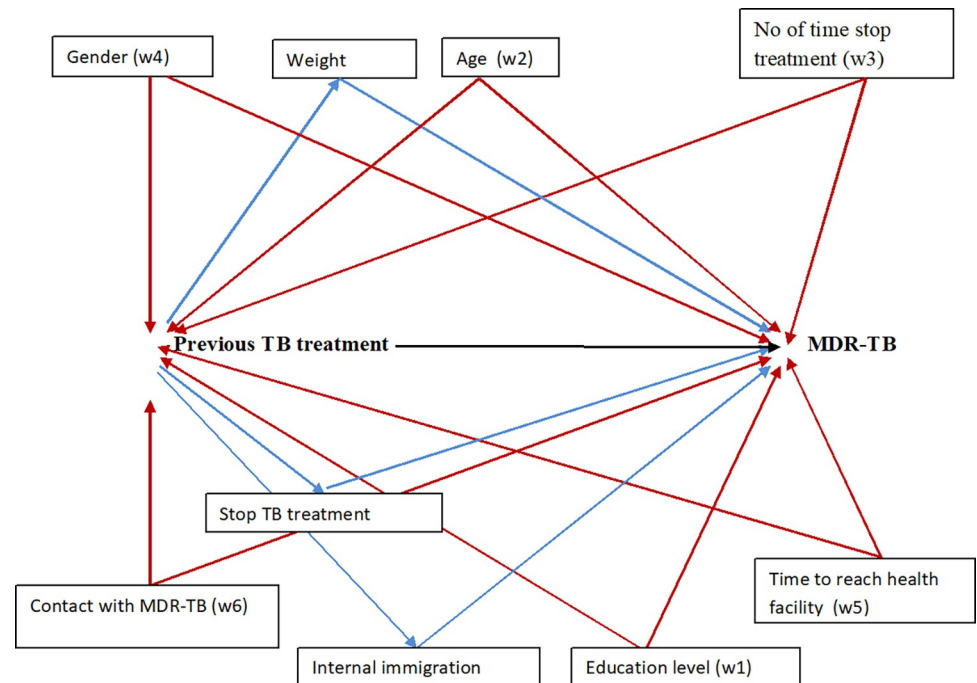


Fig 1. Direct acyclic graph (DAG) showing possible confounders in the association between previous tuberculosis treatment and Multidrug-resistant tuberculosis.

<https://doi.org/10.1371/journal.pone.0279976.g001>

process. Use of prevalence probability help to eliminate bias in case control sampling design [8]. Proper specification of a regression model requires inclusion of minimum sufficient set of confounders, selection of appropriate scale for continuous variables, and inclusion of interaction terms between predictor variables (exposure and confounders). The choice of these confounders was based on the literature and based on the result of the multiple logistic regression. The causal directed acyclic graph (causal diagrams or DAG) was used to determine a set of possible confounding variables which was included in the model (Fig 1) [17]. Direct acyclic graph (DAG) to determine minimal sufficient set (MMS) of variables that will include in the model are; Age, time to reach health facility, number of times TB treatment was stopped”, gender, education level, contact with MDR-TB.

Statistical methods

The causal method can solve these problems, and it includes three main causal approaches; inverse-probability-of-treatment weighting (IPTW), parametric g-formula, and TMLE, which incorporated these two methods.

Inverse-Probability-of-Treatment-Weighting (IPTW)

IPTW is based on propensity score (PS), which is the probability of exposure given the confounders. In IPTW, the weights of exposed group is $1/PS$, and the weight of unexposed group is $1/(1-PS)$ for the. The IPTW has the following steps:

Step 1:

Fit the exposure model either with the regression model of exposure on confounders using logistic regression or be use more advanced methods such as super learners.

Step 2:

Calculate the weight variable (W) from the exposure model fitted in step 1 as follows:
 $W = 1/PS$ for the exposed group and $W = 1/(1 - PS)$ for the unexposed group

Step 3:

Calculate the weighted mean outcome (e.g. risk) in the exposed and unexposed groups using weights (W) which was calculated in step 2, and then calculate the effect measure of interest e.g., risk difference or risk ratio.

IPTW develops a pseudo-cohort, in which confounders do not predict the exposure and the causal effects are the same as in the real cohort. So, the unadjusted (crude) analysis, which is equivalent to the weighted analysis in real cohort, yields an unbiased estimate of the exposure effect assuming that there are no unmeasured confounders and PS does not equal 0 or 1 [18]. Since the PSs are unknown and should be estimated, the IPTW is based on the exposure modeling. For instance, the exposure is regressed on the confounders and then PS and the weights are estimated as mentioned in step 1. Thus, IPTW is recommended when the exposure mechanism is known. The limitation of IPTW is that its sensitive to huge weights of strong predictors of exposure among confounders, which can introduce inefficiency and small sample bias [19].

Parametric g-formula. The parametric g-formula is based on standard outcome regression modeling and standardization which is also known as model-based standardization [20]. It has the following steps:

Step1:

Fit the regression model of outcome on the exposure and confounders. Logistic regression or more advanced methods such as super learner can be used.

Step 2:

Calculate the standardized mean outcome in the exposed group ($A = 1$ where A denote the exposure status taking (1) for the exposed and (0) for the unexposed groups by predicting the individual mean outcome for exposure (1) for all individuals, and actual values of confounders, and then average them over the individuals from the model fitted in step1. Likewise, calculate the standardized mean outcome in the unexposed group ($A = 0$) by predicting the individual mean outcome for exposure (0) for all individuals, and actual values of confounders, and then average them over the individuals from the model fitted in step 1. Then, calculate the effect measure of interest such as risk difference or risk ratio [7].

Targeted maximum likelihood estimation

The TMLE is a combination between IPTW, and Parametric g-Formula to produce a double robust causal estimation, which has two possibilities to adjust for confounders. This method provides an unbiased estimate of the causal effect if either exposure or outcome model is correct [21].

The TMLE has the following steps

First, outcome regression model on the exposure and confounders by using logistic regression for the treatment history Then we used exposure model which is the regression model of exposure on confounders variables. We calculate the variable H from the exposure model We refit the outcome regression model from step 1 by adding the variable H so that the coefficients of the model do not change. We refit the outcome regression model from step 1 by adding the variable H so that the coefficients of the model do not change. We calculate the standardized mean outcome (risk) in the exposed group as ($A = 1$). Similarly, calculate the standardized mean outcome (e.g., risk) in the unexposed group ($A = 0$).

If any of outcome regression model or exposure regression model is correct, then TMLE gives an unbiased estimates of the causal effects of the treatment history [22]. Confidence intervals can be calculated using the influence function (IF).

Model specification

There are several modeling approaches that can be used to build relationship between variables such as neural network, Bayesian methods, classification and regression trees, random forests, generalized boosted regression, and generalized additive models. A machine learning approach, which is called Super Learner (SL) is a powerful method for the modeling of exposure and outcome in the causal inference. The super learner is a combination of several regression models such as logistic regression, neural network, Bayesian approach, classification and regression trees, random forests, generalized boosted regression, generalized additive models, etc. In super learner method, a best weight for each model can obtain by the mean of square difference between observed value and the predicted value of the outcome [23].

For the analysis, we used the newly developed `ltmle` R package, which is the targeted maximum likelihood estimation (TMLE) for estimating parameters for both point estimate and longitudinal data. `ltmle` apply estimator which can able to estimate ATE, causal risk ratio, and causal odds ratio. This package is flexible and can be applied in our study design as case control [24]. First, we specified the structural causal model (SCM) in order to encode knowledge on covariates relationship. Two important assumptions that allow identification of the distribution of our causal model are the sequential randomization and positivity assumption. Under these two assumptions, the counterfactual model can be identified by the longitudinal G-computation formula. The first assumption was that no unmeasured confounders exist, while the second assumption was that each observation has a positive probability.

In addition to additive effect, risk ratio and odds ratio, `ltmle` function provides estimate for mean intervention. Our data was entered into R software by formulating a data frame, which is working platform for the `ltmle` package to work. Since our data was a case-control study, we identified treatment (exposure) as node (A) which is defined by `Anodes`, and the outcome node (Y) specified with `Y` in the `ltmle` package. We didn't have time varying variable so we put `Null` for `L` nodes. We specified (0,1) for the `abar` (binary vector), and assigned weight to our data sample by using `observation-weights = as vector (Wt)` function. We used two regression formula, a character vector `Qform`, and `qform`, where our regression model including outcome (Y), exposure (A) and set of covariates were added. We put `iptw` as `FLASE` which led to calculate both TMLE and IPTW. We used the variance method "ic" which computed a robust variance estimator using TMLE and the influence curve based variance estimator [25]. `Qform` and `qform` were working as a design matrix which allowed calling the Super Learner (SL).

We used the Super Learner algorithm to estimate the component of `Q` and `q` by implementing (SL) package. The SL algorithm chooses the best weighted estimator through a validation process. The SL has a lower bias compared to single parametric model and consistent when we estimate the `Q` and `q` components [26]. Estimating both `Q` and `q` was provided a consistent TMLE, which achieved the best possible variance. When SL was used, all variables on the right side of the formula were considered predictors in the regression model. Using `ltmle` provides a summary that returns the estimator's parameter values, estimated variance, 95% confidence interval, and p.value. Summary of `ltmle` returned also the additive treatment effect for each of the two objects in the `abar` list. Moreover, relative risk and odds ratio were calculated along with their variance, confidence interval, and p-value [27]. We used R software (version 3.5.3) for data analysis and this software used super learner to provide valid statistical inference [28].

Result

A total number of 430 cases and 860 controls with mean age 35.3 ± 13.17 for cases and 38.27 ± 16.12 for controls were included in this study. Study participants consist of 896 males (295 cases, 601 controls) and 374 females (130 cases, 294 controls). Participants with primary and secondary education were 27.1% and 27.6% respectively. Participants without education represented 19.5% of the study population. A total 292 (67.9%) of cases had history of previous TB treatment while only 26 (3.0%) of the control had received previous TB treatment (Table 1).

Table 1. Demographic and clinical characteristics of stud participants.

	Cases (430)	Control (860)	Total (1290)
Age			
Mean	35.31	38.25	
SD	13.17	16.12	
Gender			
Male	30 (69.7%)	606 (70.4%)	906
Female	130 (30.3%)	254 (29.6%)	384
Education level			
Without education	60 (14.1%)	192 (22.3%)	252
Khalwa	39 (9.2%)	55 (6.4%)	94
Primary	129 (29.9%)	220 (25.6%)	349
Secondary	117 (26.8%)	240 (27.9%)	357
Diploma	8 (1.9%)	1 (0.1%)	9
University + postgraduate	77 (18.1%)	152 (17.7%)	229
Time to reach health center (hs)			
Mean	2.2	1.9	
Median	1.2	1.5	
SD	4.4	1.2	
Number of family member (per room)			
Mean	7.5	8.3	
Median	7	8	
SD	3.6	3.6	
Contact with TB patients			
Yes	97 (22.6%)	58 (6.8%)	155
No	309 (71.9%)	783 (91.0%)	1077
I don't remember	24 (5.5%)	19 (2.2%)	43
Contact with MDR-TB patients			
Yes	64 (14.9%)	2 (0.2%)	66
No	339 (78.8%)	844 (98.2%)	1183
I don't remember	27 (6.3%)	14 (1.6%)	41
History of previous treatment			
Yes	292 (67.9%)	26 (3.0%)	318
No	138 (32.1%)	834 (97.0%)	972
Diabetes			
Yes	30 (7.0%)	46 (5.3%)	76
No	400 (93.0%)	814 (94.7%)	1214
Fever			
Yes	242 (56.3%)	379 (55.9%)	723
No	188 (43.7%)	370 (44.1%)	567

<https://doi.org/10.1371/journal.pone.0279976.t001>

Table 2. Measurement of Targeted Maximum Likelihood Estimation (TMLE).

Measurement	Estimate (95% CI)	Standard error	p-value
Risk difference	0.189 (0.16135, 0.21761)	0.014353	<0.001
Risk Ratio	16.1 (12.93, 20.00)	0.014353	0000 <0.001

<https://doi.org/10.1371/journal.pone.0279976.t002>

Previous TB treatment history was included as exposure variable, cases and controls as outcome variables, and a set of confounders were included in the analysis. These confounders included age, time to reach health facility, Number of times stop treatment, gender, education level, and contact with MDR-TB. Summary of ltmle returns treatment (exposure in our setting) estimator and the parameter estimate was (0.202, 95% CI; 0.174, 0.23009) with standard error (0.014309) and p.value (<0.001). Also, the estimator and parameter estimate of the control group was (0.012562, 95% CI; 0.010456, 0.014669) with standard error (0.0010748) and p.value (<0.001). The causal risk difference estimate was (0.18948, 95% CI; 0.16135, 0.21761) with SE 0.014353, and p. value (<0.001). Moreover, the causal Risk ratio was 21(16.083, 95% CI; 12.932, 20.001) with SE equal 0.014353 and p. value (<0.001) (Table 2). This result suggested that the risk among exposed was 16 times compared with the risk of unexposed. Patients who had been exposed to previous TB treatment history were 18% more likely develop MDR-TB compared with patients who had not exposed to tb treatment. Narrow confidence interval for the estimated parameters indicates an strong statistical association.

Discussion

In this study, we implemented a double robust method to estimate the causal effect of parameters in a case-control study. We implemented TMLE as a robust method to estimate the causal effect of the previous TB treatment history on the occurrence of the MDR-TB in Sudan. Our main finding was that the previous TB treatment was identifies as a risk factor for MDR-TB. We used TMLE as statistical method to eliminate the effect of confounding variables that can affect parameters estimation in the traditional case-control study [29]. In observational studies, such as cross-sectional studies and case-control studies, we cannot establish a causal association between exposure and outcome.

In traditional case-control study it is not able to estimate the risk difference and the risk ratio. Using ordinary logistic regression depends on the selection algorithms, such as stepwise regression, which can result in missing very important confounders. Also, logistic regression reports only the Odds ratio for the association between the exposure and the outcome. To establish causal effect association, risk ratio and risk difference need to be estimated. Thus, R ltmle package is able to estimate these causal effects association of parameters using double robust method. Therefore, we applied ltmle in this study to establish a causal association the previous TB treatment (exposure) and the development of the MDR-TB (outcome). TMLE was used by study conducted to model the treatment effect of MDR-TB patient data metanalysis. In this study targeted maximum likelihood estimate was used to provide robust estimation of parameters [30]. TMLE provided a precise estimation for the causal parameters and less prone to bias compared to other estimators such as inverse probability weighting estimator (IPW) and propensity score [31].

The advantage of using such method can help researcher to establish a causal association for observational studies. Compared with ordinary logistic regression, with TMLE can remove all biases due to model misspecification [32]. TMLE was also used by a group of researchers to assess the effect of MDR-TB outcome in 31 observational studies We have shown in this study

a successful implementation of *ltmle* in a case-control study design using an original data set [33]. Our Findings show that bias can be eliminated by using machine learner algorithms through the Super Leaner package in *ltmle*. Also, our study provided an evidence that the *ltmle* can be considered as a general template for building an efficient estimator for the targeted parameters [34].

The result of study providing a guide for people who are interested in TB and MDR-TB to conduct similar studies in their setting. Many studies reported that tuberculosis treatment interruption is the main cause of MDR-TB. In study conducted in Bangladesh, previous TB treatment was associated with the MDR-TB [35]. In a systematic review conducted in Ethiopia, the burden of the MDR-TB was high in Ethiopian setting particularly among previously treated TB patents [36]. In study conducted in Vietnam among TB patients indicated that an increase in the number of MDR-TB cases and previous TB treatment are the greatest risk of poor treatment outcome [37].

Using TMLE as statistical tool estimate risk ratio and risk difference was providing an evidence to scientists that previous treatment is a risk for the MDR-TB occurrence. Many researchers have interest to conduct observational studies, so this study provided an evidence that TMLE, as a doubled robust method, can be applied in traditional case-control study design. The existence of R Software as a free source will help researchers, particularly those in countries with limited resources, to apply TMLE in their research. In addition, further researches are needed to study the risk factors associated with MDR-TB, particularly in countries with a high burden of TB. One limitation of using TMLE is that the estimation of variance is challenging when the positivity assumption is violated. Secondly, in a rare binary outcome, to estimate the outcome models in the existence of a vector of covariates can produce a bias variance in moderate sample size. In this study we demonstrated the use of prevalence probability in weighted case-control using targeted maximum likelihood estimation to obtain efficient causal parameter. We also able to estimate risk ratio and risk difference in a traditional case-control study design. Moreover, previous history of TB treatment was identifies as a risk factor for the occurrence of MDR-TB in Sudan.

Supporting information

S1 File.
(CSV)

Acknowledgments

Ethical approval

This study was approved by the ethical committee at Tehran University of Medical Sciences under the ethical approval code: IR, TUMS.SPH.REC.1395.1718. Also, the study was approved by the National ethics committee at the Ministry of Health Sudan (NO: FMOH/rd/1/104). Written informed consent was handed to participants who were able to read and their signature was taken. For those who were not able to read the information in the informed consent was read out by the interviewers. The objectives of the study were explained in separated sheet which was attached in the questionnaire and read and explained to participants not able to read. The interviewer emphasized that the participants had the right not to participate in the study if he didn't want to.

Author Contributions

Conceptualization: Adel Hussein Elduma, Kourosh Holakouie-Naieni, Hamdan Mustafa Hamdan Ali, Mohammad Ali Mansournia.

Data curation: Adel Hussein Elduma.

Formal analysis: Abbas Rahimi Foroushani, Mohammad Ali Mansournia.

Investigation: Adel Hussein Elduma, Muatsim Ahmed Mohammed Adam, Asma Elsony.

Methodology: Adel Hussein Elduma, Kourosh Holakouie-Naieni, Abbas Rahimi Foroushani, Muatsim Ahmed Mohammed Adam, Mohammad Ali Mansournia.

Project administration: Kourosh Holakouie-Naieni.

Resources: Amir Almasi-Hashiani, Abbas Rahimi Foroushani, Hamdan Mustafa Hamdan Ali, Asma Elsony.

Supervision: Kourosh Holakouie-Naieni, Abbas Rahimi Foroushani, Mohammad Ali Mansournia.

Validation: Kourosh Holakouie-Naieni, Amir Almasi-Hashiani, Hamdan Mustafa Hamdan Ali, Asma Elsony.

Visualization: Amir Almasi-Hashiani.

Writing – original draft: Adel Hussein Elduma.

Writing – review & editing: Adel Hussein Elduma, Hamdan Mustafa Hamdan Ali, Mohammad Ali Mansournia.

References

1. World Health Organization, "Global tuberculosis report 2017," 2017. [Online]. Available: https://www.who.int/tb/publications/global_report/gtbr2017_main_text.pdf.
2. Seung K. J., Keshavjee S., and Rich M. L., "Multidrug-Resistant Tuberculosis and Extensively Drug-Resistant Tuberculosis," *Cold Spring Harb. Perspect. Med.*, vol. 5, no. 9, Sep. 2015, <https://doi.org/10.1101/cshperspect.a017863> PMID: 25918181
3. Suzuki E, Tsuda T, Mitsunashi T, Mansournia MA, Yamamoto E, "Errors in causal inference: an organizational schema for systematic error and random error," *Ann Epidemiol*, vol. 26, no. 11, pp. 788–893, 2016. <https://doi.org/10.1016/j.annepidem.2016.09.008> PMID: 27771142
4. Schuler M. S. and Rose S., "Targeted Maximum Likelihood Estimation for Causal Inference in Observational Studies," *Am. J. Epidemiol.*, vol. 185, no. 1, pp. 65–73, Jan. 2017, <https://doi.org/10.1093/aje/kww165> PMID: 27941068
5. Almasi-Hashiani Amir, Nedjat Saharnaz, Mansournia Mohammad Ali, "Causal Methods for Observational Research: A Primer," *Arch Iran Med*, vol. 21, pp. 164–169, 2018. PMID: 29693407
6. Petersen ML P.K., Gruber S, et al, "Diagnosing and responding to violations in the positivity assumption.," vol. 21, no. 1, pp. 31–54, 2011.
7. Mark J. van der Laan and Sherri Rose, "Targeted Learning: Causal Inference for Observational and Experimental Data".
8. van der Laan MJ, "Estimation based on case-control designs with known prevalence probability.," *Int J Biostat*, vol. 4, no. 1, 2008. <https://doi.org/10.2202/1557-4679.1114> PMID: 22462152
9. Rose S, van der Laan M, "A double robust approach to causal effects in case-control studies," *Am J Epidemiol*, vol. 179, no. 6, pp. 663–669, 2014. <https://doi.org/10.1093/aje/kwt318> PMID: 24488515
10. Mansournia MA, Etminan M, Danaei G, Kaufman JS, Collins G, "Handling time varying confounding in observational research," *Bmj*, vol. 359, p. j4587, 2017, <https://doi.org/10.1136/bmj.j4587> PMID: 29038130
11. Shakiba M, Mansournia MA, Salari A, Soori H, Mansournia N, Kaufman JS, "Accounting for Time-varying Confounding in the Relation between Obesity and Coronary Heart Disease: Analysis with G-

- estimation, the Atherosclerosis Risk in Communities (ARIC) study,” *Am J Epidemiol*, 2017, <https://doi.org/10.1093/aje/kwx360> PMID: 29155924
12. van der Laan MJ, and Gruber S., “Targeted minimum loss based estimation of causal effects of multiple time point interventions,” *Int J Biostat*, vol. 8, no. 1, 2012. <https://doi.org/10.1515/1557-4679.1370> PMID: 22611591
 13. “Global tuberculosis report 2019.” <https://www.who.int/publications-detail-redirect/9789241565714> (accessed Aug. 11, 2022).
 14. Ambaye G. Y. and Tsegaye G. W., “Factors Associated with Multi-Drug Resistant Tuberculosis among TB Patients in Selected Treatment Centers of Amhara Region: A Case-Control Study,” *Ethiop. J. Health Sci.*, vol. 31, no. 1, pp. 25–34, Jan. 2021, <https://doi.org/10.4314/ejhs.v31i1.4> PMID: 34158749
 15. Wakjira M. K., Sandy P. T., and Mavhandu-Mudzusi A. H., “Treatment outcomes of patients with MDR-TB and its determinants at referral hospitals in Ethiopia,” *PLOS ONE*, vol. 17, no. 2, p. e0262318, Feb. 2022, <https://doi.org/10.1371/journal.pone.0262318> PMID: 35176035
 16. Elduma A. H. et al., “Assessment of the risk factors associated with multidrug-resistant tuberculosis in Sudan: a case-control study,” *Epidemiol Health*, vol. 41, no. 0, pp. e2019014–0, Apr. 2019, <https://doi.org/10.4178/epih.e2019014> PMID: 31010280
 17. Mansournia MA, Higgins JP, Sterne JA, Hernan MA, “Biases in Randomized Trials: A Conversation Between Trialists and Epidemiologists,” *Epidemiology*, vol. 28, pp. 54–59, 2017, <https://doi.org/10.1097/EDE.0000000000000564> PMID: 27748683
 18. Hernán MA, Robins JM, *Causal Inference*. Boca Raton: Chapman & Hall/CRC, forthcoming. 2018.
 19. Petersen ML, Porter KE, Gruber S, Wang Y, van der Laan MJ, “Diagnosing and responding to violations in the positivity assumption.,” *Stat Methods Med Res*, vol. 21, pp. 31–54, 2012. <https://doi.org/10.1177/0962280210386207> PMID: 21030422
 20. Gharibzadeh S, Mansournia MA, Rahimiforoushani A, Alizadeh A, Amouzegar A, Mehrabani-Zeinabad K, et al. “Comparing different propensity score estimation methods for estimating the marginal causal effect through standardization to propensity scores,” *Commun Stat Simul Comput*, pp. 1–13, 2017, <https://doi.org/10.1080/03610918.2017.1300267>
 21. Schuler MS and Rose S, “Targeted Maximum Likelihood Estimation for Causal Inference in Observational Studies,” *Am J Epidemiol*, vol. 185, pp. 65–73, 2017, <https://doi.org/10.1093/aje/kww165> PMID: 27941068
 22. Rose S, “Causal Inference for Case-Control Studies,” University of California, Berkeley, 2011.
 23. Rose S, “Mortality risk score prediction in an elderly population using machine learning.,” *Am J Epidemiol*, vol. 177, no. 2013, pp. 443–452, 2013, <https://doi.org/10.1093/aje/kws241> PMID: 23364879
 24. Robins James M., “Commentary on ‘Using inverse weighting and predictive inference to estimate the effects of time-varying treatments on the discrete-time hazard,’” *Stat. Med.*, vol. 21, pp. 1663–1680, 2002.
 25. Lendle, J.S. Samuel D., Petersen Maya L., van der Laan Mark J., “Itmle: An R Package Implementing Targeted Minimum Loss-Based Estimation for Longitudinal Data,” *J. Stat. Softw.*, vol. 81, pp. 1–20, 2017.
 26. van der Laan Mark J., Polley Eric C, and, Hubbard Alan E., “Super Learner,” in *Statistical Applications in Genetics and Molecular Biology*, 2007, pp. 1–21. <https://doi.org/10.2202/1544-6115.1309> PMID: 17910531
 27. Schwab J, L.S., Petersen M, van der Laan M, “Itmle: Longitudinal Targeted Maximum Likelihood Estimation.,” 2017, [Online]. Available: <https://CRAN.R-project.org/package=itmle>.
 28. R Core Team, “R: A language and environment for statistical computing.” R Foundation for Statistical Computing, Vienna, Austria, 2014. [Online]. Available: <http://www.R-project.org/>.
 29. “Association of medical male circumcision and sexually transmitted infections in a population-based study using targeted maximum likelihood estimation—PMC.” <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8425067/> (accessed Aug. 02, 2022).
 30. Liu Y. et al., “Modeling treatment effect modification in multidrug-resistant tuberculosis in an individual patient data meta-analysis,” *Stat. Methods Med. Res.*, vol. 31, no. 4, p. 689, Apr. 2022, <https://doi.org/10.1177/09622802211046383> PMID: 34903098
 31. Neugebauer Romain, van der Laan Mark, “Why prefer double robust estimators in causal inference?,” *J. Stat. Plan. Inference*, vol. 129, pp. 405–426, 2005.
 32. Gruber S., Phillips R. V., Lee H., and van der Laan M. J., “Data-Adaptive Selection of the Propensity Score Truncation Level for Inverse Probability Weighted and Targeted Maximum Likelihood Estimators of Marginal Point Treatment Effects,” *Am. J. Epidemiol.*, p. kwac087, May 2022, <https://doi.org/10.1093/aje/kwac087> PMID: 35512316

33. Wang G., Schnitzer M. E., Menzies D., Viikklepp P., Holtz T. H., and Benedetti A., "Estimating treatment importance in multidrug-resistant tuberculosis using Targeted Learning: An observational individual patient data network meta-analysis," *Biometrics*, vol. 76, no. 3, pp. 1007–1016, Sep. 2020, <https://doi.org/10.1111/biom.13210> PMID: 31868919
34. Hampel Frank R., "The Influence Curve and Its Role in Robust Estimation," *J. Am. Stat. Assoc.*, vol. 69, pp. 383–393, 1974.
35. Rifat M., Hall J., Oldmeadow C., Husain A., Hinderaker S. G., and Milton A. H., "Factors related to previous tuberculosis treatment of patients with multidrug-resistant tuberculosis in Bangladesh," *BMJ Open*, vol. 5, no. 9, p. e008273, Sep. 2015, <https://doi.org/10.1136/bmjopen-2015-008273> PMID: 26351185
36. Eshetie S. et al., "Multidrug resistant tuberculosis in Ethiopian settings and its association with previous history of anti-tuberculosis treatment: a systematic review and meta-analysis," *BMC Infect. Dis.*, vol. 17, no. 1, p. 219, Mar. 2017, <https://doi.org/10.1186/s12879-017-2323-y> PMID: 28320336
37. Van L. H. et al., "Risk factors for poor treatment outcomes of 2266 multidrug-resistant tuberculosis cases in Ho Chi Minh City: a retrospective study," *BMC Infect. Dis.*, vol. 20, p. 164, Feb. 2020, <https://doi.org/10.1186/s12879-020-4887-1> PMID: 32087682