Annals of Epidemiology xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Annals of Epidemiology

journal homepage: sciencedirect.com/journal/annals-of-epidemiology

Review article

Application of targeted maximum likelihood estimation in public health and epidemiological studies: a systematic review

Matthew J. Smith, PhD [a,*,1], Rachael V. Phillips, PhD [b], Miguel Angel Luque-Fernandez, PhD [a,c,2], Camille Maringe, PhD [a,2]

[a] Inequalities in Cancer Outcomes Network, London School of Hygiene and Tropical Medicine, London, UK
[b] Division of Biostatistics, School of Public Health, University of California at Berkeley, Berkeley, CA
[c] Department of Statistics and Operations Research, University of Granada, Granada, Spain

ARTICLE INFO

Article history:
Received 3 March 2023
Received in revised form 24 May 2023
Accepted 6 June 2023
Available online xxxx

Keywords:
Targeted Maximum Likelihood Estimation (TMLE)
Epidemiology
Observational studies
Causal inference
Systematic review

ABSTRACT

Purpose: The targeted maximum likelihood estimation (TMLE) statistical data analysis framework integrates machine learning, statistical theory, and statistical inference to provide a least biased, efficient, and robust strategy for estimation and inference of a variety of statistical and causal parameters. We describe and evaluate the epidemiological applications that have benefited from recent methodological developments.

Methods: We conducted a systematic literature review in PubMed for articles that applied any form of TMLE in observational studies. We summarized the epidemiological discipline, geographical location, expertize of the authors, and TMLE methods over time. We used the Roadmap of Targeted Learning and Causal Inference to extract key methodological aspects of the publications. We showcase the contributions to the literature of these TMLE results.

Results: Of the 89 publications included, 33% originated from the University of California at Berkeley, where the framework was first developed by Professor Mark van der Laan. By 2022, 59% of the publications originated from outside the United States and explored up to seven different epidemiological disciplines in 2021–2022. Double-robustness, bias reduction, and model misspecification were the main motivations that drew researchers toward the TMLE framework. Through time, a wide variety of methodological, tutorial, and software-specific articles were cited, owing to the constant growth of methodological developments around TMLE.

Conclusions: There is a clear dissemination trend of the TMLE framework to various epidemiological disciplines and to increasing numbers of geographical areas. The availability of R packages, publication of tutorial papers, and involvement of methodological experts in applied publications have contributed to an exponential increase in the number of studies that understood the benefits and adoption of TMLE.

Background

Public health decisions across many clinical specialties are often informed by research exploring the relationship between exposures and patient health outcomes. These relationships are often susceptible to confounding bias, which requires sometimes complex statistical methodology to minimize. Randomized controlled trials (RCT) are considered the gold standard because, through randomization of subjects to a treatment, they reduce the possibility of bias. Observational data offer invaluable opportunities to study relationships in contexts where clinical trials might prove infeasible or unethical, as well as for studying groups of the population typically excluded from trials or beyond the initial target population. Under correct adjustment for selection bias, missingness, interference, and confounding, observational data complement the evidence coming from RCTs.

In both RCT and observational studies, the exposure–outcome relationship is of interest. Methodological statistical developments for causal inference attempt to produce the least biased estimate of the relationship along with accurate inference. G-computation, propensity score (PS), and inverse probability of treatment

* Corresponding author.
E-mail address: matt.smith@lshtm.ac.uk (M.J. Smith).
[1] Present address: London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, United Kingdom.
[2] Both authors contributed equally to this work.

https://doi.org/10.1016/j.annepidem.2023.06.004

Please cite this article as: M.J. Smith, R.V. Phillips, M.A. Luque-Fernandez et al., Application of targeted maximum likelihood estimation in public health and epidemiological studies: a systematic review, Annals of Epidemiology, https://doi.org/10.1016/j.annepidem.2023.06.004

weighting (IPTW) estimators rely on parametric modeling assumptions, which are susceptible to model misspecification. Double-robust methods, such as augmented inverse probability of treatment weighting (AIPTW) and targeted maximum likelihood estimation (TMLE), aim to minimize model misspecification by requiring estimation of both the outcome and exposure mechanisms. They provide a consistent estimator as long as either the outcome or exposure model is correctly specified. Double-robust methods often outperform single-robust methods in point and interval estimation [1,2].

TMLE, also known as targeted minimum loss-based estimation, was introduced by van der Laan and Rubin in 2006 [3]. In general, TMLE is a two-step process that involves (1) initial estimation of the outcome and intervention models, and then (2) in a "targeting" step, uses information from them to optimize the bias-variance trade-off for the target estimand (e.g., average treatment effect [ATE]), rather than the whole outcome probability distribution. Furthermore, to avoid model misspecification, ensemble machine learning algorithms are used to estimate the initial models. In particular, the SuperLearner (SL) algorithm for stacked ensemble machine learning is most commonly used as it is theoretically grounded and proven to perform optimally in large samples [4].

We lightly detail the technical steps involved in the TMLE of the ATE, i.e., the effect of a binary exposure A on a postexposure outcome Y, adjusted by baseline covariates W [5]. The prediction function for the mean outcome Y, given exposure A and covariates W, is estimated, most commonly, using SL. We could use this estimated prediction function, $\hat{E}[Y|A, W]$, to arrive at an estimate of the ATE. Specifically, we would obtain predicted outcomes under a counterfactual scenario where all subjects receive the exposure/treatment versus another scenario where no one receives it. The average difference between these predicted counterfactual outcomes is an estimate of the ATE. However, formal statistical inference (i.e., confidence intervals and P-values) cannot be obtained for this estimate and it is susceptible to residual confounding; the latter can be reduced by using the information on how each individual was assigned or allocated to each level of the exposure. We, therefore, estimate the function for predicting the probability of being exposed, given the covariates W, using SL (exposure model, i.e., PS). These first steps are common to other double-robust estimators of the ATE, such as AIPTW. We then calculate the so-called "clever covariate" for the ATE, which is the individual values of the binary exposure weighted by the predicted probabilities of the exposure, given W. This is similar to IPTW, except here we weight the predicted probability of each exposure level instead of the outcome. The fluctuation parameter (ϵ) describes the difference between the observed outcome Y and the initial predictions of the outcome from the outcome model. It is calculated through maximum likelihood estimation by regressing the clever covariate on the observed outcome. When the fluctuation parameter is estimated to be close to 0, there is little difference between the observed and predicted outcomes; thus, the PS does not provide additional information for the initial estimate of the outcome model because it was correctly specified. If the fluctuation parameter is not close to 0, then this indicates the presence of residual confounding in the initial estimate. The initial outcome model's predictions for each level of the binary exposure are updated using the fluctuation parameter ϵ as a weight, and the final ATE estimate is calculated from these updated estimates. The functional delta method based on the influence function can be used to derive the standard error of the ATE and construct Wald-type confidence intervals.

Since 2006, the TMLE framework has experienced a growing number of theoretical and applied developments, and it has expanded further after a book that shared the TMLE framework to the international community of applied researchers was published in 2011 [2]. Targeting specifically applied researchers, efforts were made to provide lay-language descriptions of the framework and exemplify its applications [5–7]. Furthermore, in 2018, a second book was published disseminating more advanced applications of the TMLE framework to data scientists with a particular focus on longitudinal settings [8]. TMLE is a robust framework for statistical analysis in clinical, observational, and randomized studies. Since 2016, the American Causal Inference Conference has hosted a data challenge in which teams compete to estimate a causal effect in simulated datasets based on real-world data, such as from healthcare or education [9]. The competition is a proving ground for cutting-edge causal inference methods that have the potential to transform program evaluation. TMLE has consistently been a top-performing method [10].

The use of robust statistical methods is key to obtaining reliable results for public health and epidemiological research and maximizing their benefit to society. Evidence shows that TMLE, by blending flexible machine learning methods and causal inference framework, is one such step toward robust causal claims that bear significant and practical effects. We reviewed the literature around public health and epidemiological applications of TMLE to date, alongside key TMLE developments over the last 20 years. We highlight the speed at which the field has developed and spread through the scientific community, and identify areas for further development to increase the utility of the TMLE framework in epidemiological and applied research.

Methods

Protocol registration and reporting standards

This study is reported using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guideline. We registered this systematic review with PROSPERO (ID: CRD42022328482).

Information sources

We searched the PubMed medical literature database for published epidemiological studies using TMLE in any epidemiological field (i.e., observational settings in biomedical sciences, including clinical epidemiology and public health). We searched for publications from any time up to December 31, 2022, the date the search was executed. The search strategy comprised two specific groups of search terms focusing on TMLE and epidemiology. Relevant Mesh headings were included along with free-text terms, which were searched for in the title, abstract, and keyword fields. We used general and specific TMLE search terms, such as "targeted maximum likelihood estimation," "targeted minimum loss-based estimation," and "targeted machine learning." Epidemiological search terms included "epidemiology," "public health," "population," or "treatment." The two specific groups of terms were combined with "AND" to retrieve the final set of results. Search strategies were developed with an information specialist (MALF). The full search strategy is shown in Table 1.

Eligibility criteria

We excluded studies that did not report the use of TMLE as a tool to explore their estimand of interest. We also excluded experimental studies, such as RCTs (n = 18, Appendix Table A.1), [11–28] because they are designed to minimize confounding bias through randomization, which makes them fundamentally different from observational studies that heavily rely on the use of statistical methods to minimize confounding bias. By focusing on observational studies, we provide a more detailed and nuanced understanding of how TMLE

Table 1
Boolean search queries

Query	Boolean terms	Results
#1	(epidemiology OR (public AND health) OR population OR treat*)	11,459,953
#2	("targeted maximum likelihood estimation") OR ("targeted minimum loss based estimation") OR ("targeted minimum loss-based estimation") OR ("TMLE") OR ("targeted machine learning") OR ("targeted learning") OR ("targeted machine-learning")	315
#3	#1 AND #2	254

can be used to address confounding bias and to identify gaps in the existing literature on TMLE in observational studies, which can help to guide future research.

We did not consider manuscripts that compared the performance of TMLE to other estimators when there was no new development proposed, even if there was an applied question of interest [29–31]. Studies were restricted to the English language and primary research studies. Secondary research studies, such as reviews and comments of TMLE, conference abstracts and brief reports, and preprints, were not searched. We classified the retained manuscripts into observational, methodological, and tutorial articles. TMLE methodological development articles and tutorials were considered separately, even if they contained a methodological development specifically designed to investigate an epidemiological question within the same article. We make reference to these methodological articles throughout this review, as they underpin the applied publications.

Study selection, data extraction, and management

All retrieved publications were imported into the Endnote reference software where they were checked for duplication. Two of the three lead researchers (authors MJS, MALF, and CM) were randomly allocated two-thirds of the 254 articles, to screen titles and abstracts of each publication independently and classify them into (1) observational, (2) methodological developments, (3) tutorial, (4) systematic review, (5) RCT, or (6) not relevant (Fig. 1). Discordant classifications were discussed and adjudicated by the third independent reviewer, where necessary. Two researchers (authors MJS and CM) independently reviewed the full text of all eligible observational publications for data extraction.

Results

We found 254 unique publications published prior to December 31, 2022, in PubMed (Fig. 1). Of these, 102 articles were methodological development (including theoretical- or software-based), eight were tutorials, five were systematic reviews, and 18 were RCTs. Of the 32 articles that were not relevant, three mentioned "TMLE" only in the author fields, one was a discussion of currently existing methods, some were assessments of *learning* (educational) programs that are *targeted* toward clinical environments, and others were comparisons of machine learning algorithms to other prediction models. Overall, we focused on 89 observational studies in this systematic review for which full texts were available; six publications were not open-access and full texts were obtained from the corresponding authors. For the interested reader, information extracted on RCTs is presented in Appendix Table A.1.

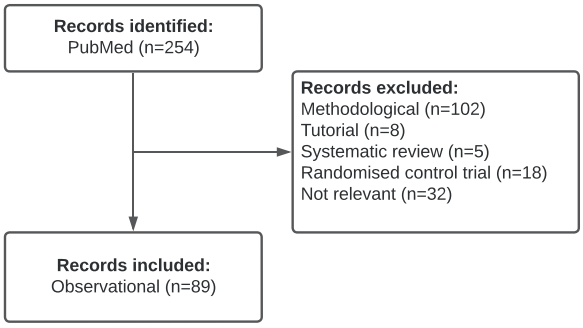

Fig. 1. Flow diagram of studies included in the systematic review.

Dissemination and uptake of the TMLE framework

There has been a growing uptake of the TMLE framework over time, with five or fewer applied publications per year until 2017, and up to 21 in 2021. The majority (66, 74%) of publications using TMLE were published in the last 4 years (2019–2022). Most studies (85, 95.5%) cited the authors of particular TMLE methods they apply, whereas four (4.5%) did not cite any TMLE references. The large majority of these first epidemiological studies benefitted from the expert knowledge of an author who is (or was) part of Professor Mark van der Laan's lab (Table 2).

Of the 89 studies included, two-thirds were conducted in the United States (US) (58, 65.2%, Fig. 2) [32–89], with 100% of articles before 2017 being published in the US, down to 41% of all 2022 articles. Publications from Europe (13, 14.6%) [90–102], Africa (4, 4.5%) [103–106], the Middle East (5, 5.6%) [107–111], and Oceania or Asia (8, 9.0%) [112–119] represent between 25% (in 2017) and 69% (in 2022) of all applied studies published in the last 6 years (Fig. 2, Table 2). In the US, the majority of publications (29) were from California, including 20 from the University of California at Berkeley, where TMLE was first described.

In the early years, the first authors tended to be qualitative academic experts, but we saw more variety in expertize and a larger number of practising clinicians leading observational studies in many epidemiological and public health fields in recent publications. The most common epidemiological subdiscipline was non-communicable diseases (27, 30.3%)[36,47,51,52,54,56,58,67,72,75,80,82,83,85,87,90,95,96,98,100,102,106–110,115], followed by behavioral epidemiology (17, 19.1%) [34,41,48,60,63–65,68,69,73,74,78,79,92,112,117,118], and then infectious disease epidemiology (13, 14.6%)[37,44,66,70,71,94,99,101,104,105,111,116,120]. Through time we see an uptake of TMLE in many more disciplines, such as pharmaco-epidemiology [46,81,97,103], policy [43,45,50,61,76,84,86,119], biomarker epidemiology [32,33,39,40,42,62,113,114], environmental epidemiology [35,49,55,59,77,88,89,91,93], occupational epidemiology [38,53], and health economy [57].

We also studied the evolution of citations. When only methodological overviews of the TMLE framework were available, these were cited despite their heavy statistical requisite. Since 2016, tutorials were published and started to be cited alongside references for statistical packages [1,5–7,121–124] (Table 2).

Of the epidemiological study designs, a cohort study [33,34,37–39,45–47,50,52,53,56,58–61,63,66,69,70,72,75,78–81,83,85,87–89,94,95,98–101,103,105,106,112–119] was the most commonly used design (48, 53.9%), which was followed by cross-sectional (34, 38.2%) (Appendix Table A.2) [32,35,90,36,41–44,91,48,49,51,54,55,57,62,64,65,67,68,71,73,74,76,82,86,92,93,96,97,102,104,107,109]. Other types of commonly used epidemiological study designs included case–control [40,108,110,111,120] and ecological [77,84].

Many articles reported results from other statistical methods, in addition to reporting those obtained from TMLE. Over one-quarter of the studies used adjusted parametric regression (24, 27.0%) [35–37,43,46,51,53,56,58,62,64,70,75,92,93,96–98,105–107,111,115,120],

Table 2
Distribution of observational papers by year of publication and selected characteristics

	Year of publication																										Total
	2009		2010		2011		2012		2013		2014		2015		2016		2017		2018		2019		2020		2021		2022		
	N	(%)	N	(%)	N	(%)	N	(%)	N	(%)	N	(%)	N	(%)	N	(%)	N	(%)	N	(%)	N	(%)	N	(%)	N	(%)	N	(%)	N
Publications	2				1		1		1		2		1		5		4		6		12		16		21		17		89
TMLE expert (author)*	2	(100)			1	(100)	1	(100)	1	(100)	1	(50)	1	(100)	3	(60)	1	(25)	4	(67)	6	(50)	6	(38)	5	(24)	2	(12)	34
USA-based publication	2	(100)			1	(100)	1	(100)			2	(100)	1	(100)	5	(100)	3	(75)	6	(100)	9	(75)	10	(63)	11	(52)	7	(41)	58
Discipline																													
Behavioral epi					1	(100)									1	(20)			1	(17)	2	(17)	6	(38)	4	(19)	2	(12)	17
Biomarker	2	(100)													3	(60)							1	(6)	2	(10)			8
Environmental epi							1	(100)									1	(25)	1	(17)	3	(25)			1	(5)	2	(12)	9
Health economy																					1	(8)							1
Infectious disease											1	(50)					1	(25)			1	(8)	4	(25)	2	(10)	4	(24)	13
Noncommunicable disease									1	(100)	1	(50)							3	(50)	3	(25)	4	(25)	9	(43)	6	(35)	27
Occupational epi													1	(100)							1	(8)							2
Pharmaco-epi																	1	(25)					1	(6)	2	(10)			4
Policy															1	(20)	1	(25)	1	(17)	1	(8)			1	(5)	2	(12)	7
Motivations†																													
Bias					1	(100)	1	(100)	1	(100)	1	(50)			5	(100)	3	(75)	1	(17)	4	(33)	10	(63)	9	(43)	5	(29)	41
Double-robust	2	(100)					1	(100)			1	(50)	1	(100)			1	(25)	2	(33)	4	(33)	5	(31)	5	(24)	5	(29)	27
Efficient											1	(50)	1	(100)	2	(40)	1	(25)			1	(8)	3	(19)	5	(24)			14
Finite sample					1	(100)																	1	(6)			1	(6)	3
Model misspecification																			2	(33)	2	(17)	2	(13)	6	(29)	5	(29)	17
Positivity assumption															1	(20)									2	(10)			3
Power																							1	(6)					1
Time-varying confounding																											2	(12)	2
None specified																	1	(25)	1	(17)	4	(33)	2	(13)	4	(19)	4	(24)	16
Expertize first author																													
Biostatistician	1	(50)									1	(50)	1	(100)			1	(25)	2	(33)	2	(17)	2	(13)	4	(19)	5	(29)	19
Epidemiologist					1	(100)	1	(100)			1	(50)			3	(60)			1	(17)	4	(33)	6	(38)	7	(33)	4	(24)	28
MD															1	(20)	2	(50)			1	(8)	1	(6)	5	(24)	3	(18)	13
MD, MPH																					1	(8)	1	(6)					2
MD, PhD									1	(100)											3	(25)	2	(13)	3	(14)	4	(24)	13
Other																							2	(13)	2	(10)			4
PhD	1	(50)															1	(25)	2	(33)			1	(6)					5
Not known															1	(20)			1	(17)	1	(8)	1	(6)			1	(6)	5
Citations‡																													
Overall TMLE method	2	(100)			1	(50)	1	(100)			5	(56)	1	(33)	4	(33)	4	(50)	3	(33)	7	(21)	13	(41)	20	(38)	11	(26)	72
Specific TMLE method					1	(50)			1	(100)	4	(44)	1	(33)	6	(50)	1	(13)	3	(33)	15	(45)	8	(25)	12	(23)	12	(28)	64
Tutorial																			1	(11)	7	(21)	7	(22)	15	(29)	11	(26)	41
R software													1	(33)	2	(17)	3	(38)	2	(22)	4	(12)	4	(13)	5	(10)	9	(21)	30

* TMLE expert is a current or past member of M.J. van der Laan's Lab.
† Proportions calculated over the number of publications within that year.
‡ Proportions calculated over the total number of citations within that year.

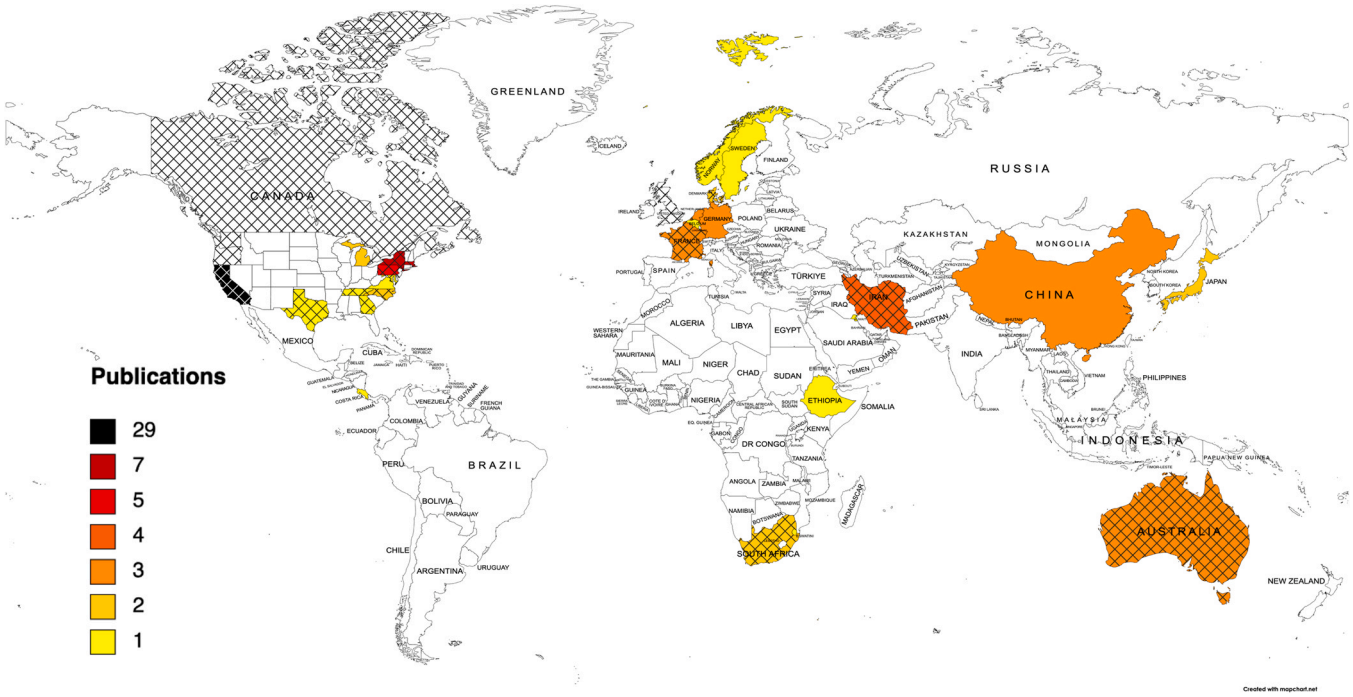

Fig. 2. World map of publications using targeted maximum likelihood estimation by the geographical location of the first author (2006 to mid-2022). Colors represent the number of observational studies and the crosshatch pattern identifies where at least one methodological publication stem from.

one sixth (12, 13.5%) used IPTW [34,37,49,54,63,73,74,77,79,81,103,104], one (1.1%) used AIPTW [76], three (3.4%) used non-parametric methods (e.g., Kaplan Meier) [37,53,75], and seven (7.9%) used unadjusted regression [63,68,72,83,84]. Some studies included more than one comparative method.

The SL package provides a flexible machine learning approach to the estimation of the initial outcome and intervention models (such as the PS). Of the 89 articles, more than half (55, 61.8%) used the SL algorithm [43,45,47,49–52,54,58,60–63,65–67,69–72,74–77,79,81–89,91,92,94,96,97,99–102,104,105,109–120], 18 (20.2%) used logistic regression [32–38,41,46,48,53,55,56,64,73,103,108], and 16 (18.0%) did not specify the approach for the estimation of either the outcome or intervention model [39,40,42,44,57,59,68,78,80,90,93,95,98,106,107]. The average number of machine-learning algorithms considered by the SL was 6.3 (range 1–16), 19 different machine-learning algorithms were used across the articles (a machine-learning algorithm is a wrapper included within the SL [2] library in R software).

The variances (standard errors) of point estimates obtained from TMLE were estimated using differing approaches such as the influence function (n = 21, 23.6%) [35–38,40,43,48,56,71,84,86,90,96,97,101,107–111,115], bootstrap (n = 6, 6.7%) [33,34,46,49,54,73], and Wald tests (n = 2, 2.2%) [41,45], while 60 (67.4%) studies did not specify how standard errors were obtained [32,39,42,44,47,51–53,55,57–70,72,74–83,85,87–89,91–95,98–100,102–106,112–114,116–120,125].

The Causal Inference Roadmap [126] contains seven recommended criteria to define a causal effect: (i) specify the scientific question, (ii) specify the causal model, (iii) define the target causal quantity, (iv) link the observed data to the causal model, (v) assess identifiability assumptions, (vi) estimate the target statistical parameters, and (vii) interpretation of the results. On average, 5.4 (SD 0.9) criteria were complete per article. We considered a version of the Targeted Learning Roadmap [127] that contains five criteria: (i) specify the observed data and describe the data-generating experiment, (ii) specify a statistical model representing a set of realistic assumptions about the underlying true probability distribution of the data, (iii) define a target estimand of the data distribution that "best" approximates the answer to the scientific question of interest, (iv) given statistical model and target estimand, construct an optimal plug-in estimator of the target estimand of the observed data distribution, while respecting the model, and (v) construct a confidence interval by estimating the sampling distribution of the estimator. When the scientific question of interest is causal, step (iii) of the Targeted Learning Roadmap incorporates steps (ii)–(v) of the Causal Inference Roadmap [127]. On average, 3.4 (SD 0.9) criteria were complete per article. Most studies have room to state the necessary content for at least one more criteria.

Most publications (85, 95.5%) used R software to perform TMLE [34–41,43–54,56–69,71,72,74–78,83–92,94–105,107–114,118–120] except four that used STATA [82,93,106,115]. Nonetheless, 10 articles reported using another software tool (i.e., Stata/SAS/SPSS/Python) alongside R for TMLE [34,51,57,65,83,97,100,103,105,110]. The most commonly used R software packages were *tmle* [128] (41, 46.1%) and *ltmle* [129] (18, 20.2%).

Showcase of the TMLE framework

(i) Showcase by motivations

In all disciplines and applications, applying the TMLE framework to their specific research question encouraged authors to review the strengths and limitations of their data and carefully consider how their data and setting might violate identifiability assumptions, which are assumptions necessary for causal inference but not TMLE. If, and only if, the identifiability assumptions are assumed to hold, the estimated effect is a causal effect. However, for observational studies, it cannot be known whether identifiability assumptions hold. Therefore, if an estimate is interpreted as a causal effect, then this interpretation should be accompanied by a discussion of the plausibility of identifiability assumptions. All disciplines and disease areas highlight issues with missing data and measurement errors and incorporate subject-matter knowledge (Appendix Table A.2). We review in turn each characteristic, highlighted by authors, that motivated their use of TMLE.

Three-quarters of the studies (n = 68, 76.4%) provided at least one justification for using TMLE compared to another method (Table 2). The targeting step of the TMLE, aimed to account for any residual confounding due to the model selection, leads to **bias reduction**, that is, an estimated parameter closest to the true value of our quantity of interest. This feature of TMLE was, by far, the most appealing to applied researchers in their observational analyses, with 41 articles (46.1%) mentioning bias reduction [25,34,35,39–44,46,50,54,56,58,64–68,73,74,77,80,90,91,95–97,103,104,106,107,110–112,114,117,119]. **Double-robustness**, meaning that only one of the initial outcome or exposure models needs to be correctly specified, was also a property that attracted authors, cited by 27 articles (30.3%) [32,33,35,38,44,51,52,54,56,60,62,63,67,69,70,73,74,80,85,86,88,89,100,106–108]. **Model misspecification**, which might result from imposing constraints that are unrealistic or not informed by subject-matter knowledge, is reduced in TMLE thanks to machine-learning algorithms used in modeling the outcome and the exposure. Reduced model misspecification was a highly specified driver for using the TMLE framework, cited by 17 articles (19.1%) [48,57,62,72,74,79,81,82,84,88,99–101,109,115,117]. Standard regression techniques in settings with low incidence [44,101], rare outcomes [103], or low sample size [73,92] may over-fit the data or not converge: careful SL specifications overcome these limitations [44,57]. TMLE is also less sensitive than IPW to **positivity violation**, due to the use of machine learning [43,98,109]. **Efficiency**, meaning that fewer observations may be required to achieve a given error performance, is a motivation cited in 14 articles (15.7%) across all disciplines [37–39,41,63,64,77,91,96,97,104,108,110].

(ii)Showcase by disciplines

There was a range of disease areas covered in the 27 **non-communicable disease epidemiology** studies. The appealing property of TMLE was that it is a semiparametric estimator, allowing the use of machine learning algorithms to minimize model misspecification [54,67,80,82,85,98,106,109,115]. Additionally, extensions of TMLE have developed ways to appropriately handle the dual nature of time-varying confounding, which have been utilized in longitudinal studies analyzing data on depression [52], survival from acute respiratory distress syndrome [75], caries arising from intake of soda [56], effects of smoking on rheumatoid arthritis [58], effects of asthma medication on asthma symptoms [95], and reduction of pain after knee replacement surgery [85]. Improved predictive performance [90] and adjusting for informative censoring [75] were additional reasons for using TMLE. Furthermore, the extension of TMLE to case–control studies, in which sampling is biased with respect to the disease status, provided a platform for analyzing the causal effect of reproductive factors on breast cancer by using case–control weighted TMLE [110]. Real-world data overcome limitations of RCTs, such as under-power and evaluation of long-term interventions [85]. Recent applications claimed their use of TMLE in real-world data provided results that were more generalizable than what RCTs would provide [100].

In **infectious disease epidemiology** articles, most were concerned with having a flexible modeling approach that does not impose assumptions on the functional form of the exposure–outcome relationship [66,71,94,104,116,120]. A key feature of the infectious disease epidemiology subdiscipline is that baseline confounders and exposures may change over time and can obscure the causal effect of interest [37]. Standard survival modeling assumes that censoring and survival processes are independent, which is likely violated in this setting, and it assumes there is no time-dependent confounding [37]. TMLEs with a working marginal structural model and for time-to-event outcomes permit evaluation of the effect of an exposure at multiple time points, which is beneficial when the interpretation of causal effects from hazard models is often difficult [130]. Other studies have overcome this issue by using TMLE in a target trial framework or case–cohort studies [70,111].

In **behavioral epidemiology** manuscripts, the behavioral nature of the topics covered implied that RCTs are mostly unethical, bear prohibitive costs or have very small sample sizes. There are several key challenges for using observational data to study the causal effects of childhood adversities [41,48], physical activity [34,63,118], alcohol consumption [65] or supply [112] on various outcomes, including fractures [34], mental health [41,78,79], asthma [92], and pregnancy outcomes [63,64]. They include a risk for reverse causation [75,78,79]; high dimensional data and in particular, multidimensional exposures [41,48]; and measurement error resulting from self-reported exposures or outcomes [64,79,117,118]. Longitudinal relationships and time-varying confounding, where confounders of the effect of an exposure on an outcome can themselves be affected by prior exposures, as well as sample attrition [60,112,117,118], are particular challenges faced by survey data that are collected in consecutive waves [60,74,79,112,117,118]. TMLE adjusts for time-varying confounders affected by prior exposure and employs a doubly robust estimation approach that allows for flexible model fitting. Additionally, as pointed out in 2016 by Ahern et al. [41], “TMLE with machine learning addresses the challenge of a multidimensional exposure because it facilitates ‘learning’ from the data the strength of the relations between each adversity [dimensions of the exposure] and outcome, incorporating any interactions or nonlinearity, specific to each [subgroup].”

The field of **biomarker epidemiology** is driven by the search for sets of candidate biomarkers that are important in determining given outcomes. Ranking the contributions of these candidate biomarkers is also of interest. Some studies used TMLE to measure variable importance in biomarker research [32,40,114] and in other fields [90]. Dimension reduction for the estimation of causal effects is an aim in some biomarker examples [42,62,113]. In the models presented in the publications, there are complex joint effects to consider in large correlated data, as well as longitudinal patterns and time-dependent confounding [42,62,113]. Furthermore, two manuscripts present covariate selection algorithms ahead of causal effect estimation [113,114].

Research published in **environmental epidemiology** highlights challenges around the selection of key variables of interest [86,93], clear definitions of exposure and outcomes [35], as there are likely many proxy and surrogate measures of exposure [91], coupled with potential exposure misclassification and measurement errors [35,55]. Nonetheless, TMLE was successfully applied to determine either causal attributable risk [35,91] or risk differences [49]. Mediation effects were studied in Casey et al. [59] looking at adverse birth outcomes.

The only observational study of TMLE in **health economics** explored the relationship between financial resources leading to food insecurity and healthcare expenditure in a pay-to-access healthcare system. It uses ecological measures of exposure and outcome and leads to evidence for policy [57].

Two publications focused on **occupational epidemiology** [38,53]. A key aspect of occupational epidemiology is accounting for the healthy worker survivor effect: a bias arising due to healthier workers accruing more exposure over time. These studies looked at exposure to particulate matter from aluminum or metalworking fluids in metal factory workers, which varied depending on the length of employment. Both studies benefited from TMLE’s flexibility to allow for time-varying confounding of the exposure.

The field of **pharmacoepidemiology** is concerned with assessing treatment’s efficacy in real-world settings and monitoring long-term side effects of treatments. Both objectives would be either impractical or too costly to study in RCTs, given the limited follow-up time available in clinical trials. TMLE has been used in this setting, as it provides a robust procedure for estimation [46,81,97,103]. In

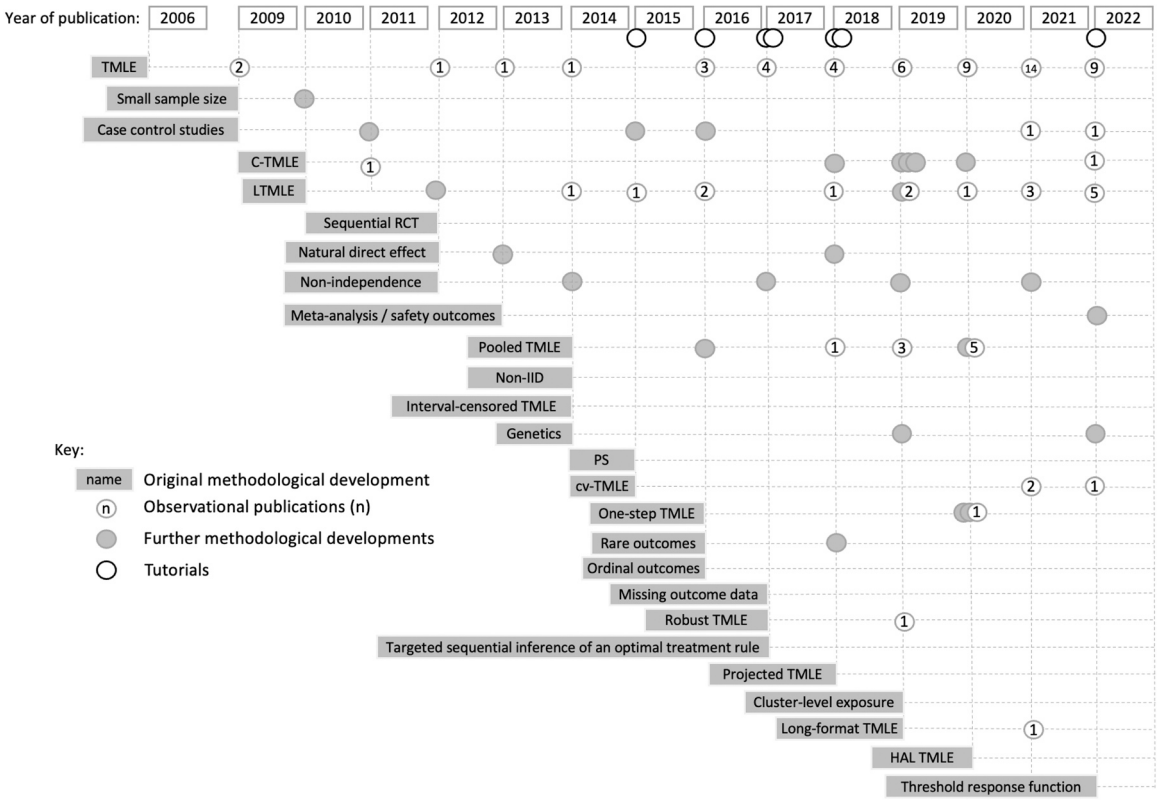

Fig. 3. Applied clinical and epidemiological research by year of publication and TMLE method implemented.

particular, the flexibility of TMLE, provided through the specification of a diverse and rich set of machine learning algorithms in the SL, is crucial for appropriately adjusting for confounding in observational studies [131].

Policy epidemiology assesses the effects of population programs or mandates. Lack of randomization, such as in studies examining the association between specialty probation and public safety outcomes [45,50], leads to an imbalance in the covariate distribution by exposure levels. Studies of cost-effectiveness may involve dealing with outliers which can be addressed with TMLE [50,76]. Other challenges include zero-inflation, such as the assessment of the effect of primary care physician density on arthroplasty outcomes, in which some areas had zero density [76]. This is dealt with by using a mixture of models to assess the probability of non-exposure (i.e., very low density) [76]. Other policy studies presented challenges around missing data [76], reliance on epidemic modeling assumptions [84], target trial emulation [119], dimensionality reduction and spatial associations [86], or infeasible randomization process [61].

Methodological developments and their implementation

Over the years since the TMLE framework was first laid out [3], many contributions have been made to expand the settings in which TMLE is used, provide tools for implementation in standard software, and describe the TMLE framework and application in lay language. Thanks to this, the community of public health researchers and epidemiologists have started implementing the TMLE framework and its latest developments to obtain double robust, least biased and efficient estimates and statistical inference from studies. The properties of TMLE, in contrast to other estimators commonly used for causal inference, include that it is loss-based, well-defined, unbiased, efficient and can be used as a substitution estimator.

Figure 3 shows schematically when and why extensions of TMLE have happened in the last 15 years, as well as extensions and uptake. The 89 applied epidemiological studies are classified by methodological development used during the study. In Appendix Table A.3, the main methodological references are listed and grouped by methodological developments highlighted in Figure 3.

TMLE's superior efficiency and power are evidenced in small sample size settings where marginal effects from logistic regression models adjusted for (possibly many) covariates would not be recommended [132]. The implementation of TMLE in complex causal effect estimation problems is discussed in many publications, such as in settings with multiple time point interventions [133,134], longitudinal data [135,136], postintervention effect modifiers [137], dependence of the treatment assignment between units [138] or censoring [139], causally connected units [140,141], hierarchical data structures [142], randomization at the cluster level [143], large electronic health record data [144], and in meta-analyses [145,146].

The TMLE framework is extended and discussed in the setting of case–control studies. One study matched cases to controls [147], and another used two-stage sampling and nested case–control design [148]. Other studies required the design to be adaptive to possibly invalid assumptions of independent units [149] or if the sample population differs from the (possibly ill-defined) target population [150].

The collaborative TMLE (C-TMLE), introduced in 2010 [151], is an extension of TMLE, in which information on the causal parameter of interest is used when estimating and selecting the initial model(s). C-TMLE aims to improve the robustness and efficiency of the TMLE. Schnitzer et al. [139] highlight the pitfalls and the consequences of automated variable selection in causal inference, such as in the PS model, and how C-TMLE corrects for this. C-TMLE was later extended to measure variable importance [152] and to longitudinal data settings [153]. Proposals to enhance the C-TMLE algorithm include ordering covariates to decrease C-TMLE time complexity [154], using LASSO with C-TMLE for the estimation of the PSs [155], and adaptive truncation of the PSs with C-TMLE to ensure positivity [156].

The pooled TMLE [157] was developed for the context of longitudinal data structures with baseline covariates, time-dependent intervention nodes, intermediate time-dependent covariates, and a possibly time-dependent outcome. Extensions include advice for the optimal discretization of time [158] and to the hazard function [159].

The one-step TMLE aims to preserve the performance of the original two-step TMLE and achieves bias reduction in one step (i.e., without additional iterations of the TMLE update step and possible overfitting in finite samples) [160]. This one-step TMLE was later extended to counterfactual average survival curves [161] and heterogeneous treatment effects [162].

Causal mediation analyses in the nonlongitudinal and longitudinal settings have been developing fast since the end of the 1990s. TMLE was used to propose estimators of natural direct effect [163], or in settings with time-varying mediators and exposures [164], estimates of the complier stochastic direct effect [165], transported interventional effects with multiple, high-dimensional mediators [166] and stochastic (in)direct effects with intermediate confounders [167]. Robust TMLE was proposed in 2017 for transporting intervention effects from one population to another [168].

The cross-validated TMLE (CV-TMLE) provides asymptotic inference under minimal conditions (i.e., nonparametric smoothness [169]) keeping the bounds of the parameter estimates. It is also used in the estimation of data-adaptive target parameters, such as optimal treatment regimes. Recently, TMLE was shown to be useful in defining thresholds and marking specified levels of risks [170].

The set of observational articles that use TMLE in their main or sensitivity analyses shows that TMLE has successfully been used to examine associations [32,34,40–42,44,46,48,49,51,52,54,55,57,59,60,62–64,66,68,71–74,76–83,90,97–99,103,104,107,109,113,116–118], causation [33,35–39,43,45,53,56,58,65,69,70,75,84,91–93,95,96,105,106,108,110–112,114,115,119,125], and variable importance [32,40,90,114]. It has been used to analyze data with varying numbers of observations, from less than 100 to over hundreds of thousands, from clinical trials, cohort studies [33,34,37–39,45,46,52,53,56,58–60,63,66,69,70,72,75,78–81,83,95,98,99,103,105,106,112–119], and observational studies.

Discussion

We aimed to investigate the use of the TMLE framework in epidemiology and public health research and to describe the uptake of its methodological developments since its inception in 2006. We focused on TMLEs for point treatment, time-to-event/survival, and longitudinal exposure-outcome relationships. We found that the TMLE framework and its different estimators were implemented in at least 89 epidemiological observational studies. The majority of these studies have come from the US, many of which are from the University of California, Berkeley. Recently, the use of TMLE has spread across the world. Until 2016, TMLE in observational studies was used by select groups of researchers, such as biostatisticians or epidemiologists in academia exploring noncommunicable and infectious diseases, or behavioral epidemiology. From 2016 onward, there was a faster uptake among a wider range of researchers. There is potential for even wider dissemination and acceptance, both geographically and in some specific disease areas or epidemiological disciplines. From the end of 2022 up to the time of writing, and using the same Boolean search terms for 2023, we found a further 18 observational studies [27,171–187], and 5 methodological studies [188–192] that use some form of TMLE. We hope this review of explicit and applied examples will contribute to enhancing the relevance of the TMLE framework and increasing its uptake and acceptance in settings where challenges with regard to data, unrealistic assumptions, or subject-matter knowledge lend themselves to the framework.

Initially, causal inference methods and estimators relied on parametric modeling assumptions but, to quote Box (1976), “all models are wrong but some are useful” [193]. It highlights that model misspecification was and remains a challenge, even with ever-growing datasets and computing power. Semiparametric and nonparametric estimators, such as AIPTW, double-debiased [194], and TMLE [3] aim to provide the least biased estimate of the effect of an exposure on an outcome [1,195]. Maximum likelihood estimation-based methods (stratification, PS, and parametric regression adjustment) and other estimating equations (AIPTW) do not have all of the properties of TMLE, and evidence shows that they underperform in comparison to TMLE in specific settings [3,148,5,1]. Augmented inverse probability weighting (AIPW) is the closest equivalent methodology to TMLE (e.g., both utilize the efficient influence function, are double robust for the ATE, and are asymptotically unbiased). However, AIPW has different statistical properties. Notably, AIPW/IPW aim to solve an estimating equation, unlike TMLE which uses the log-likelihood as a criterion. As discussed in van der Laan and Rose (2011), estimators based on estimating equations might be nonunique due to the existence of multiple solutions, do not respect known statistical model constraints (i.e., are not substitution estimators), and are sensitive to how the nuisance parameter is estimated [2]. These issues are not present in TMLE and instead TMLE solves the efficient influence curve estimating equation but is not defined by it and is a substitution estimator and thus respects the global constraints of the statistical model [2]. TMLE augments the initial estimates to obtain an optimal bias-variance trade-off for the target estimand of interest and produces a well-defined, unbiased, efficient substitution estimator. Furthermore, the targeting step (i.e., update of the initial estimate) may remove finite sample bias. Lastly, the TMLE framework can be tailored to specific research questions that are difficult to answer using other causal inference methods, such as rare diseases [196,197], ordinal [198] or continuous exposures [199], dynamic treatment regimes [157], and missing outcome data [200]. These are the reasons why we focused on analyses of observational data that used TMLE and did not consider other estimators.

We argue that dissemination of any new statistical methodology relies on five key factors: (i) software availability, (ii) accessibility of available material (e.g., quality of software help files, language used in publications, etc.), (iii) number of experts in the area, (iv) teaching, and (v) collaborations. In the following, we discuss the dissemination of TMLE with regard to each of them.

(i) Software availability

Various TMLEs have been developed for complex study designs, such as those with time-to-event outcomes, case-control studies, hierarchical data structures (including cluster randomized trials), longitudinal data, and time-dependent confounding. These methodological developments were accompanied by the release of R software packages, increasing the usability of TMLE. Such software developments include the *SuperLearner* [4] R package in 2007 and the *tmle* R package in 2012 [3,201]. TMLE software for survival analysis (*survtmle*) [202,203], longitudinal data (*ltmle*) [129,204], double-robust confidence intervals (*drtmle*) [205,206], and estimation of the survival curve under static, dynamic, and stochastic interventions (*stremr*) [207,208] were implemented in 2017. To match the expanding framework, further software developments occurred in the following years, such as the *tlverse* suite of software packages for Targeted Learning (https://tlverse.org/tlverse-handbook/), which includes R packages for cross-validation (*origami*) [209,210], highly adaptive lasso (HAL, *hal9001*) [211–213], super learning (*sl3*) [4,214], and TMLEs for a range of target estimands, such as effects under static interventions on an exposure (*tmle3*) [215], optimal dynamic treatment regimes for binary and categorical exposures (*tmle3mopttx*) [169,216], and stochastic treatment regimes that shift the treatment mechanism of a continuous exposure (*tmle3shift*) [217,218]. Additional

recently developed packages in R include *ctmle* for collaborative TMLE [151,219], *haldensify* for conditional density estimation with HAL [220,221], *txshift* for estimating causal effects of stochastic interventions [222–224], and *lmtp* for longitudinal modified treatment policies [199,225].

Although the TMLE framework is well developed in the R software, applied epidemiological research is performed in several other software languages, such as Stata, Python, and SAS. TMLE implementations for binary point exposure and outcome studies are available in all of these languages. A SAS macro for the general implementation of TMLE was programmed in 2016 [121]. TMLE has been developed for the Python software language in the library *zEpid* [226]. The number of applied researchers in epidemiological studies using Python is relatively low but is increasing; thus, this tool is not currently widely used among applied health sciences researchers. Further development could improve on software packages in the widely used statistical software in health sciences and econometrics, such as Stata [227]. Nonetheless, the development version of the user-written Stata command *eltmle* is currently available to Stata users [227]. Not all features of TMLE are available in this Stata command, such as longitudinal analysis and cross-validated TMLE. Additionally, *eltmle* provides ensemble learning capabilities by accessing the *SuperLearner* R package. Lastly, any new software development needs to have a friendly user interface, together with standard programming features to be easily disseminated and quickly adopted.

(ii) Accessibility of available material

The TMLE framework is a series of potentially statistically-complex modeling approaches and computational algorithms, grounded in statistical theory that requires a solid understanding of highly advanced statistics (i.e., theory for semiparametric estimation, asymptotics, efficiency, empirical processes, functional analyses, and statistical inference). Tutorials in a more lay language targeting applied researchers and epidemiologists have become more common over the past 5 years and the uptake of TMLE is expected to increase in the future because of them [1,5–7,121–124,131,148]. Their beneficial impact is evident from this review, as these articles are highly referenced in applied work, from the year of their publication, alongside more methodologically heavy contributions to start with, and as sole references in later years. This shows evidence of the importance of speaking the language of the target audience and disseminating advanced mathematical statistics and algorithms from an applied perspective.

Additionally, the gradual dissemination of the TMLE framework was evident from our systematic review of the methods sections of the 89 selected manuscripts. We observed that papers published in the early years lay out their TMLE strategy and carefully describe each step in the methods section; whereas, more recently, publications of applied research have placed details of the methods in appendices (or supplementary material) and only cite tutorials and software packages. This shows that the community (e.g., authors, editors, reviewers, readers, etc.) is now aware of the TMLE framework, its utility, and its advantages. A wide range of journals have published the applied research articles studied here, from non-specific public health journals to statistical or disease-specific journals.

(iii) Experts

Dissemination outside the US needs further work, as evidenced in our systematic review. We have shown that the TMLE framework appears to be well consolidated in the US, and adoption from Europe and other regions are lower in comparison. This may be related to the delayed introduction of causal inference education outside the US. Fostering targeted local seminars and dedicated short courses for the interested applied audience could be a useful strategy to disseminate the framework. Disease- or discipline-specific experts would be useful for the wider distribution of the methods in specific areas that would benefit from improved methodology.

TMLE remains dominant in noncommunicable or infectious disease epidemiology compared to other disciplines, but it has high applicability in many disciplines and its use has increased in several of them. The slower uptake of the TMLE framework in other disciplines might be due to a lack of empirical examples of how one performed and interpreted statistical analyses using TMLE in a specific disease area. We aimed to provide such a showcase of the application of the methods in specific settings, based on the available literature, and we demonstrated how the framework was successfully used to advance research by providing robust results. We believe interested readers will find it useful to refer to the studies that faced similar challenges, or were based in settings comparable to theirs.

(iv) Teaching

There have been tremendous efforts of dissemination of causal inference methods across disciplines, with a particular emphasis on epidemiology and econometrics sciences in the US during the last 20 years. Most graduate programs in epidemiology have included the teaching of causal inference as a leading topic in the field. In Europe, the trends have not been as fast-paced and for a long time, introductions to causal inference methods have mainly been provided through week-long intensive short courses and at international conferences. These different approaches have major impacts on how quickly the methods are adopted by the community of researchers, journal editors, public health groups, and regulatory agencies. In recent years, there has been a development and acceptance of real-world evidence in various public-health fields, such as the Food and Drug Administration's 21st Century Cures Act of 2016 in the US, which specifically promotes the use of causal inference methodology and designs, such as the emulated trial and TMLE frameworks [228–230].

(v) Collaborations

The Center for Targeted Machine Learning and Causal Inference (CTML) is an interdisciplinary research center at the University of California at Berkeley that is focused on applications of causal inference and targeted learning. The CTML mission is to advance, implement, and disseminate methodology to address problems arising in public health and clinical medicine (https://ctml.berkeley.edu/home). CTML provides a great resource for courses, ongoing research, partners, collaborators, and Berkeley faculty members involved in TMLE. CTML sponsors include the Danish multinational pharmaceutical company, Novo Nordisk A/S, the Patient-Centered Outcomes Research Institute (pcori), Kaiser Permanente, the US National Institutes of Health, and the Bill and Melinda Gates Foundation. Academic partners include the University of Washington, University of Copenhagen, UCLA David Geffen School of Medicine, University of California at San Francisco, and Monash University.

Conclusions

Evidence shows that cross-validated, double-robust, efficient, and unbiased estimators are at the forefront of causal inference and statistics, as they aim to avoid model misspecification, bias, and invalid inference. The TMLE framework for causal and statistical inference was first developed in 2006 and its expansion in applied studies arose in 2018 via applied epidemiological work, tutorials, and user-friendly software. The theoretical properties and practical

benefits of the TMLE framework have been highlighted across different fields of applied research (such as various epidemiological, public health, and clinical disciplines). More can be done to reach a wider audience across varied disease areas and scientific fields (e.g., genomics, econometrics, political, and sociological sciences), including the development of software packages outside the R software, tutorial articles as well as seminars and courses targeted to audiences in specific disciplines, lay-language demonstration, such as by example, of the benefits of TMLE in improving epidemiological output, to name only a few ideas. Many recent TMLE developments answer a variety of methodological problems that expand across scientific disciplines and further efforts can be made to disseminate the framework. This would facilitate the conscientious application of TMLE for causal inference and statistical data analyses, so more researchers could use it in their applied work to minimize the risk of reporting misleading results that are biased due to misspecification.

Authors contributions

The article arose from the motivation to disseminate the principles of modern epidemiology among clinicians and applied researchers. All authors developed the concept and wrote the first draft of the article. MJS and CM reviewed the literature. MJS, RVP, MALF, and CM drafted and revised the manuscript. RVP provided comments on the draft manuscript. RVP contributed to drafting some sections. All authors read and approved the final version of the manuscript. CM is the guarantor of the article.

Funding

This work was supported by the Medical Research Council [grant number MR/W021021/1]. A CC BY or equivalent license is applied to the Author Accepted Manuscript (AAM) arising from this submission, in accordance with the grant's open access conditions. Camille Maringe is supported by a Cancer Research UK Population Research Committee Program Award (C7923/A29018).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The motivation and some parts of the manuscript come from MALF's work in a visiting academic position in the Division of Biostatistics at the Berkeley School of Public Health in 2019. We acknowledge Dr Susan Gruber for her careful read over one draft manuscript and Professor Mark J. van der Laan for his additional guidance.

References

[1] Smith MJ, Mansournia MA, Maringe C, Zivich PN, Cole SR, Leyrat C. Introduction to computational causal inference using reproducible Stata, R, and Python code: a tutorial. Stat Med 2022;41(2):407–32. https://doi.org/10.1002/sim.9234
[2] van der Laan MJ, Rose S. Targeted learning. Springer Series in Statistics. New York: Springer; 2011. https://doi.org/10.1007/978-1-4419-9782-1
[3] van der Laan MJ, Rubin D. Targeted maximum likelihood learning. Int J Biostat 2006;2(1):1–38. https://doi.org/10.2202/1557-4679.1043
[4] van der Laan MJ, Polley EC, Hubbard AE. Super learner. Stat Appl Genet Mol Biol 2007;6(1). https://doi.org/10.2202/1544-6115.1309
[5] Luque-Fernandez MA, Schomaker M, Rachet B, Schnitzer ME. Targeted maximum likelihood estimation for a binary treatment: a tutorial. Stat Med 2018;37(16):2530–46. https://doi.org/10.1002/sim.7628
[6] Gruber S. Targeted learning in healthcare research. Big Data 2015;3(4):211–8. https://doi.org/10.1089/big.2015.0025
[7] Schuler MS, Rose S. Targeted maximum likelihood estimation for causal inference in observational studies. Am J Epidemiol 2017;185(1):65–73. https://doi.org/10.1093/aje/kww165
[8] van der Laan MJ, Rose S. Targeted learning in data science: causal inference for complex longitudinal studies. 2011. doi:10.1007/978-3-319-65304-4.
[9] UC Berkeley. American Causal Inference Conference. ⟨https://ctml.berkeley.edu/american-causal-inference-conference-2022⟩; 2022. (accessed Jan 1, 2023).
[10] UC Berkeley. American Causal Inference Conference. Results. ⟨https://acic2022.mathematica.org/results⟩; 2022. (accessed Jan 1, 2023).
[11] Arnold B, Arana B, Mäusezahl D, Hubbard A, Colford J. Evaluation of a pre-existing, 3-year household water treatment and handwashing intervention in rural Guatemala. Int J Epidemiol 2009;38(6):1651–61. https://doi.org/10.1093/ije/dyp241
[12] Moore KL, Neugebauer R, Valappil T, Laan MJ. Robust extraction of covariate information to improve estimation efficiency in randomized trials. Stat Med 2011;30(19):2389–408. https://doi.org/10.1002/sim.4301
[13] Hubbard A, Jamshidian F, Jewell N. Adjusting for perception and unmasking effects in longitudinal clinical trials. Int J Biostat 2012;8(2):7. https://doi.org/10.2202/1557-4679.1376
[14] Wester CW, Stitelman OM, deGruttola V, Bussmann H, Marlink RG, van der Laan MJ. Effect modification by sex and baseline CD4+ cell count among adults receiving combination antiretroviral therapy in Botswana: results from a clinical trial. AIDS Res Hum Retroviruses 2012;28(9):981–8. https://doi.org/10.1089/aid.2011.0349
[15] Decker AL, Hubbard A, Crespi CM, Seto EY, Wang MC. Semiparametric estimation of the impacts of longitudinal interventions on adolescent obesity using targeted maximum-likelihood: accessible estimation with the ltmle package. J Causal Inference 2014;2(1):95–108. https://doi.org/10.1515/jci-2013-0025
[16] Balzer L, Staples P, Onnela JP, DeGruttola V. Using a network-based approach and targeted maximum likelihood estimation to evaluate the effect of adding pre-exposure prophylaxis to an ongoing test-and-treat trial. Clin Trials 2017;14(2):201–10. https://doi.org/10.1177/1740774516679666
[17] BL P, PB G, MJ t. Estimation of the optimal surrogate based on a randomized trial. Biometrics 2018;74(4):1271–81.
[18] Sridhar S, Luedtke A, Langevin E, Zhu M, Bonaparte M, Machabert T, et al. Effect of dengue serostatus on dengue vaccine safety and efficacy. N Engl J Med 2018;379(4):327–40. https://doi.org/10.1056/NEJMoa1800820
[19] Havlir DV, Balzer LB, Charlebois ED, Clark TD, Kwarisiima D, Ayieko J, et al. HIV testing and treatment with the use of a community health approach in rural Africa. N Engl J Med 2019;381(3):219. https://doi.org/10.1056/NEJMOA1809866
[20] Nguyen M, Pirracchio R, Kornblith LZ, Callcut R, Fox EE, Wade CE, et al. Dynamic impact of transfusion ratios on outcomes in severely injured patients: targeted machine learning analysis of the PROPPR randomized clinical trial. J Trauma Acute Care Surg 2020;89(3):505. https://doi.org/10.1097/TA.0000000000002819
[21] Dayan GH, Langevin E, Gilbert PB, Wu Y, Moodie Z, Forrat R, et al. Assessment of the long-term efficacy of a dengue vaccine against symptomatic, virologically-confirmed dengue disease by baseline dengue serostatus. Vaccine 2020;38(19):3531–6. https://doi.org/10.1016/j.vaccine.2020.03.029
[22] Potter K, Masteller B, Balzer LB. Examining obedience training as a physical activity intervention for dog owners: findings from the stealth pet obedience training (SPOT) pilot study. Int J Environ Res Public Health 2021;18(3). https://doi.org/10.3390/ijerph18030902
[23] Nyabuti MN, Petersen ML, Bukusi EA, Kamya MR, Mwangwa F, Kabami J, et al. Characteristics of HIV seroconverters in the setting of universal test and treat: Results from the SEARCH trial in rural Uganda and Kenya. PLoS One 2021;16(2):e0243167. https://doi.org/10.1371/journal.pone.0243167
[24] Hickey MD, Ayieko J, Owaraganise A, Sim N, Balzer LB, Kabami J, et al. Effect of a patient-centered hypertension delivery strategy on all-cause mortality: secondary analysis of SEARCH, a community-randomized trial in rural Kenya and Uganda. PLoS Med 2021;18(9):e1003803. https://doi.org/10.1371/journal.pmed.1003803
[25] Amato HK, Hemlock C, Andrejko KL, Smith AR, Hejazi NS, Hubbard AE, et al. Biodigester Cookstove interventions and child diarrhea in Semirural Nepal: a causal analysis of daily observations. Environ Health Perspect 2022;130(1):17002. https://doi.org/10.1289/ehp9468
[26] Hickey MD, Owaraganise A, Sang N, Opel FJ, Mugoma EW, Ayieko J, et al. Effect of a one-time financial incentive on linkage to chronic hypertension care in Kenya and Uganda: A randomized controlled trial. PLOS One 2022;17(11). https://doi.org/10.1371/JOURNAL.PONE.0277312
[27] Marquez C, Chen Y, Atukunda M, Chamie G, Balzer LB, Kironde J, et al. The association between social network characteristics and tuberculosis infection among adults in 9 rural Ugandan communities. Clin Infect Dis 2023;76(3):e902. https://doi.org/10.1093/CID/CIAC669
[28] Kahler CW, Janssen T, Gruber S, Howe CJ, Laws BM, Walthers J, et al. Change talk subtypes as predictors of alcohol use following brief motivational intervention. Psychology of addictive behaviors: journal of the Society of

Psychologists in. Psychol Addict Behav 2022. https://doi.org/10.1037/ADB0000898

[29] Luque-Fernandez MA, Belot A, Valeri L, Cerulli G, Maringe C, Rachet B. Data-adaptive estimation for double-robust methods in population-based cancer epidemiology: risk differences for lung cancer mortality by emergency presentation. Am J Epidemiol 2018;187(4):871–8. https://doi.org/10.1093/AJE/KWX317

[30] Nebel MB, Lidstone DE, Wang L, Benkeser D, Mostofsky SH, Risk BB. Accounting for motion in resting-state fMRI: What part of the spectrum are we characterizing in autism spectrum disorder? NeuroImage 2022;257:119296. https://doi.org/10.1016/J.NEUROIMAGE.2022.119296

[31] Sciannameo V, Fadini GP, Bottigliengo D, Avogaro A, Baldi I, Gregori D, et al. Assessment of glucose lowering medications' effectiveness for cardiovascular clinical risk management of real-world patients with type 2 diabetes: targeted maximum likelihood estimation under model misspecification and missing outcomes. Int J Environ Res Public Health 2022;19(22):14825. https://doi.org/10.3390/IJERPH192214825/S1

[32] Bembom O, Petersen ML, Rhee SY, Fessel WJ, Sinisi SE, Shafar RW, et al. Biomarker discovery using targeted maximum-likelihood estimation: application to the treatment of antiretroviral-resistant HIV infection. Stat Med 2009;28(1):152–72. https://doi.org/10.1002/sim.3414

[33] Rosenblum M, Deeks SG, van der Laan MJ, Bangsberg DR. The risk of virologic failure decreases with duration of HIV suppression, at greater than 50% adherence to antiretroviral therapy. PLoS One 2009;4(9):e7196. https://doi.org/10.1371/journal.pone.0007196

[34] Mackey DC, Hubbard AE, Cawthon PM, Cauley JA, Cummings SR, Tager IB. Usual physical activity and hip fracture in older men: an application of semiparametric methods to observational data. Am J Epidemiol 2011;173(5):578–86. https://doi.org/10.1093/aje/kwq405

[35] Padula AM, Mortimer K, Hubbard A, Lurmann F, Jerrett M, Tager IB. Exposure to traffic-related air pollution during pregnancy and term low birth weight: estimation of causal associations in a semiparametric model. Am J Epidemiol 2012;176(9):815–24. https://doi.org/10.1093/aje/kws148

[36] Leslie HH, Karasek DA, Harris LF, Chang E, Abdulrahim N, Maloba M, et al. Cervical cancer precursors and hormonal contraceptive use in HIV-positive women: application of a causal model and semi-parametric estimation methods. PLoS One 2014;9(6):e101090. https://doi.org/10.1371/journal.pone.0101090

[37] Schnitzer ME, Moodie EE, van der Laan MJ, Platt RW, Klein MB. Modeling the impact of hepatitis C viral clearance on end-stage liver disease in an HIV co-infected cohort with targeted maximum likelihood estimation. Biometrics 2014;70(1):144–52. https://doi.org/10.1111/biom.12105

[38] Brown DM, Petersen M, Costello S, Noth EM, Hammond K, Cullen M, et al. Occupational exposure to PM2.5 and incidence of ischemic heart disease: longitudinal targeted minimum loss-based estimation. Epidemiology 2015;26(6):806–14. https://doi.org/10.1097/ede.0000000000000329

[39] Gianfrancesco MA, Balzer L, Taylor KE, Trupin L, Nititham J, Seldin MF, et al. Genetic risk and longitudinal disease activity in systemic lupus erythematosus using targeted maximum likelihood estimation. Genes Immun 2016;17(6):358–62. https://doi.org/10.1038/gene.2016.33

[40] Hsu LI, Briggs F, Shao X, Metayer C, Wiemels JL, Chkalingam AP, et al. Pathway analysis of genome-wide association study in childhood leukemia among Hispanics. Cancer Epidemiol Biomarkers Prev 2016;25(5):815–22. https://doi.org/10.1158/1055-9965.Epi-15-0528

[41] Ahern J, Karasek D, Luedtke AR, Bruckner TA, van der Laan MJ. Racial/ethnic differences in the role of childhood adversities for mental disorders among a nationally representative sample of adolescents. Epidemiology 2016;27(5):697–704. https://doi.org/10.1097/ede.0000000000000507

[42] Salihu HM, Das R, Morton L, Huang H, Paothong A, Wilson RE, et al. Racial differences in DNA-methylation of CpG sites within preterm-promoting genes and gene variants. Matern Child Health J 2016;20(8):1680–7. https://doi.org/10.1007/s10995-016-1967-3

[43] Tran L, Yiannoutsos CT, Musick BS, Wools-Kaloustian KK, Siika A, Kimaiyo S, et al. Evaluating the Impact of a HIV low-risk express care task-shifting program: a case study of the targeted learning roadmap. Epidemiol Methods 2016;5(1):69–91. https://doi.org/10.1515/em-2016-0004

[44] Davis FM, Sutzko DC, Grey SF, Mansour MA, Jain KM, Nypaver TJ, et al. Predictors of surgical site infection after open lower extremity revascularization. J Vasc Surg 2017;65(6):1769–1778.e3. https://doi.org/10.1016/j.jvs.2016.11.053

[45] Skeem JL, Manchak S, Montoya L. Comparing public safety outcomes for traditional probation vs specialty mental health probation. JAMA Psychiatry 2017;74(9):942–8. https://doi.org/10.1001/jamapsychiatry.2017.1384

[46] Sukul D, Seth M, Dixon SR, Khandelwal A, LaLonde TA, Gurm HS. Contemporary trends and outcomes associated with the preprocedural use of oral p2y12 inhibitors in patients undergoing percutaneous coronary intervention: insights from the Blue Cross Blue Shield of Michigan Cardiovascular Consortium (BMC2). J Invasive Cardiol 2017;29(10):340–51.

[47] Akosile M, Zhu H, Zhang S, Johnson NP, Lai D, Zhu H. Reassessing the effectiveness of right heart catheterization (RHC) in the initial care of critically Ill patients using targeted maximum likelihood estimation. Int J Clin Biostat Biomet 2018;4(1). https://doi.org/10.23937/2469-5831/1510018

[48] Platt JM, McLaughlin KA, Luedtke AR, Ahern J, Kaufman AS, Keyes KM. Targeted estimation of the relationship between childhood adversity and fluid intelligence in a US population sample of adolescents. Am J Epidemiol 2018;187(7):1456–66. https://doi.org/10.1093/aje/kwy006

[49] Pearl M, Ahern J, Hubbard A, Laraia B, Shrimali BP, Poon V, et al. Life-course neighbourhood opportunity and racial-ethnic disparities in risk of preterm birth. Paediatr Perinat Epidemiol 2018;32(5):412–9. https://doi.org/10.1111/ppe.12482

[50] Skeem JL, Montoya L, Manchak SM. Comparing costs of traditional and specialty probation for people with serious mental illness. Psychiatr Serv 2018;69(8):896–902. https://doi.org/10.1176/appi.ps.201700498

[51] Mosconi L, Rahman A, Diaz I, Wu X, Scheyer O, Hristov HW, et al. Increased Alzheimer's risk during the menopause transition: a 3-year longitudinal brain imaging study. PLoS One 2018;13(12):e0207885. https://doi.org/10.1371/journal.pone.0207885

[52] Torres JM, Rudolph KE, Sofrygin O, Glymour MM, Wong R. Longitudinal associations between having an adult child migrant and depressive symptoms among older adults in the Mexican Health and Aging Study. Int J Epidemiol 2018;47(5):1432–42. https://doi.org/10.1093/ije/dyy112

[53] Izano MA, Sofrygin OA, Picciotto S, Bradshaw PT, Eisen EA. Metalworking fluids and colon cancer risk: longitudinal targeted minimum loss-based estimation. Environ Epidemiol 2019;3(1):e035. https://doi.org/10.1097/ee9.0000000000000035

[54] Yu YH, Bodnar LM, Brooks MM, Himes KP, Naimi AI. Comparison of parametric and nonparametric estimators for the association between incident prepregnancy obesity and stillbirth in a population-based cohort study. Am J Epidemiol 2019;188(7):1328–36. https://doi.org/10.1093/aje/kwz081

[55] Rudolph KE, Shev A, Paksarian D, Merikangas KR, Mennitt DJ, James P, et al. Environmental noise and sleep and mental health outcomes in a nationally representative sample of urban US adolescents. Environ Epidemiol 2019;3(4):e056. https://doi.org/10.1097/ee9.0000000000000056

[56] Lim S, Tellez M, Ismail AI. Estimating a dynamic effect of soda intake on pediatric dental caries using targeted maximum likelihood estimation method. Caries Res 2019;53(5):532–40. https://doi.org/10.1159/000497359

[57] Berkowitz SA, Basu S, Gundersen C, Seligman HK. State-level and county-level estimates of health care costs associated with food insecurity. Prev Chronic Dis 2019;16:E90. https://doi.org/10.5888/pcd16.180549

[58] Gianfrancesco MA, Trupin L, Shiboski S, van der Laan MJ, Graf J, Imboden J, et al. Smoking is associated with higher disease activity in rheumatoid arthritis: a longitudinal study controlling for time-varying covariates. J Rheumatol 2019;46(4):370–5. https://doi.org/10.3899/jrheum.180262

[59] Casey JA, Goin DE, Rudolph KE, Schwartz BS, Mercer D, Elser H, et al. Unconventional natural gas development and adverse birth outcomes in Pennsylvania: the potential mediating role of antenatal anxiety and depression. Environ Res 2019;177:108598. https://doi.org/10.1016/j.envres.2019.108598

[60] Torres JM, Rudolph KE, Sofrygin O, Wong R, Walter LC, Glymour MM. Having an adult child in the United States, physical functioning, and unmet needs for care among older Mexican adults. Epidemiology 2019;30(4):553–60. https://doi.org/10.1097/ede.0000000000001016

[61] You Y, Doubova SV, Pinto-Masis D, Pérez-Cuevas R, Borja-Aburto VH, Hubbard A. Application of machine learning methodology to assess the performance of DIABETIMSS program for patients with type 2 diabetes in family medicine clinics in Mexico. BMC Med Inform Decis Mak 2019;19(1). https://doi.org/10.1186/S12911-019-0950-5

[62] Izano MA, Cushing LJ, Lin J, Eick SM, Goin DE, Epel E, et al. The association of maternal psychosocial stress with newborn telomere length. PLoS One 2020;15(12):e0242064. https://doi.org/10.1371/journal.pone.0242064

[63] Ehrlich SF, Neugebauer RS, Feng J, Hedderson MM, Ferrara A. Exercise during the first trimester and infant size at birth: targeted maximum likelihood estimation of the causal risk difference. Am J Epidemiol 2020;189(2):133–45. https://doi.org/10.1093/aje/kwz213

[64] Bodnar LM, Cartus AR, Kirkpatrick SI, Himes KP, Kennedy EH, Simhan HN, et al. Machine learning as a strategy to account for dietary synergy: an illustration based on dietary intake and adverse pregnancy outcomes. Am J Clin Nutr 2020;111(6):1235–43. https://doi.org/10.1093/ajcn/nqaa027

[65] Kagawa RMC, Pear VA, Rudolph KE, Keyes KM, Cerdá M, Wintemute GJ. Distress level and daily functioning problems attributed to firearm victimization: sociodemographic-specific responses. Ann Epidemiol 2020;41:35–42.e3. https://doi.org/10.1016/j.annepidem.2019.12.002

[66] Kempker RR, Mikiashvili L, Zhao Y, Benkeser D, Barbakadze K, Bablishvili N, et al. Clinical outcomes among patients with drug-resistant tuberculosis receiving bedaquiline- or delamanid-containing regimens. Clin Infect Dis 2020;71(9):2336–44. https://doi.org/10.1093/cid/ciz1107

[67] Yu YH, Bodnar LM, Himes KP, Brooks MM, Naimi AI. Association of overweight and obesity development between pregnancies with stillbirth and infant mortality in a cohort of multiparous women. Obstet Gynecol 2020;135(3):634–43. https://doi.org/10.1097/aog.0000000000003677

[68] Puryear SB, Balzer LB, Ayieko J, Kwarisiima D, Hahn JA, Charlebois ED, et al. Associations between alcohol use and HIV care cascade outcomes among adults undergoing population-based HIV testing in East Africa. Aids 2020;34(3):405–13. https://doi.org/10.1097/qad.0000000000002427

[69] Torres JM, Sofrygin O, Rudolph KE, Haan MN, Wong R, Glymour MM. US migration status of adult children and cognitive decline among older parents who remain in Mexico. Am J Epidemiol 2020;189(8):761–9. https://doi.org/10.1093/aje/kwz277

[70] Westling T, Cowden C, Mwananyanda L, Kapasa ML, Machona S, Pierre C, et al. Impact of chlorhexidine baths on suspected sepsis and bloodstream infections in hospitalized neonates in Zambia. Int J Infect Dis 2020;96:54–60. https://doi.org/10.1016/j.ijid.2020.03.043

[71] Aiemjoy K, Aragie S, Wittberg DM, Tadesse Z, Callahan EK, Gwyn S, et al. Seroprevalence of antibodies against *Chlamydia trachomatis* and enteropathogens and distance to the nearest water source among young children in the Amhara Region of Ethiopia. PLoS Negl Trop Dis 2020;14(9):e0008647. https://doi.org/10.1371/journal.pntd.0008647
[72] Reilly ME, Conti MS, Day J, MacMahon A, Chrea B, Caolo KC, et al. Modified lapidus vs scarf osteotomy outcomes for treatment of hallux valgus deformity. Foot Ankle Int 2021;42(11):1454–62. https://doi.org/10.1177/10711007211013776
[73] Kang L, Vij A, Hubbard A, Shaw D. The unintended impact of helmet use on bicyclists' risk-taking behaviors. J Safety Res 2021;79:135–47. https://doi.org/10.1016/j.jsr.2021.08.014
[74] Torres JM, Mitchell UA, Sofrygin O, Rudolph KE, Lopez-Ortega M, Sharif MZ, et al. Associations between spousal caregiving and health among older adults in Mexico: a targeted estimation approach. Int J Geriatr Psychiatry 2021;36(5):775–83. https://doi.org/10.1002/gps.5477
[75] Torres LK, Hoffman KL, Oromendia C, Diaz I, Harrington JS, Schenck EJ, et al. Attributable mortality of acute respiratory distress syndrome: a systematic review, meta-analysis and survival analysis using targeted minimum loss-based estimation. Thorax 2021;76(12):1176–85. https://doi.org/10.1136/thoraxjnl-2020-215950
[76] Mehta B, Brantner C, Williams N, Szymonifka J, Navarro-Millan I, Mandl LA, et al. Primary care provider density and elective total joint replacement outcomes. Arthroplast Today 2021;10:73–8. https://doi.org/10.1016/j.artd.2021.05.010
[77] Nardone A, Rudolph KE, Morello-Frosch R, Casey JA. Redlines and greenspace: the relationship between historical redlining and 2010 greenspace across the United States. Environ Health Perspect 2021;129(1):17006. https://doi.org/10.1289/ehp7495
[78] Lee JO, Kapteyn A, Clomax A, Jin H. Estimating influences of unemployment and underemployment on mental health during the COVID-19 pandemic: who suffers the most? Public Health 2021;201:48–54. https://doi.org/10.1016/j.puhe.2021.09.038
[79] Shiba K, Torres JM, Daoud A, Inoue K, Kanamori S, Tsuki T, et al. Estimating the impact of sustained social participation on depressive symptoms in older adults. Epidemiology 2021;32(6):886–95. https://doi.org/10.1097/ede.0000000000001395
[80] Goel AR, Bruce HA, Williams N, Alexiades G. Long-term effects of hearing aids on hearing ability in patients with sensorineural hearing loss. J Am Acad Audiol 2021;32(6):374–8. https://doi.org/10.1055/s-0041-1731592
[81] Kahkoska AR, Abrahamsen TJ, Alexander GC, Bennett TD, Chute CG, Haendel MA, et al. Association between glucagon-like peptide 1 receptor agonist and sodium-glucose cotransporter 2 inhibitor use and COVID-19 outcomes. Diabetes Care 2021;44(7):1564–72. https://doi.org/10.2337/dc21-0065
[82] Beydoun HA, Beydoun MA, Huang S, Eid SM, Zonderman AB. Hospitalization outcomes among brain metastasis patients receiving radiation therapy with or without stereotactic radiosurgery from the 2005-2014 Nationwide Inpatient Sample. Sci Rep 2021;11(1):19209. https://doi.org/10.1038/s41598-021-98563-y
[83] Crowner JR, Marston WA, Freeman NLB, Sperger J, Austin HD, Steffan M, et al. The Society for Vascular Surgery Objective Performance Goals for Critical Limb Ischemia are attainable in select patients with ischemic wounds managed with wound care alone. Ann Vasc Surg 2022;78:28–35. https://doi.org/10.1016/j.avsg.2021.06.034
[84] Wong AK, Balzer LB. State-level masking mandates and COVID-19 outcomes in the United States: a demonstration of the causal roadmap. Epidemiology 2022;33(2):228–36. https://doi.org/10.1097/ede.0000000000001453
[85] Jafarzadeh SR, Neogi T, White DK, Felson DT. The relationship of pain reduction with prevention of knee replacement under dynamic intervention strategies. Arthritis Rheumatol 2022;74(10):1668–75. https://doi.org/10.1002/ART.42272
[86] Kagawa R, Calnin B, Smirniotis C, Cerdá M, Wintemute G, Rudolph KE. Effects of building demolitions on firearm violence in Detroit, Michigan. Prev Med 2022;165:107257. https://doi.org/10.1016/J.YPMED.2022.107257
[87] Rajan L, Kim J, An T, Fuller R, Zhu J, Pinney SJ, et al. The influence of concomitant hammertoe correction on postoperative outcomes in patients undergoing hallux valgus correction. Foot Ankle Surg 2022;28(7):1100–5. https://doi.org/10.1016/J.FAS.2022.03.008
[88] Shiba K, Daoud A, Hikichi H, Yazawa A, Aida J, Kondo K, et al. Uncovering heterogeneous associations between disaster-related trauma and subsequent functional limitations: a machine-learning approach. Am J Epidemiol 2023;192(2):217–29. https://doi.org/10.1093/AJE/KWAC187
[89] Shiba K, Hikichi H, Okuzono SS, Vanderweele TJ, Arcaya M, Daoud A, et al. Long-term associations between disaster-related home loss and health and well-being of older survivors: nine years after the 2011 great east Japan earthquake and tsunami. Environ Health Persp 2022;130(7). https://doi.org/10.1289/EHP10903
[90] Legrand M, Pirracchio R, Rosa A, Peterson ML, van der Laan MJ, Fabiani JN, et al. Incidence, risk factors and prediction of post-operative acute kidney injury following cardiac surgery for active infective endocarditis: an observational study. Crit Care 2013;17(5):R220. https://doi.org/10.1186/cc13041
[91] Herrera R, Berger U, Ehrenstein voS, Diaz I, Huber S, Moraga Munoz D, et al. Estimating the causal impact of proximity to gold and copper mines on respiratory diseases in Chilean children: an application of targeted maximum likelihood estimation. Int J Environ Res Public Health 2017;15(1). https://doi.org/10.3390/ijerph15010039
[92] Rodríguez-Molina D, Barth S, Herrera R, Rossmann C, Radon K, Karnowski V. An educational intervention to improve knowledge about prevention against occupational asthma and allergies using targeted maximum likelihood estimation. Int Arch Occup Environ Health 2019;92(5):629–38. https://doi.org/10.1007/s00420-018-1397-1
[93] Papadopoulou E, Haug LS, Sakhi AK, Andrusaityte S, Basagana X, Brantsaeter AL, et al. Diet as a source of exposure to environmental contaminants for pregnant women and children from six European countries. Environ Health Perspect 2019;127(10):107005. https://doi.org/10.1289/ehp5324
[94] Vauchel T, Pirracchio R, Chaussard M, Lafaurie M, Rouveau M, Rousseau C, et al. Impact of an *Acinetobacter baumannii* outbreak on kidney events in a burn unit: a targeted machine learning analysis. Am J Infect Control 2019;47(4):435–8. https://doi.org/10.1016/j.ajic.2018.09.010
[95] Veit C, Herrera R, Weinmayr G, Genuneit J, Windstetter D, Vogelberg C, et al. Long-term effects of asthma medication on asthma symptoms: an application of the targeted maximum likelihood estimation. BMC Med Res Methodol 2020;20(1):307. https://doi.org/10.1186/s12874-020-01175-9
[96] Decruyenaere A, Steen J, Colpaert K, Benoit DD, Decruyenaere J, Vansteelandt S. The obesity paradox in critically ill patients: a causal learning approach to a casual finding. Crit Care 2020;24(1):485. https://doi.org/10.1186/s13054-020-03199-5
[97] Rossides M, Kullberg S, DiGiuseppe D, Eklund A, Grunewald J, Askling J, et al. Infection risk in sarcoidosis patients treated with methotrexate compared to azathioprine: a retrospective 'target trial' emulated with Swedish real-world data. Respirology 2021;26(5):452–60. https://doi.org/10.1111/resp.14001
[98] Zou R, ElMarroun H, Cecil C, Jaddoe VWV, Hillegers M, Tiemeier H, et al. Maternal folate levels during pregnancy and offspring brain development in late childhood. Clin Nutr 2021;40(5):3391–400. https://doi.org/10.1016/j.clnu.2020.11.025
[99] Isfordink CJ, Smit C, Boyd A, de Regt MJA, Rijnders BJA, Van Crevel R, et al. Low hepatitis C virus-viremia prevalence yet continued barriers to direct-acting antiviral treatment in people living with HIV in the Netherlands. Aids 2022;36(6):773–83. https://doi.org/10.1097/qad.0000000000003159
[100] Bruun-Rasmussen P, Andersen PK, Banasik K, Brunak S, Johansson PI. Estimating the effect of donor sex on red blood cell transfused patient mortality: a retrospective cohort study using a targeted learning and emulated trials-based approach. eClinicalMedicine 2022;51:101628. https://doi.org/10.1016/J.ECLINM.2022.101628
[101] De La Court F, Boyd A, Davidovich U, Hoornenborg E, Schim Van Der Loeff MF, De Vries HJC, et al. Eligibility criteria vs. need for pre-exposure prophylaxis: a reappraisal among men who have sex with men in Amsterdam, the Netherlands. Epidemiol Infect 2022:150. https://doi.org/10.1017/S0950268822001741
[102] Møller AL, Rytgaard HCW, Mills EHA, Christensen HC, Blomberg SNF, Folke F, et al. Hypothetical interventions on emergency ambulance and prehospital acetylsalicylic acid administration in myocardial infarction patients presenting without chest pain. BMC Cardiovasc Dis 2022;22(1). https://doi.org/10.1186/S12872-022-03000-1
[103] Bell-Gorrod H, Fox MP, Boulle A, Prozesky H, Wood R, Tanser F, et al. The impact of delayed switch to second-line antiretroviral therapy on mortality, depending on definition of failure time and CD4 count at failure. Am J Epidemiol 2020;189(8):811–9. https://doi.org/10.1093/aje/kwaa049
[104] Amusa L, Zewotir T, North D, Kharsany ABM, Lewis L. Association of medical male circumcision and sexually transmitted infections in a population-based study using targeted maximum likelihood estimation. BMC Public Health 2021;21(1):1642. https://doi.org/10.1186/s12889-021-11705-9
[105] Kerschberger B, Boulle A, Kuwengwa R, Ciglenecki I, Schomaker M. The impact of same-day antiretroviral therapy initiation under the World Health Organization treat-all policy. Am J Epidemiol 2021;190(8):1519–32. https://doi.org/10.1093/aje/kwab032
[106] Dadi AF, Miller ER, Woodman RJ, Azale T, Mwanri L. Effect of perinatal depression on risk of adverse infant health outcomes in mother-infant dyads in Gondar town: a causal analysis. BMC Pregnancy Childbirth 2021;21(1):255. https://doi.org/10.1186/s12884-021-03733-5
[107] MozafarSaadati H, Mehrabi Y, Sabour S, Mansournia MA, HashemiNazari SS. Estimating the effects of body mass index and central obesity on stroke in diabetics and non-diabetics using targeted maximum likelihood estimation: Atherosclerosis Risk in Communities study. Obes Sci Pract 2020;6(6):628–37. https://doi.org/10.1002/osp4.447
[108] Abdollahpour I, Nedjat S, Almasi-Hashiani A, Nazemipour M, Mansournia MA, Luque-Fernandez MA. Estimating the marginal causal effect and potential impact of waterpipe smoking on risk of multiple sclerosis using the targeted maximum likelihood estimation method: a large, population-based incident case-control study. Am J Epidemiol 2021;190(7):1332–40. https://doi.org/10.1093/aje/kwab036
[109] MozafarSaadati H, Sabour S, Mansournia MA, Mehrabi Y, HashemiNazari SS. Effect modification of general and central obesity by sex and age on cardiovascular outcomes: targeted maximum likelihood estimation in the atherosclerosis risk in communities study. Diabetes Metab Syndr 2021;15(2):479–85. https://doi.org/10.1016/j.dsx.2021.02.024
[110] Almasi-Hashiani A, Nedjat S, Ghiasvand R, Safiri S, Nazemipour M, Mansournia N, et al. The causal effect and impact of reproductive factors on breast cancer using super learner and targeted maximum likelihood estimation: a case-control study in Fars Province, Iran. BMC Public Health 2021;21(1):1219. https://doi.org/10.1186/s12889-021-11307-5
[111] Akhtar S, El-Muzaini H, Alroughani R. Recombinant hepatitis B vaccine uptake and multiple sclerosis risk: a marginal structural modeling approach. Mult Scler Relat Disord 2022;58:103487. https://doi.org/10.1016/j.msard.2022.103487

[112] Clare PJ, Dobbins T, Bruno R, Peacock A, Boland V, Yuen WS, et al. The overall effect of parental supply of alcohol across adolescence on alcohol-related harms in early adulthood-a prospective cohort study. Addiction 2020;115(10):1833–43. https://doi.org/10.1111/add.15005
[113] Wang L, Sun X, Jin C, Fan Y, Xue F. Identification of tumor microenvironment-related prognostic biomarkers for ovarian serous cancer 3-year mortality using targeted maximum likelihood estimation: a TCGA data mining study. Front Genet 2021;12:625145. https://doi.org/10.3389/fgene.2021.625145
[114] Sun X, Wang L, Li H, Jin C, Yu Y, Hou L, et al. Identification of microenvironment related potential biomarkers of biochemical recurrence at 3 years after prostatectomy in prostate adenocarcinoma. Aging (Albany NY) 2021;13(12):16024–42. https://doi.org/10.18632/aging.203121
[115] Chavda MP, Bihari S, Woodman RJ, Secombe P, Pilcher D. The impact of obesity on outcomes of patients admitted to intensive care after cardiac arrest. J Crit Care 2022;69:154025. https://doi.org/10.1016/j.jcrc.2022.154025
[116] Chen C, Chen X, Chen J, Xing J, Hei Z, Zhang Q, et al. Association between preoperative hs-crp/Albumin ratio and postoperative sirs in elderly patients: a retrospective observational cohort study. J Nutr Health Aging 2022;26(4):352–9. https://doi.org/10.1007/s12603-022-1761-4
[117] Ikeda T, Cooray U, Murakami M, Osaka K. Maintaining moderate or vigorous exercise reduces the risk of low back pain at 4 years of follow-up: evidence from the english longitudinal study of ageing. J Pain 2022;23(3):390–7. https://doi.org/10.1016/j.jpain.2021.08.008
[118] Ikeda T, Tsuboya T. Effects of changes in depressive symptoms on handgrip strength in later life: a four-year longitudinal study in England. J Affect Disord 2022;299:67–72. https://doi.org/10.1016/j.jad.2021.11.057
[119] Moreno-Betancur M, Lynch JW, Pilkington RM, Schuh HS, Gialamas A, Sawyer MG, et al. Emulating a target trial of intensive nurse home visiting in the policy-relevant population using linked administrative data. Int J Epidemiol 2022;52(1):119–31. https://doi.org/10.1093/ije/dyac092
[120] Figueroa SC, Kennedy CJ, Wesseling C, Wiemels JM, Morimoto L, Mora AM. Early immune stimulation and childhood acute lymphoblastic leukemia in Costa Rica: a comparison of statistical approaches. Environ Res 2020;182:109023. https://doi.org/10.1016/J.ENVRES.2019.109023
[121] Pang M, Schuster T, Filion KB, Eberg M, Platt RW. Targeted maximum likelihood estimation for pharmacoepidemiologic research. Epidemiology 2016;27(4):570–7. https://doi.org/10.1097/ede.0000000000000487
[122] Pearl M, Balzer L, Ahern J. Targeted estimation of marginal absolute and relative associations in case-control data: an application in social epidemiology. Epidemiology 2016;27(4):512–7. https://doi.org/10.1097/ede.0000000000000476
[123] Kreif N, Tran L, Grieve R, De Stavola B, Tasker RC, Petersen M. Estimating the comparative effectiveness of feeding interventions in the pediatric intensive care unit: a demonstration of longitudinal targeted maximum likelihood estimation. Am J Epidemiol 2017;186(12):1370–9. https://doi.org/10.1093/aje/kwx213
[124] Almasi-Hashiani A, Nedjat S, Mansournia MA. Causal methods for observational research: a primer. Arch Iran Med 2018;21(4):164–9.
[125] Skeem JL, Montoya L, Manchak SM. Comparing costs of traditional and specialty probation for people with serious mental illness. Psychiatr Serv 2018;69(8):896–902. https://doi.org/10.1176/APPI.PS.201700498
[126] Saddiki H, Balzer LB. A primer on causality in data science. arXiv 2018. https://doi.org/10.48550/arxiv.1809.02408
[127] Coyle JR, Hejazi NS, Malenica I, Phillips RV, Arnold BF, Mertens A, et al. Targeting learning: robust statistics for reproducible research. 2020. https://doi.org/10.48550/arXiv.2006.07333
[128] Gruber S, van der Laan MJ. tmle: An R package for targeted maximum likelihood estimation. J Stat Softw 2012;51(13):1–35. https://doi.org/10.18637/jss.v051.i13
[129] Lendle SD, Schwab J, Petersen ML, van der Laan MJ. ltmle: An R package implementing targeted minimum loss-based estimation for longitudinal data. J Stat Softw 2017;81:1–21. https://doi.org/10.18637/JSS.V081.I01
[130] Hernán MA. The hazards of hazard ratios. Epidemiology (Cambridge, Mass.) 2010;21(1):13. https://doi.org/10.1097/EDE.0B013E3181C1EA43
[131] Phillips RV, Laan vdMJ, Lee H, Gruber S. Practical considerations for specifying a super learner. arXiv 2022.10.48550/arXiv.2204.06139.
[132] Moore KL, van der Laan MJ. Covariate adjustment in randomized trials with binary outcomes: targeted maximum likelihood estimation. Stat Med 2009;28(1):39–64. https://doi.org/10.1002/sim.3445
[133] van der Laan MJ. Targeted maximum likelihood based causal inference: Part, I. Int J Biostat 2010;6(2):2. https://doi.org/10.2202/1557-4679.1211
[134] van der Laan MJ. Targeted maximum likelihood based causal inference: Part II. Int J Biostat 2010;6(2):2. https://doi.org/10.2202/1557-4679.1241
[135] van der Laan MJ, Balzer LB, Petersen ML. Adaptive matching in randomized trials and observational studies. J Stat Res 2012;46(2):113–56.
[136] Schomaker M, Luque-Fernandez MA, Leroy V, Davies MA. Using longitudinal targeted maximum likelihood estimation in complex settings with dynamic interventions. Stat Med 2019;38(24):4888–911. https://doi.org/10.1002/sim.8340
[137] Zheng W, Luo Z, van der Laan MJ. Marginal structural models with counterfactual effect modifiers. Int J Biostat 2018;14(1). https://doi.org/10.1515/ijb-2018-0039
[138] van der Laan MJ, Gruber S. Targeted minimum loss based estimation of causal effects of multiple time point interventions. Int J Biostat 2012;8(1). https://doi.org/10.1515/1557-4679.1370
[139] Schnitzer ME, Lok JJ, Bosch RJ. Double robust and efficient estimation of a prognostic model for events in the presence of dependent censoring. Biostatistics 2016;17(1):165–77. https://doi.org/10.1093/biostatistics/kxv028
[140] van der Laan MJ. Causal inference for a population of causally connected units. J Causal Inference 2014;2(1):13–74. https://doi.org/10.1515/jci-2013-0002
[141] Sofrygin O, van der Laan MJ. Semi-parametric estimation and inference for the mean outcome of the single time-point intervention in a causally connected population. J Causal Inference 2017;5(1). https://doi.org/10.1515/jci-2016-0003
[142] Balzer LB, Zheng W, van der Laan MJ, Petersen ML. A new approach to hierarchical data analysis: targeted maximum likelihood estimation for the causal effect of a cluster-level exposure. Stat Methods Med Res 2019;28(6):1761–80. https://doi.org/10.1177/0962280218774936
[143] Balzer LB, van der Laan M, Ayieko J, Kamya M, Chamie G, Schwab J, et al. Two-stage TMLE to reduce bias and improve efficiency in cluster randomized trials. Biostatistics 2021. https://doi.org/10.1093/biostatistics/kxab043
[144] Sofrygin O, Zhu Z, Schmittdiel JA, Adams AS, Grant RW, van der Laan MJ, et al. Targeted learning with daily EHR data. Stat Med 2019;38(16):3073–90. https://doi.org/10.1002/sim.8164
[145] Gruber S, van der Laan MJ. An application of targeted maximum likelihood estimation to the meta-analysis of safety data. Biometrics 2013;69(1):254–62. https://doi.org/10.1111/j.1541-0420.2012.01829.x
[146] Liu Y, Schnitzer ME, Wang G, Kennedy E, Viiklepp P, Vargas MH, et al. Modeling treatment effect modification in multidrug-resistant tuberculosis in an individual patientdata meta-analysis. Stat Methods Med Res 2022;31(4):689–705. https://doi.org/10.1177/09622802211046383
[147] Rose S, Laan MJ. Why match? Investigating matched case-control study designs with causal effect estimation. Int J Biostat 2009;5(1):1. https://doi.org/10.2202/1557-4679.1127
[148] Rose S, van der Laan MJ. A targeted maximum likelihood estimator for two-stage designs. Int J Biostat 2011;7(1):17. https://doi.org/10.2202/1557-4679.1217
[149] Balzer LB, Petersen ML, van der Laan MJ. Adaptive pair-matching in randomized trials with unbiased and efficient effect estimation. Stat Med 2015;34(6):999–1011. https://doi.org/10.1002/sim.6380
[150] Balzer LB, Petersen ML, van der Laan MJ. Targeted estimation and inference for the sample average treatment effect in trials with and without pair-matching. Stat Med 2016;35(21):3717–32. https://doi.org/10.1002/sim.6965
[151] van der Laan MJ, Gruber S. Collaborative double robust targeted maximum likelihood estimation. Int J Biostat 2010;6(1):17. https://doi.org/10.2202/1557-4679.1181
[152] Pirracchio R, Yue JK, Manley GT, van der Laan MJ, Hubbard AE. Collaborative targeted maximum likelihood estimation for variable importance measure: illustration for functional outcome prediction in mild traumatic brain injuries. Stat Methods Med Res 2018;27(1):286–97. https://doi.org/10.1177/0962280215627335
[153] Schnitzer ME, Sango J, FerreiraGuerra S, van der Laan MJ. Data-adaptive longitudinal model selection in causal inference with collaborative targeted minimum loss-based estimation. Biometrics 2020;76(1):145–57. https://doi.org/10.1111/biom.13135
[154] Ju C, Gruber S, Lendle SD, Chambaz A, Franklin JM, Wyss R, et al. Scalable collaborative targeted learning for high-dimensional data. Stat Methods Med Res 2019;28(2):532–54. https://doi.org/10.1177/0962280217729845
[155] Ju C, Schwab J, van der Laan MJ. On adaptive propensity score truncation in causal inference. Stat Methods Med Res 2019;28(6):1741–60. https://doi.org/10.1177/0962280218774817
[156] Ju C, Wyss R, Franklin JM, Schneeweiss S, Häggström J, van der Laan MJ. Collaborative-controlled LASSO for constructing propensity score-based estimators in high-dimensional data. Stat Methods Med Res 2019;28(4):1044–63. https://doi.org/10.1177/0962280217744588
[157] Petersen M, Schwab J, Gruber S, Blaser N, Schomaker M, van der Laan MJ. Targeted maximum likelihood estimation for dynamic and static longitudinal marginal structural working models. J Causal Inference 2014;2(2):147–85. https://doi.org/10.1515/jci-2013-0007
[158] FerreiraGuerra S, Schnitzer ME, Forget A, Blais L. Impact of discretization of the timeline for longitudinal causal inference methods. Stat Med 2020;39(27):4069–85. https://doi.org/10.1002/sim.8710
[159] Zheng W, Petersen M, van der Laan MJ. Doubly robust and efficient estimation of marginal structural models for the hazard function. Int J Biostat 2016;12(1):233–52. https://doi.org/10.1515/ijb-2015-0036
[160] van der Laan M, Gruber S. One-step targeted minimum loss-based estimation based on universal least favorable one-dimensional submodels. Int J Biostat 2016;12(1):351–78. https://doi.org/10.1515/IJB-2015-0054
[161] Cai W, van der Laan MJ. One-step targeted maximum likelihood estimation for time-to-event outcomes. Biometrics 2020;76(3):722–33. https://doi.org/10.1111/biom.13172
[162] Zhu J, Gallego B. Targeted estimation of heterogeneous treatment effect in observational survival analysis. J Biomed Inform 2020;107:103474. https://doi.org/10.1016/j.jbi.2020.103474
[163] Zheng W, van der Laan MJ. Targeted maximum likelihood estimation of natural direct effects. Int J Biostat 2012;8(1). https://doi.org/10.2202/1557-4679.1361
[164] Zheng W, Laan Mvd. Longitudinal mediation analysis with time-varying mediators and exposures, with application to survival outcomes. J Causal Inference 2017;5(2). https://doi.org/10.1515/JCI-2016-0006
[165] Rudolph KE, Sofrygin O, van derLaan MJ. Complier stochastic direct effects: identification and robust estimation. J Am Stat Assoc 2021;116(535):1254. https://doi.org/10.1080/01621459.2019.1704292

[166] Rudolph KE, Díaz I. Efficiently transporting causal direct and indirect effects to new populations under intermediate confounding and with multiple mediators. Biostatistics (Oxford, England) 2022;23(3):789. https://doi.org/10.1093/BIOSTATISTICS/KXAA057
[167] Hejazi NS, Rudolph KE, Van Der Laan MJ, Díaz I. Nonparametric causal mediation analysis for stochastic interventional (in)direct effects. Biostatistics 2022;24(3):686–707. https://doi.org/10.1093/BIOSTATISTICS/KXAC002
[168] Rudolph KE, van der Laan MJ. Robust estimation of encouragement-design intervention effects transported across sites. J R Stat Soc Series B Stat Methodol 2017;79(5):1509–25. https://doi.org/10.1111/rssb.12213
[169] van der Laan MJ, Luedtke AR. Targeted learning of the mean outcome under an optimal dynamic treatment rule. J Causal Inference 2015;3(1):61–95. https://doi.org/10.1515/jci-2013-0022
[170] van der Laan L, Zhang W, Gilbert PB. Nonparametric estimation of the causal effect of a stochastic threshold-based intervention. Biometrics 2022;79(2):1014–28. https://doi.org/10.1111/biom.13690
[171] Lea AN, Leyden WA, Sofrygin O, Marafino BJ, Skarbinski J, Napravnik S, et al. Human immunodeficiency virus status, tenofovir exposure, and the risk of poor coronavirus disease 19 (COVID-19) outcomes: real-world analysis from 6 United States cohorts before vaccine rollout. Clin Infect Dis 2023;76(10):1727–34. https://doi.org/10.1093/CID/CIAD084
[172] Okuzono SS, Wilkinson R, Shiba K, Yazawa A, VanderWeele T, Slopen N. Residential instability during adolescence and health and wellbeing in adulthood: a longitudinal outcome-wide study. Health Place 2023;80:102991. https://doi.org/10.1016/J.HEALTHPLACE.2023.102991
[173] Rivera AS, Al-Heeti O, Petito LC, Feinstein MJ, Achenbach CJ, Williams J, et al. Association of statin use with outcomes of patients admitted with COVID-19: an analysis of electronic health records using superlearner. BMC Infect Dis 2023;23(1). https://doi.org/10.1186/S12879-023-08026-0
[174] DelliCarpini G, Giannella L, Carboni M, Fichera M, Pizzagalli D, Segnalini N, et al. Cervical pessary for preterm birth prevention after an episode of arrested preterm labor: a retrospective cohort study with targeted maximum likelihood estimation of the average treatment effect. Eur Rev Med Pharmacol Sci 2023;27(3):1058–68. https://doi.org/10.26355/EURREV_202302_31202
[175] Matsuyama Y. Time-varying exposure analysis of the relationship between sustained natural dentition and cognitive decline. J Clin Periodontol 2023;50(6). https://doi.org/10.1111/JCPE.13786
[176] Dias A, Nunes HrdC, Ruiz-Frutos C, Gomez-Salgado J, Sproesser Alonso M, Bernardes JM, et al. The impact of disease changes and mental health illness on readapted return to work after repeated sick leaves among Brazilian public university employees. Front Public Health 2023. https://doi.org/10.3389/FPUBH.2022.1026053
[177] Elduma AH, Holakouie-Naieni K, Almasi-Hashiani A, Rahimi Foroushani A, Mustafa Hamden Ali H, Adam MAM, et al. The targeted maximum likelihood estimation to estimate the causal effects of the previous tuberculosis treatment in multidrug-resistant tuberculosis in Sudan. PLOS ONE 2023;18(1). https://doi.org/10.1371/JOURNAL.PONE.0279976
[178] Nugent JR, Balzer LB. A demonstration of modified treatment policies to evaluate shifts in mobility and COVID-19 case rates in US counties. Am J Epidemiol 2023;192(5):762–71. https://doi.org/10.1093/AJE/KWAD005
[179] Shiba K, Daoud A, Hikichi H, Yazawa A, Aida J, Kondo K, et al. Uncovering heterogeneous associations between disaster-related trauma and subsequent functional limitations: a machine-learning approach. Am J Epidemiol 2023;192(2):217–29. https://doi.org/10.1093/AJE/KWAC187
[180] Ellison JE, Kumar S, Steingrimsson JA, Adhikari D, Charlesworth CJ, McConnell KJ, et al. Comparison of low-value care among commercial and medicaid enrollees. J Gen Intern Med 2023;38(4):954–60. https://doi.org/10.1007/S11606-022-07823-8/METRICS
[181] Moreno-Betancur M, Lynch JW, Pilkington RM, Schuh HS, Gialamas A, Sawyer MG, et al. Emulating a target trial of intensive nurse home visiting in the policy-relevant population using linked administrative data. Int J Epidemiol 2023;52(1):119. https://doi.org/10.1093/IJE/DYAC092
[182] Cooray U, Tsakos G, Heilmann A, Watt RG, Takeuchi K, Kondo K, et al. Impact of teeth on social participation: modified treatment policy approach. J Dent Res 2023;102(8):887–94. https://doi.org/10.1177/00220345231164106
[183] Lee SY, Modzelewski KL, Law AC, Walkey AJ, Pearce EN, Bosch NA. Comparison of propylthiouracil vs methimazole for thyroid storm in critically Ill patients. JAMA Netw Open 2023;6(4):e238655. https://doi.org/10.1001/JAMANETWORKOPEN.2023.8655
[184] Ackerman KS, Hoffman KL, Díaz I, Simmons W, Ballman KV, Kodiyanplakkal RP, et al. Effect of sepsis on death as modified by solid organ transplantation. Open Forum Infec Dis 2023;10(4). https://doi.org/10.1093/OFID/OFAD148
[185] Bosch NA, Teja B, Law AC, Pang B, Jafarzadeh SR, Walkey AJ. Comparative effectiveness of fludrocortisone and hydrocortisone vs hydrocortisone alone among patients with septic shock. JAMA Intern Med 2023;183(5):451–9. https://doi.org/10.1001/JAMAINTERNMED.2023.0258
[186] Li Y, Yang Y, Zhao P, Wang J, Mi B, Zhao Y, et al. Longitudinal associations between specific types/amounts social contact and cognitive function among middle-aged and elderly Chinese: a causal inference and longitudinal targeted maximum likelihood estimation analysis. J Affect Dis 2023;331:158–66. https://doi.org/10.1016/J.JAD.2023.03.039
[187] Hatano Y, VanWagner LB, Carnethon MR, Bancks MP, Carson AP, Lloyd-Jones DM, et al. Racial difference in the association between non-alcoholic fatty liver disease and incident type 2 diabetes: findings from the CARDIA study. Diabetologia 2023;66(7):1235–46. https://doi.org/10.1007/S00125-023-05903-W/METRICS. 1–12.
[188] Rytgaard HCW, Eriksson F, van der Laan MJ. Estimation of time-specific intervention effects on continuously distributed time-to-event outcomes by targeted maximum likelihood estimation. Biometrics 2023. https://doi.org/10.1111/BIOM.13856
[189] Petersen JM, Naimi AI, Bodnar LM. Does heterogeneity underlie differences in treatment effects estimated from SuperLearner versus logistic regression? An application in nutritional epidemiology. Ann Epidemiol 2023;83:30–4. https://doi.org/10.1016/J.ANNEPIDEM.2023.04.017
[190] Balzer LB, Laan Mvd, Ayieko J, Kamya M, Chamie G, Schwab J, et al. Two-stage TMLE to reduce bias and improve efficiency in cluster randomized trials. Biostatistics (Oxford, England) 2023;24(2):502. https://doi.org/10.1093/BIOSTATISTICS/KXAB043
[191] Tackney MS, Morris T, White I, Leyrat C, Diaz-Ordaz K, Williamson E. A comparison of covariate adjustment approaches under model misspecification in individually randomized trials. Trials 2023;24(1). https://doi.org/10.1186/S13063-022-06967-6
[192] Ren J, Cislo P, Cappelleri JC, Hlavacek P, DiBonaventura M. Comparing g-computation, propensity score-based weighting, and targeted maximum likelihood estimation for analyzing externally controlled trials with both measured and unmeasured confounders: a simulation study. BMC Med Res Methodol 2023;23(1):18. https://doi.org/10.1186/S12874-023-01835-6
[193] Box GE. Science and statistics. J Am Stat Assoc 1976;71(356):791–9. https://doi.org/10.1080/01621459.1976.10480949
[194] Chernozhukov V, Chetverikov D, Demirer M, Duflo E, Hansen C, Newey W, et al. Double/debiased machine learning for treatment and structural parameters. Econom J 2018;21(1):C1–68. https://doi.org/10.1111/ECTJ.12097
[195] Zivich PN, Breskin A. Machine learning for causal inference: on the use of cross-fit estimators. Epidemiology 2021;32(3):393–401. https://doi.org/10.1097/ede.0000000000001332
[196] Balzer L, Ahern J, Galea S, van der Laan M. Estimating effects with rare outcomes and high dimensional covariates: knowledge is power. Epidemiol Methods 2016;5(1):1–18. https://doi.org/10.1515/em-2014-0020
[197] Benkeser D, Carone M, Gilbert PB. Improved estimation of the cumulative incidence of rare outcomes. Stat Med 2018;37(2):280–93. https://doi.org/10.1002/sim.7337
[198] Díaz I, Colantuoni E, Rosenblum M. Enhanced precision in the analysis of randomized trials with ordinal outcomes. Biometrics 2016;72(2):422–31. https://doi.org/10.1111/biom.12450
[199] Díaz I, Williams N, Hoffman KL, Schenck EJ. Nonparametric causal effects based on longitudinal modified treatment policies. J Am Stat Assoc 2021. https://doi.org/10.1080/01621459.2021.1955691
[200] Díaz I, van der Laan MJ. Doubly robust inference for targeted minimum loss-based estimation in randomized trials with missing outcome data. Stat Med 2017;36(24):3807–19. https://doi.org/10.1002/sim.7389
[201] Gruber S, van der Laan MJ. tmle: An R package for targeted maximum likelihood estimation. J Stat Softw 2012;51(13):1–35.
[202] Benkeser DC, Carone M, Gilbert PB. Improved estimation of the cumulative incidence of rare outcomes. Stat Med 2017;37(2):280–93. https://doi.org/10.1002/sim.7337
[203] Benkeser DC, Hejazi NS. survtmle: Targeted Minimum Loss-Based Estimation for Survival Analysis in R.2017.10.5281/zenodo.835868.
[204] Petersen M, Schwab J, Gruber S, Blaser N, Schomaker M, Laan Mvd. Targeted maximum likelihood estimation for dynamic and static longitudinal marginal structural working models. J Causal Inference 2014;2(2):147. https://doi.org/10.1515/JCI-2013-0007
[205] Benkeser D. drtmle: Doubly-Robust Nonparametric Estimation and Inference. 2017.10.5281/zenodo.844836.
[206] Benkeser D, Carone M, Laan MJD, Gilbert PB. Doubly robust nonparametric inference on the average treatment effect. Biometrika 2017;104(4):863–80. https://doi.org/10.1093/BIOMET/ASX053
[207] Sofrygin O, van der Laan MJ, Neugebaueer R. stremr: Streamlined estimation for static, dynamic and stochastic treatment regimes in longitudinal data. R package 2017. ⟨https://github.com/osofr/stremr⟩ (accessed Jan 1, 2023).
[208] Luedtke AR, Sofrygin O, van der Laan MJ, Carone M. Sequential double robustness in right-censored longitudinal models. arXiv 2017. https://doi.org/10.48550/arXiv.1705.02459
[209] Coyle JR, Hejazi NS. origami: A generalized framework for cross-validation in R. J Open Source Softw 2018;3(21). https://doi.org/10.21105/joss.00512
[210] Zheng W, van der Laan MJ. Asymptotic theory for cross-validated targeted maximum likelihood estimation. U.C. Berkeley Division of Biostatistics Working Paper Series 2010.
[211] Coyle JR, Hejazi NS, Phillips RV, van der Laan L, van der Laan MJ. hal9001: the scalable highly adaptive lasso.doi:10.5281/zenodo.3558313; 2022. R package version 0.4.2.
[212] Hejazi NS, Coyle JR, van der Laan MJ. hal9001: Scalable highly adaptive lasso regression in R. J Open Source Softw 2020. https://doi.org/10.21105/joss.02526
[213] Van Der Laan M. A generally efficient targeted minimum loss based estimator based on the highly adaptive Lasso. Int J Biostat 2017;13(2). https://doi.org/10.1515/IJB-2015-0097/MACHINEREADABLECITATION/RIS
[214] Coyle JR, Hejazi NS, Malenica I, Phillips RV, Sofrygin O. sl3: modern pipelines for machine learning and super learning. ⟨https://github.com/tlverse/sl3⟩; 2021. R package version 1.4.2. (accessed Jan 1, 2023).
[215] Coyle JR. tmle3: the extensible TMLE framework. 2021. doi:10.5281/zenodo.4603358.

[216] Malenica I, Coyle J, van der Laan, MJ. tmle3mopttx: targeted learning and variable importance with optimal individualized categorical treatment. ⟨https://github.com/tlverse/tmle3mopttx⟩; 2022. R package version 1.0.0. (accessed Jan 1, 2023).
[217] Hejazi NS, Coyle JR, van der Laan, MJ. tmle3shift: targeted learning of the causal effects of stochastic interventions. ⟨https://github.com/tlverse/tmle3shift⟩; 2021. R package version 0.2.0. (accessed Jan 1, 2023).
[218] Diaz I, van der Laan M. Population intervention causal effects based on stochastic interventions. Biometrics 2012;68(2):541–9. https://doi.org/10.1111/j.1541-0420.2011.01685.x
[219] Ju C, Gruber S, van der Laan MJ. ctmle: collaborative targeted maximum likelihood estimation. R package version 0.1.1 2017. https://CRAN.R-project.org/package=ctmle. (accessed Jan 1, 2023).
[220] Hejazi NS, van der Laan MJ, Benkeser D. haldensify: Highly adaptive lasso conditional density estimation in R. J Open Source Softw 2022;7(77):4522. https://doi.org/10.21105/JOSS.04522
[221] Ertefaie A, Hejazi NS, van der Laan MJ. Nonparametric inverse probability weighted estimators based on the highly adaptive lasso. Biometrics 2022;79(2):1029–41. https://doi.org/10.1111/biom.13719
[222] Hejazi NS, van der Laan MJ, Janes HE, Gilbert PB, Benkeser DC. Efficient nonparametric inference on the effects of stochastic interventions under two-phase sampling, with applications to vaccine efficacy trials. Biometrics 2020;77(4):1241–53. https://doi.org/10.1111/biom.13375
[223] Hejazi NS, Benkeser DC. txshift: Efficient estimation of the causal effects of stochastic interventions in R. J Open Source Softw 2020. https://doi.org/10.21105/joss.02447
[224] Hejazi NS, Benkeser DC. txshift: Efficient estimation of the causal effects of stochastic interventions. J Open Source Softw 2022. https://doi.org/10.5281/zenodo.4070042
[225] Williams NT, Díaz I. lmtp: non-parametric causal effects of feasible interventions based on modified treatment policies. 2021. R package version 1.3.1.
[226] Zivich P, Davidson-Pilon C, Diong J, Reger D, Badger TG. zEpid. ⟨https://github.com/pzivich/zepid⟩; 2022. Python package version 0.9.1. (accessed Jan 1, 2023).
[227] Luque-Fernandez MA. ELTMLE: Stata module to provide ensemble learning targeted maximum likelihood estimation. Statistical Software Components 2021⟨https://ideas.repec.org/c/boc/bocode/s458337.html⟩.
[228] Gruber S, Phillips RA, Lee H, Ho M, Concato J, van der Laan MJ. Targeted learning: towards a future informed by real-world evidence. Arxiv. 2022. arXiv:2205.08643.
[229] Sentinel Innovation Center. Sentinel Innovation Center Master Plan. 2021. ⟨https://www.sentinelinitiative.org/news-events/publications-presentations/innovation-center-ic-master-plan⟩. (accessed Jan 1, 2023).
[230] Gruber S, Lee H, Phillips R, Ho M, van der Laan MJ. Developing a targeted learning-based statistical analysis plan. Stat Biopharm Res 2022:1–8. https://doi.org/10.1080/19466315.2022.2116104

Table A.1
Articles on randomized control trials

Authors	Year	Journal	Disease area	Research question	TMLE method	Challenges	Contribution to research	Development of understanding aided by TMLE?	
Arnold et al.	2009	Int J Epidemiol	Diarrhea, household water treatment, handwashing	Health effectiveness of behavior-based water and hygiene interventions	TMLE	Residual confounding	Minimal sustained water treatment and handwashing behavior, which consequently led to no impacts on acute gastrointestinal, respiratory or anthropometric measures	Findings are consistent with efficacy trials of household water treatment that have found that health impacts are contingent on compliance	Confirmed
Moore et al.	2011	Stat Med	Drug-to-drug interaction, all-cause mortality	Evaluate safety based on mortality because of drug-to-drug interaction	TMLE	Efficiency in the estimation of marginal effects using logistic regression models	Covariate adjustment for binary outcomes using logistic models canincrease the estimation efficiency (precision) for the marginal effect of treatment when the probability of receiving treatment is 50%	The gain in efficiency can have real implications in phase III RCT as was demonstrated with the fact that the test for superiority would provide different conclusions using either the unadjusted or adjusted estimation approaches	New
Hubbard et al.	2012	Int J Biostat	Gabapentin, painful neuropathy, diabetes	Effectiveness of gabapentin among Type I and Type II diabetic patients	LTMLE	Time of onset of treatment related side effects	The treatment effect on average final pain scores was estimated to be reduced to 0.78 for a population where no unmasking occurred	The methods described here provide a methodology to use such data in estimating causal treatment effects that are not influenced by perception	Expanded
Wester et al.	2012	AIDS Res Hum Retroviruses	Antiretroviral therapy, CD4+ cell count, nevirapine, efavirenz	Retrospectively evaluated the causal effect of assigned nonnucleoside reverse transcriptase inhibitor (NNRTI) on time to virologic failure or death [intent-to-treat] and time to minimum of virologic failure, death, or treatment modifying toxicity	TMLE with effect modification	Effect modification	Efavirenz-treated women and Nevirapine-treated men had more favorable combination antiretroviral therapy (cART) outcomes	TMLE appears to be an efficient technique that allows for the clinically meaningful delineation and interpretation of the causal effect of NNRTI treatment and effect modification by sex and baseline CD4+ cell count strata in this study	Expanded
Decker et al.	2014	J Causal Inference	CD4+ cell count, antiretroviral therapy, nonnucleoside reverse transcriptase inhibitor	Assess if either gender or baseline CD4 level modify the effect of two combination antiretroviral therapy (cART) therapies of interest, efavirenz (EFV) and nevirapine (NVP), on the progression of HIV	LTMLE (Longitudinal TMLE)	Model misspecification	Early, sustained intervention on total calories had a greater impact than a physical activity intervention or nonsustained interventions	Multivariable linear regression yielded inflated effect estimates compared to estimates based on targeted maximum likelihood estimation and data-adaptive super learning	Acknowledged strengths of methodology
Balzer et al.	2017	Clin Trials	Test-and-treat strategy, human immuno-deficiency virus, pre-exposure prophylaxis	How to target pre-exposure prophylaxis to high-risk groups and how to maximize power to detect the individual and combined effects of universal test-and-treat and pre-exposure prophylaxis strategies?	TMLE, with and without pair matching	Limited to the size of the adjustment set due to few independent units (i.e., villages)	Nesting a pre-exposure prophylaxis study within an ongoing trial can lead to combined intervention effects greater than those of universal test-and-treat alone and can provide information about the efficacy of pre-exposure prophylaxis in the presence of high coverage of treatment for HIV+ persons	Data-adaptively adjusting for baseline covariates measured at both the individual and clusters levels, the latter consistently leads to notable gains in attained power, while maintaining nominal confidence intervals	Expanded

(*continued on next page*)

Table A.1 (*continued*)

Authors	Year	Journal	Disease area	Research question	TMLE method	Challenges	Contribution to research	Development of understanding aided by TMLE?	
Price et al.	2018	Biometrics	Vaccine efficacy, tetravalent dengue vaccine	Primary analyses assessed vaccine efficacy (VE) against symptomatic, virologically confirmed dengue (VCD) occurring at least 28 days after the third immunization through to the month 25 visit	TMLE	Longitudinal setting, surrogate outcome	A third useful feature of the proposed approach is that the estimated optimal surrogate-in being built by super-learner followed by a TMLE update-contains all information about the average clinical treatment effect in the original trial.	None specified	
Sridhar et al.	2018	N Engl J Med	Dengue serostatus, dengue vaccine safety and efficacy	Newly developed dengue anti-nonstructural protein 1 (NS1) IgG enzyme-linked immunosorbent assay (ELISA) to differentiate between anti-NS1 antibodies induced by wild-type dengue infection and those induced by vaccination to infer baseline dengue serostatus and reanalyze vaccine safety and efficacy according to serostatus	TMLE	Case–cohort study	The tetravalent dengue vaccine protected against severe virologically confirmed dengue (VCD) and hospitalization for VCD for 5 y in persons who had exposure to dengue before vaccination, and there was evidence of a higher risk of these outcomes in vaccinated persons who had not been exposed to dengue	Results were consistent with studies of other analytical approaches	Confirmed
Havlir et al.	2019	N Engl J Med	Human immunodeficiency virus (HIV), test-and-treat, community health approach	Test the hypothesis that universal HIV treatment and annual testing delivered with a community-based, multidisease, patient-centered approach would result in a lower number of new HIV infections and better community health than the current standard of care	TMLE	Cluster-randomized trial, pair matched	Universal HIV treatment did not result in a significantly lower incidence of HIV infection than standard care, probably owing to the availability of comprehensive baseline HIV testing and the rapid expansion of antiretroviral therapy (ART) eligibility in the control group	None specified	
Nguyen et al.	2020	J Trauma Acute Care Surg	Severely injured patients, transfusion ratios	Relationship between the actual transfusion ratios in specific intervals of time and outcomes (mortality and hemostasis) at different time-points during the first 24 h	TMLE	Residual confounding	Transfusion ratios had no significant impact on mortality over time. However, receiving higher ratios of platelets and plasma relative to red blood cells hastens hemostasis in subjects who have yet to achieve hemostasis within 3 h after hospital admission	None specified	
Dayan et al.	2020	Vaccine	Dengue vaccine, disease, and serostatus	Assessment of the long-term efficacy of a dengue vaccine against symptomatic, virologically-confirmed dengue disease by baseline dengue serostatus	TMLE	Not known	The tetravalent dengue vaccine was shown to maintain efficacy against symptomatic virologically confirmed dengue (VCD) in seropositive participants aged ≥9 y up to 6 y after the first dose. Persistence of efficacy was also observed in seropositive participants aged 6–8 y	None specified	

(*continued on next page*)

Table A.1 (*continued*)

Authors	Year	Journal	Disease area	Research question	TMLE method	Challenges	Contribution to research	Development of understanding aided by TMLE?	
Potter et al.	2021	Int J Environ Res Public Health	Physical activity intervention, dog owners, obedience training	Examining Obedience Training as a Physical Activity Intervention for Dog Owners: Findings from the Stealth Pet Obedience Training (SPOT) Pilot Study	TMLE	Not known	Attending a basic dog obedience training course may lead dog owners to walk more and sit less	None specified	
Nyabuti et al.	2021	PLoS One	Human immunodeficiency virus, seroconverters, test-and-treat	Characterize seroconverters and risk factors of HIV infection where high levels of population level viral suppression were acheived	TMLE	Residual confounding	Some demographic groups such as young girls, alcohol users, mobile populations, men who engage in transactional sex as well as women in intergenerational sexual relationships continue to record high HIV incidence rates. Consequently, in order to achieve control of the HIV epidemic, there is need for expansion of existing preventive interventions like PrEP and development of other targeted prevention interventions that are tailored not only to the unique needs of these populations but also to their contextual regional differences	None specified	
Hickey et al.	2021	PLoS Med	Hypertension, all-cause mortality	Effect of a patient-centered hypertension delivery strategy on all-cause mortality	TMLE	Clusters due to community-level randomization	Implementation of a patient-centered hypertension care model was associated with a 21% reduction in all-cause mortality and a 22% improvement in hypertension control compared to standard care among adults with baseline uncontrolled hypertension	None specified	
Amato et al.	2022	Environ Health Perspect	Diarrhea, biodigester cookstove interventions	Effect of daily reported biogas cookstove use on incident diarrhea among children < 5 y old in the Kavrepalanchok District of Nepal	CV-TMLE	Address bias induced by the use of a proxy exposure variable, we employed doubly robust estimation methods with an additional layer of sample-splitting (cross-validation).	This analysis provides new evidence that child diarrhea may be an unintended health risk of biogas cookstove use	Evidence that child diarrhea is an unintended health risk associated with biogas cookstove use in rural Nepal	New
Hickey et al.	2022	PLoS One	Chronic hypertension care, one-time financial incentive	Effect of a one-time financial incentive on linkage to chronic hypertension care	TMLE	Imbalance of baseline confounders in the randomization process	One-time financial incentives and phone-based follow-up to ensure linkage are effective strategies for increasing linkage to hypertension care following community-based screening in rural East Africa	None specified	

(*continued on next page*)

Table A.1 (*continued*)

Authors	Year	Journal	Disease area	Research question	TMLE method	Challenges	Contribution to research	Development of understanding aided by TMLE?	
Marquez et al.	2022	Clin Infect Dis	Social network characteristics, tuberculosis, rural locations	Association between social network characteristics and prevalent tuberculosis infection	Longitudinal TMLE	Cluster-randomized trial	Social networks with higher centrality, more men, contacts with HIV, and tuberculosis infection were positively associated with tuberculosis infection. Tuberculosis transmission within measurable social networks may explain prevalent tuberculosis not associated with a household contact	Utilizing longitudinal TMLE allowed for flexible nonparametric adjustment of covariates that occur during the follow-up	New
Kahler et al.	2022	Psychol Addict Behav	Alcohol use, motivational intervention, behavioral intervention	Examine relative important of client change language subtypes as predictors of alcohol use following motivational interviewing	TMLE	Variable importance analyses to rank order change language subtypes	Alcohol counseling, clients' expressions of concern about taking certain steps or setting certain goals around their drinking—which could be considered arguments against behavior change—actually predicted better drinking outcomes. Thus, as long as a client has an intention to change their drinking, frank discussions about what goals and steps they may or may not find acceptable may facilitate behavior change	None specified	

TMLE: targeted maximum likelihood estimation; LTMLE: Longitudinal TMLE; CV-TMLE =cross-validated TMLE; PrEP =pre-exposure prophylaxis.

Table A.2
Articles by discipline

Authors	Year	Journal	Disease area	Research question	Challenges	Contribution to research	Development of understanding aided by TMLE?	
Behavioral epidemiology								
Mackey DC et al.	2011	Am J Epidemiol	Healthy living, older age	Physical activity and hip fracture	No randomized control trials	Little difference in hip fracture risk for men with moderate or high physical activity levels relative to those with low physical activity level	TMLE analyses, which [can be] regarded [as] the least biased because the estimation procedure was doubly robust, indicated little difference in hip fracture risk for men with moderate or high physical activity levels relative to those with low physical activity level	Acknowledge strengths of methodology
Ahern J et al.	2016	Epidemiology	Mental health, childhood hardship, adolescent health, ethnicity	Childhood adversities and mental disorders in adolescents	Multidimensional exposure	Childhood adversities play an important role in the burden of mental disorders in adolescents, particularly for behavior disorders and to some extent for distress and substance disorders. However, despite the substantially higher burden of adversities experienced by black and Hispanic adolescents, they have a minor role in the patterns of disparities between racial/ethnic groups in mental disorders. Notably, fear disorders were the most common in all racial/ethnic groups, highest in black youth, and largely unrelated to childhood adversities	Findings that childhood adversities had the largest population attributable risks differences for behavior disorders and the smallest for fear disorders are consistent with analyses of these data combined across racial/ethnic groups	Confirmed
Platt JM et al.	2018	Am J Epidemiol	Childhood hardship, IQ	Association between 11 childhood adversities and intelligence	Multidimensional exposure	Interventions attempting to support and improve cognition in individuals who report childhood adversity can be a useful complement to interventions for emotional and behavioral disturbances	Using TMLE, we identified several significant relationships between diverse forms of childhood adversity and lower levels of fluid intelligence. These were identified by estimating differences in mean Kaufman-Brief Intelligence Test (K-BIT) scores in the presence and absence of the adversities, as well as the risk of low vs. average K-BIT score	Expanded

(*continued on next page*)

Table A.2 (*continued*)

Authors	Year	Journal	Disease area	Research question	Challenges	Contribution to research	Development of understanding aided by TMLE?	
Rodrıguez-Molina D et al.	2019	Int Arch Occup Environ Health	Occupational epi, respiratory disease	Improve knowldedge about prevention against occupational asthma	Treatment effect heterogeneity bias, low sample size, high prop missing values	We found that using an instructional video as educational intervention is an effective approach to improve knowledge about preventive measures against occupational asthma and allergies in Bavarian farm apprentices	Findings suggest that our educational intervention improves the knowledge about preventive measures against asthma and allergies in about 20% of young Bavarian farm apprentices, and that TMLE is an efficient double-robust and semiparametric method able to provide causal effect estimates where traditional regression methods cannot	Expanded
Torres JM et al.	2019	Epidemiology	Migration, older age, care, deprivation	Is family-member migration associated with unmet caregiving needs among older adults who remain in low and middle-income settings?	Longitudinal relationships, sample attrition, time-varying covariates	It may be that the long-term - but not short-term – absence of adult children had adverse consequences for women's physical functioning as they aged into older adulthood. We also found entirely null associations between having an adult child in the US and physical functioning for men	This research extends the primarily cross-sectional or short-term follow-up studies used in prior research to a large cohort followed over an 11-y period [enabled through use of TMLE]	Expanded
Ehrlich SF et al.	2020	Am J Epidemiol	Pregnancy/prenatal exposure, healthy living, birth outcomes	Risk of small or large for gestational age (S/LGA) according to exercise during 1st trimester of pregnancy	Discussion of assumptions, model misspecification	In underweight and normal-weight women only, meeting the lower exercise threshold recommended by the Physical Activity Guidelines for Americans also appears to increase the risk of small for gestational age (SGA) and decrease the risk of large for gestational age (LGA)	The present study is the first to report an association between exercise performed during the first trimester and infant size at birth, and to observe that meeting the Physical Activity Guidelines for Americans during early pregnancy increases the risk of SGA and decreases the risk of LGA in underweight and normal-weight women	New
Bodnar LM et al.	2020	Am J Clin Nutr	Nutritional epi, pregnancy	Associations between fruit and vegetable intake relative to total energy intake and adverse pregnancy outcomes	Dichotomization of exposure, complex interactive effects between exposure and outcome, multidimensional exposure, curse of dimensionality	TMLE produced effect estimates with less variation that suggested protective associations for diets high in fruits and vegetables relative to energy on risk of preterm birth, SGA birth, and preeclampsia	The differences in results between Super Learner with TMLE and logistic regression suggest that dietary synergy, which is accounted for in machine learning, may play a role in pregnancy outcomes. This innovative methodology for analyzing dietary data has the potential to advance the study of diet patterns	Expanded

(*continued on next page*)

Table A.2 (*continued*)

Authors	Year	Journal	Disease area	Research question	Challenges	Contribution to research	Development of understanding aided by TMLE?	
Kagawa RMC et al.	2020	Ann Epidemiol	Social epi, deprivation, mental health	Effect of fire arm involvement during violent victimization on the level of distress experienced and daily functioning within sociodemographic subgroups	Missing data	Victimization with a firearm is more distressing than victimization with another weapon or no weapon and that this response is almost universal across age, sex, race, and socioeconomic position	Results are consistent with research supporting a specific association between exposure to firearm violence and negative mental health outcomes	Confirmed
Puryear SB et al.	2020	AIDS	HIV, alcohol	Assessing the effect of alcohol use across the entire cascade, from diagnosis to viral suppression	Longitudinal effects	Via the multiple steps of the cascade, HIV-positive drinkers had significantly worse viral suppression outcomes than nondrinkers	Alcohol use was associated with decreased ART use at study entry (i.e., prior to the implementation of the SEARCH treatment intervention) is consistent with earlier studies in general populations in South Africa, Uganda, and the United States	Confirmed
Torres JM et al.	2020	Am J Epidemiol	Migration, oder age, physical activity	Association between adult child US migration status and change in cognitive performance scores			This is the first study to have evaluated the relationship between adult child migration status and cognitive decline. Departing from most prior studies of adult child migration status and health, we evaluated longitudinal associations with a doubly robust estimation approach that accounted for respondent attrition	New
Clare PJ et al.	2020	Addiction	Alcohol, adolescent health	Effect of parental supply of alcohol on alcohol-related outcomes in early adulthood	Time-varying confounding, causal effects	Further evidence that parental supply of alcohol in adolescence has effects on a number of negative outcomes in early adulthood, including binge drinking and alcohol-related harm, leading not only to increased risk of binge drinking and harm but also increased frequency of binge drinking and number of harms experienced. Analysis of earlier initiation of supply showed that the magnitude of the effect of parental supply increased the earlier that supply was initiated	Robust statistical techniques were used to account for the complex sources of bias that can be introduced by longitudinal analysis of observational data	Acknowledge strengths of methodology

(*continued on next page*)

Table A.2 (*continued*)

Authors	Year	Journal	Disease area	Research question	Challenges	Contribution to research	Development of understanding aided by TMLE?	
Kang L et al.	2021	J Safety Res	Helmet use of cyclists, risk taking behaviors	Risk-taking behaviors under various urban-street conditions, as a function of helmet use	Self selection, survey, heterogeneous effects, model misspecification	Based on 131 survey participants, a significant positive risk compensation effect has been identified using the TMLE estimator and the size of effect is estimated to be about 15.6%	Our analysis also demonstrates the significance of using SuperLearner-based approach to account for possible model misspecification error. In our case, if we choose to use traditional linear models, no statistically significant results would be obtained	Expanded
Torres JM et al.	2021	Int J Geriatr Psychiatry	Older age, caring responsibilities, health outcomes	Evaluated the effect of spousal caregiving on multiple health outcomes in middle-aged and older adults in Mexico	Reverse causality, survey waves, longitudinal data	Select evidence of adverse associations between spousal caregiving and past-week depressive symptoms: These adverse associations are generally described as the result of the emotional and physical burden of caregiving, which may have negative consequences for sleep, time for leisure and health promoting activities, and social isolation	While our analysis improves upon prior methods used to evaluate the health effects of spousal caregiving in observational studies, we are not able to rule out residual unmeasured confounding; we therefore interpreted estimates as associations	Expanded
Lee JO et al.	2021	Public Health	Employment, mental health, Covid-19	Examined the association of employment insecurity with two mental health measures, depression and anxiety	Reverse causality, missing data, sample weights	Employment insecurity has threatened mental health in the United States during the pandemic, and mental health repercussions are not felt equally across the population	Causal interpretation of the results from stratified analyses warrants particular caution because the smaller sample size may threaten the assumptions needed to interpret coefficients from TMLE as causal effects	Acknowledge strengths of methodology and highlight further sources of concern
Shiba K et al.	2021	Epidemiology	Older age, mental health	We present an analysis that estimates and compares prevalence of depressive symptoms under alternattive hypothetical interventions in social participation	Longitudinal data, survey waves, reverse causation, measurement bias/mis-classification	Past studies linking social participation and depressive symptoms in late life have not rigorously considered the time-varying nature of social participation. First study that explicitly estimated and compared the effects of alternative hypothetical interventions in social participation at 2 time points on subsequent depressive symptoms	Past studies linking social participation and depressive symptoms in late life have not rigorously considered the time-varying nature of social participation. Applied an analytic approach that addressed time-dependent confounders and performed doubly robust estimation with a machine learning-based ensemble estimator	Expanded

(*continued on next page*)

Table A.2 (*continued*)

Authors	Year	Journal	Disease area	Research question	Challenges	Contribution to research	Development of understanding aided by TMLE?	
Ikeda T et al.	2022	J Affect Disord	Older age, mental health, strength	Examine the magnitude of the association between depressive symptoms over 2 y and weak handgrip strength among English people in 4 y of follow-up	Survey waves, sample attrition - selection bias, self-reported exposure	The main finding of our study was that people who maintained nondepressive symptoms or improved depressive symptoms were less likely to have weak handgrip strength than those with persistent depressive symptoms	We applied the TMLE model in which time-variant variables (i.e., exposure, covariates, and outcome) and time-invariant variables were concurrently taken into account to obtain a more robust estimation. By contrast, we demonstrated the reverse, that is, worsening mental health led to poorer physical function	Expanded
Ikeda T et al.	2022	J Pain	Older age, physical activity	We hypothesized that older individuals who maintained physical activity over time tend to have a lower risk of low back pain, whereas those who discontinued activity were not	Self reported exposure, small sample, sample attrition	Overall, the present study confirmed that maintaining physical activity reduced the risk of low back pain at the follow-up survey. Conversely, discontinuing activity (engaged only at the baseline survey) was not beneficial	Discrepancies between studies can be explained by the differences in the method of analysis, that is, whether changes in physical activity or other time-varying covariates were taken into account. Thus, time-varying exposures are considered essential to avoid erroneous conclusions	Expanded
Biomarker epidemiology								
Bembom O et al.	2009	Stat Med	Mutations, virology	Determine which of a set of candidate viral mutations affects clinical virologic response to the antiretroviral drug lopinavir + rank the importance of these mutations for drug-specific resistance	Identify subset of relevant biomarkers, biomarker importance	The subset of mutations identified by this approach as significant contributors to lopinavir resistance was in better agreement with the current knowledge than the subsets identified by an unadjusted analyses or the G-computation approach. In addition, the specific ranking provided by targeted variable important measures (VIM) estimation also agreed better with the current understanding than did the rankings generated with alternative methods	Our analysis suggests that targeted maximum-likelihood estimation of VIM represents a promising new approach for studying the effects of HIV mutations on clinical virologic response to antiretroviral therapy	Expanded

(*continued on next page*)

Table A.2 (*continued*)

Authors	Year	Journal	Disease area	Research question	Challenges	Contribution to research	Development of understanding aided by TMLE?	
Rosenblum M et al.	2009	PLoS One	HIV, virology	Effect of adherence on viral load after different durations of viral suppression	Selection bias, unmeasured confounding	These data suggest that for adherence proportions greater than 50%, the probability of virologic failure decreases with longer duration of viral suppression	The estimation method we used relied on having included all confounders of adherence and virologic failure in our analyses, and on our marginal structural model and other models used being correctly specified. While we included many of the known predictors of adherence and virologic failure, unmeasured confounders may lead to bias in our estimates	Acknowledge strengths of methodology and highlight further sources of concern
Gianfrancesco MA et al.	2016	Genes Immun	Genetics, ethnicity, skin disease	Examine clinically important marginal effects of a Genetic Risk Score (GRS) composed of 41 established genetic risk loci on systemic lupus erythematosus activity over a period of 9 y	Little awareness of how genetic profiles have an impact on disease activity, longitudinal effects, time-dependent confounding, sample attrition	Results from individual single-nucleotide polymorphism (SNP) analyses provide important insight to the overall GRS findings; specifically, evidence for significant associations between certain SNPs and Systemic Lupus Activity Questionnaire (SLAQ) score at 2 time points during the longitudinal study was demonstrated	Using a robust method of statistical analysis, our findings do not support a strong causal relationship between an overall GRS comprised of established Systemic lupus erythematous (SLE) SNPs and disease activity as measured by the validated self-reported SLAQ	New
Hsu LI et al.	2016	Cancer Epidemiol Biomarkers Prev	Ethnicity, cancer, children health	Identify a list of candidate genes within each significantly enriched pathway in childhood leukemia, while accounting in models, for the complex correlation between single-nucleotide polymorphisms (SNPs)	Large, correlated data, data reduction, gene selection, variable importance	The results demonstrate that newly developed bioinformatics tools and causal inference methods may illuminate new and biologically relevant pathways and genes to improve current understanding of pathogenesis in childhood leukemia	The results demonstrate that newly developed bioinformatics tools and causal inference methods [TMLE] may illuminate new and biologically relevant pathways and genes to improve current understanding of pathogenesis in childhood leukemia	Expanded
Salihu HM et al.	2016	Matern Child Health J	Ethnicity, pregnancy outcomes	Describe the methylation patterns of 20 candidate genes associated with preterm birth and evaluate their role in preterm births in African-American women	Residual confounding	This study examined 42 CpG (5'—C—phosphate—G—3') sites within 20 candidate genes previously linked to preterm birth and identified 3 CpG sites on 2 distinct genes (TNF-α and PON1) that were differentially methylated between black and nonblack newborns	An additional strength is the robust methodology applied in our analysis. To avoid spurious associations, we employed both Receiver Operating Curve (ROC) and TMLE to identify levels of methylation differences across CpG sites. Results were validated and replicated through bootstrapping	Acknowledge strengths of methodology

(*continued on next page*)

Table A.2 (*continued*)

Authors	Year	Journal	Disease area	Research question	Challenges	Contribution to research	Development of understanding aided by TMLE?	
Izano MA et al.	2020	PLoS One	Birth, mental health	Evaluate the relationship between newborn telomere length and a comprehensive suite of chronic maternal stressors	Complex joint effects	We found that a greater proportion of Latina mothers reported financial strain, food insecurity, and high job strain, while a greater proportion of Black mothers reported poor neighborhood quality, experiencing stressful/ traumatic life events, or having an unplanned pregnancy than other racial/ ethnic groups	Our flexible modeling approach, correction for multiple testing, and consideration of joint effects differ from previous approaches, and may account for some differences in findings. The use of cross-validated ensemble learners improved model fit and reduced bias due to model misspecification, allowing us to establish associations not observed with traditional regression approaches	Acknowledge strengths of methodology, expanded
Wang L et al.	2021	Front Genet	Cancer, environmental epidemiology	Identify tumor microenvironment-related genes to estimate their effects on the 3-y mortality of ovarian cancer	Variable selection, complex joint effects, model misspecification	ARID3C, CROCC2, FREM2, and PTF1A were identified as prognostic biomarkers for OSC patients. Two of them (FREM2 and PTF1A), alongside CROCC, were successfully validated in 3 Gene Expression Omnibus (GEO) datasets	The prognostic biomarkers were driven from a causal inference framework-based TMLE algorithm. Such methodologies can be used to better inform future clinical therapy.	Acknowledge strengths of methodology
Sun X et al.	2021	Aging (Albany NY)	Cancer	Identify the potential prognostic genes in the prostate adenocarcinoma microenvironment and estimate the causal effects simultaneously	Causal effects, variable selection	Based on this strategy, we identified 14 genes involved in the prognosis of prostate adenocarcinoma. The interaction between Prostate adenocarcinoma (PRAD) and tumor microenvironment (TME) might have serious effects on tumor evolution, further influencing tumor resistance, recurrence, and overall prognosis	Using the traditional approaches, such as linear or logistic regression models, confounding factors and complex associations among covariates might bias the results and lead to fallacious conclusions. Whereas, robust TMLE was demonstrated to help reduce the risk of spurious findings [17]. Although TMLE optimizes the bias-variance trade off for the estimated causal effects, a rough trend could still be observed for the individual effects of patients. Based on this strategy, we identified 14 genes involved in the prognosis of PRAD	Acknowledge strengths of methodology, expanded

(*continued on next page*)

Table A.2 (*continued*)

Authors	Year	Journal	Disease area	Research question	Challenges	Contribution to research	Development of understanding aided by TMLE?	
Environmental epidemiology								
Padula AM et al.	2012	Am J Epidemiol	Prenatal exposure, birth outcomes	Estimate, at the population level, the predicted probability of term low birth weight (LBW) had everyone been exposed to each quartile of traffic density	Clear definitions of outcome and exposure + previous inconsistent findings, measurement error, misclassification of exposure	The results did not show a clear exposure–response relation across the quartiles of traffic density; however, there was a significant difference in the predicted probability of LBW between the highest and lowest quartiles of exposure, showing that higher traffic density is associated with increased probability of LBW	None specified	
Herrera R et al.	2017	Int J Environ Res Public Health	Respiratory conditions, children health	Quantify the causal attributable risk of living close to the mines on asthma or allergic rhinoconjunctivitis risk burden in children	Assumptions discussed/checked, proxy measures of exposure, questionnaire, small sample size, missing outcome values	Results indicated that a hypothetical intervention intending to increase the distance from children's home to the mines could result in a reduction of rhinoconjunctivitis prevalence in the studied population by up to 4.7 percentage points (95% CI: −8.4%; −1.1%)	Our method estimates the public health impact of such an intervention, which could not be done using the standard statistical approaches (e.g., logistic regression estimates)	Expanded
Pearl M et al.	2018	Paediatr Perinat Epidemiol	Life course epidemiology, pregnancy outcomes	Evaluate the separate and combined relations of neighborhood opportunity in early-life and adulthood with preterm birth risk in Black, White, and Latina women + assess the contribution of neighborhood opportunity to racial-ethnic disparities in risk of pre-term birth (PTB)	Longitudinal information	Our results do not point to a common susceptible period across racial/ethnic groups, possibly reflecting unmeasured heterogeneity of experiences represented by the poverty measure, or other unmeasured cofactors	Models account for time-dependent confounding often overlooked in life-course studies of birth outcomes, and the targeted estimation may reduce bias in outcome regression models	Acknowledge strengths of methodology
Rudolph KE et al.	2019	Environ Epidemiol	Adolescent health	Examine the relation between environmental noise and adolescent health in the US	Missing data - multiple imputation, sampling weights, propensity score matching, different definitions for exposure, unmeasured confounding/ measurement error	We found that living in a community where average day-night noise exceeded the US Environmental Protection Agency's safety guideline of 55 dB was associated with approximately 30-40 min later bedtimes. Associations with Diagnostic and Statistical Manual of Mental Disorders 4th edition mental disorders were mixed, generally with wide CIs, and not robust across sensitivity analyses	None specified	

(*continued on next page*)

Table A.2 (*continued*)

Authors	Year	Journal	Disease area	Research question	Challenges	Contribution to research	Development of understanding aided by TMLE?	
Casey JA et al.	2019	Environ Res	Prenatal exposure	Do lower Socio-economic status (SES) pregnant women have a heightened response to nconventional natural gas development (UNGD) activity due to coexposure to other environmental and social stressors?	Mediation analysis, missing data - multiple imputation, exposure and outcome measure, assumptions discussed/ checked	Our findings revealed an association between living in the highest quartile of a cumulative metric of UNGD activity during pregnancy and increased risk of antenatal anxiety or depression. This increased risk, however, did not appear to mediate the observed association between UNGD activity and preterm birth or reduced term birth weight, as we found no relationship between antenatal anxiety or depression and these outcomes in our sample	None specified	
Papadopoulou E et al.	2019	Environ Health Perspect	Prenatal exposure, biomarkers	Study the association between diet and measured blood and urinary levels of environmental contaminants in mother–child pairs from 6 European birth cohorts	Variable selection	We estimated that adherence to the dietary recommendations (pregnant women: ≤3 servings = week, children ≤2 servings = week) for fish intake would result in lower exposure to per- and polyfluoroalkyl substances (PFASs), arsenic (As), and mercury (Hg) compared with those exceeding these recommendation. Fruit consumption was associated with increased levels of urinary organophosphate pesticide (OP) metabolites concentrations in both pregnant women and children. Using TMLE analysis, we found that consuming more than 2 fruits could increase the exposure of pregnant women to OPs, compared with lower fruit intake	None specified	Expanded
Nardone A et al.	2021	Environ Health Perspect	Historical epi	Assess the association between Home Owners' Loan Corporation (HOLC) grade and 2010 normalized difference vegetation index, a measure of overall greenness.		We found evidence of an association between worse historical HOLC grade and less 2010 greenspace using data from 102 US urban metropolitan areas.	None specified	

(*continued on next page*)

Table A.2 (*continued*)

Authors	Year	Journal	Disease area	Research question	Challenges	Contribution to research	Development of understanding aided by TMLE?	
Shiba et al. (2022)	2022	Am J Epidemiol	Physical health	Association between disaster-related trauma and functional limitations in a cohort of older survivors of the 2011 Great East Japan Earthquake and Tsunami	Heterogeneous effects of home loss on functional limitations	There was strong evidence of population average effects of home loss on increased functional limitations across all indicators 2.5 y and 5.5 y after the disaster. There was evidence of heterogeneity in the associations between home loss and functional limitation. There were patterns in predisaster characteristics of subgroups particularly vulnerable to functional impairment following home loss	Our finding for the population average effects of home loss on increased functional limitations is consistent with what has been reported previously. Our inductive approach for assessing effect heterogeneity provided potentially new insights that could have been missed with a deductive approach	Expanded
Shiba et al. (2022)	2022	Environ Health Perspect	Physical health, health and well-being	Longitudinal associations between disaster-related home loss and a comprehensive array of subsequent health and well-being outcomes	Model misspecification	Home loss was consistently associated with persistent mental health problems; there was robust evidence for increased posttraumatic stress symptoms (PTSS), and somewhat more modest evidence for increased depressive symptoms and risk of hopelessness at the 9-y follow-up point after the disaster. Home loss was associated with broader indices of well-being that prior epidemiologic studies have not examined. There was modest evidence linking home loss with increased chronic conditions, higher BMI, and decreased happiness	This analytic approach was used because we conditioned on many covariates, and a conventionalestimation approach using parametric outcome regression would be prone to model misspecification.We extended the previous evidence by leveraging the natural experiment design and adjusted for a comprehensive set of predisaster characteristics, including preexposure outcome levels	Expanded
Health economy								
Berkowitz S et al.	2019	Prev Chronic Dis	Financial resources and food insecurity, healthcare expenditure	Estimate the association between state- and county-level healthcare expenditures and food insecurity	Unreliable assumptions in Generalized Linear Models (GLM), ecological measures of exposure and outcome estimated	Adults who were food insecure had annual healthcare expenditures that were $1834 (95% CI, $1073–$2595) higher than adults who were food secure ($P < 0.001$). In children, the model-based estimate for health care costs associated with food insecurity was $80 annually, but this finding was not significant ($P = 0.53$, 95% CI, −$171 to $329)	None specified	

(*continued on next page*)

Table A.2 (*continued*)

Authors	Year	Journal	Disease area	Research question	Challenges	Contribution to research	Development of understanding aided by TMLE?	
Noncommunicable disease epidemiology								
Legrand M et al.	2013	Crit Care	Postoperative, surgery, kidney	To determine the risk factors for postoperative acute kidney injury (AKI) in patients operated on for inefective endocarditis	Multifactorial reasons for the exposure–outcome association	Postoperative AKI following cardiopulmonary bypass for IE results from additive hits to the kidney. We identified several potentially modifiable risk factors such as treatment with vancomycin or aminoglycosides or preoperative anemia	Despite the SuperLearner procedure, which is intended to optimize the prediction, our predictive performance was in fact limited	Acknowledge strengths of methodology and highlight other sources of concern
Leslie HH et al.	2014	PLoS One	Hormonal contrception, cervical cancer	To determine whether potential risk factors such as hormonal contraception are true causes for cervical cancer amongst human immunodeficiency virus-positive women in developing countries	Timing and duration of the exposure, type and measurement of the outcome	Although selected results suggest an increased prevalence of cervical intraepithelial neoplasia 2 or greater (CIN2+) associated with combined oral contraceptive (COC), evidence is insufficient to conclude causality	None specified	
Mosconi L et al.	2018	PLoS One	Menopause, Alzheimer's disease (AD), cognitive performance	Examines the impact of the menopausal transition on Alzheimer's disease (AD) biomarker changes and cognitive performance in midlife	Alzheimer's disease is progressive and cross-sectional studies does not capture the ongoing AD process	The optimal window of opportunity for therapeutic intervention to prevent or delay progression of AD endophenotype in women is early in the endocrine aging process	None specified	
Torres JM et al.	2018	Int J Epidemiol	Adult child migrant, mental health	We examined longitudinal associations between having an adult child migrant and mental health, for middle-aged and older Mexican adults accounting for complex time-varying confounding	A doubly robust, semiparametric estimation strategy that reduces reliance on correct specification of multiple parametric models. Adjust for time-varying confounders affected by prior exposure	Women with at least 1 adult child in the US had a higher adjusted baseline prevalence of elevated depressive symptoms (Risk difference (RD): 0.063, 95% CI: 0.035, 0.091) compared to women with no adult children in the US. Men with at least 1 child in another Mexican city at all three study waves had a lower adjusted prevalence of elevated depressive symptoms at 11-y follow-up (RD: 0.042, 95% CI: 0.082, 0.003) compared to those with no internal migrant children over those waves	The present study utilized longitudinal targeted maximum likelihood estimation to robustly estimate the associations between having an adult child migrant and depressive symptoms over an 11-y period for a national sample of older adults in Mexico. In the overall sample, there was limited evidence of meaningful associations between having an adult migrant child and depressive symptoms	Acknowledge strengths of methodology

(*continued on next page*)

Table A.2 (*continued*)

Authors	Year	Journal	Disease area	Research question	Challenges	Contribution to research	Development of understanding aided by TMLE?	
Yu YH et al.	2019	Am J Epidemiol	Obesity, stillbirth	Estimated the association between incident obesity and stillbirth in a cohort constructed from linked birth and death records in Pennsylvania (2003–2013)	Time of onset of obesity is unknown, unclear whether estimated associations are driven largely by persons with long-standing obesity, recently developed obesity, or combination of both. Cause-effect interpretations rely heavily on assumptions if exposure cannot be randomized	We found positive associations of incident prepregnancy obesity with stillbirth	In this study, we showed that when strong relationships between confounders and exposures exist, nonparametric TMLE can yield highly volatile estimates. In our study, body weight characteristics from prior pregnancies were strongly associated with obesity in the pregnancy of interest. These strong associations resulted in positivity violations that were not apparent when examining the distribution of stabilized inverse probability weights, a commonly used diagnostic strategy	Highlight sources of concern
Lim S et al.	2019	Caries Res	Soda in-take, pediatric caries	Effect of soda intake on dental caries in young children (birth to 5 y)	Longitudinal, time-varying data	Consistent soda intake during the early childhood led to one additional caries tooth surface	The traditional regression analyses showed a positive association between soda intake and dental caries, but itwas not statistically significant. [results from TMLE were significant]	Expanded
Gianfrancesco MA et al.	2019	J Rheumatol	Smoking, rheumatoid arthritis	Examined the association between smoking and rheumatoid arthritis (RA)	Heterogeneous study designs, measurement error in key variables, biases in statistical anaylsis	Smoking is associated with higher levels of disease activity in RA	Our findings show that interpretations surrounding the association between smoking and RA disease activity may differ dramatically depending on the type of statistical analysis conducted. We found current smoking status to be associated with higher levels of disease activity as measured by patient's global assessment (PtGA) score and swollen joint counts (SJC) using Longitudinal TMLE	Expanded

(*continued on next page*)

Table A.2 (*continued*)

Authors	Year	Journal	Disease area	Research question	Challenges	Contribution to research	Development of understanding aided by TMLE?	
Mozafar Saadati H et al.	2020	Obes Sci Pract	BMI, obesity, stroke, diabetics	Compared the effects of body mass index and central obesity on stroke in diabetics and nondiabetics	Model misspecification, different distributions of the exposure between covariates	Among diabetics, body shape index (BSI) and waist-to-hip ratio (WHR) indices were associated with a higher incidence of stroke	With respect to the effects in males, females, and all participants, the results of the TMLE method (the fourth model) were more precise than those based on conventional models and showed that the strongest effect was related to body shape index (BSI) and Body mass index (BMI) for all participants: waist circumference (WC), BSI, and waist-to-hip ratio (WHR) for males and BSI and body roundness index (BRI) for females	Expanded
Veit C et al.	2020	BMC Med Res Methodol	Asthma, control medication	Calculate the long-term risk of reporting asthma symptoms in relation to control medication use in a real-life setting from childhood to adulthood	Randomized clinical trials (RCT) may not represent the general population	We did not observe a beneficial effect of asthma control medication on asthma symptoms	By using a marginal structural models approach, we could account for time-varying treatment and confounding. While we could confirm the targeted maximum likelihood estimation to be a usable and robust statistical tool, from a clinical perspectice we did not observe the desired beneficial effect of asthma control medication on asthma symptoms	Acknowledge strengths of methodology, expanded
Yu YH et al.	2020	Obstet Gynecol	Newly overweight and obesity, stillbirth	Identify the association of newly developed prepregnancy overweight and obesity with stillbirth and infant mortality	Time-varying effects of covariates, timing of exposure	Transitioning from normal weight to overweight or obese between pregnancies was associated with an increased risk of stillbirth and neonatal mortality	None specified	
Decruyenaere A et al.	2020	Crit Care	Obesity, survival of critically ill patients	Association between obesity and improved survival among critically ill patients	Failure to account for confounding and collider stratification bias	TMLE mitigates the obesity paradox observed in critically ill patients, whereas a traditional approach results in even more paradoxical findings	The robust approach that combined targeted learning with multiple imputation to deal with both types of biases yielded an average treatment effect in the untreated (ATU) of −0.59% (95% CI −2.77 to 1.60%, P = 0.599) and thereby mitigated the obesity paradox	Acknowledge strengths of methodology, expanded
Abdollahpour I et al.	2021	Am J Epidemiol	Waterpipe smoke (vape), multiple scelrosis	Role of lifetime waterpipe smoking in the etiology of multiple sclerosis (MS)	Estimate marginal effects	These results suggest that waterpipe use, or strongly related but undetermined factors, increases the risk of multiple sclerosis	None specified	

(*continued on next page*)

Table A.2 (*continued*)

Authors	Year	Journal	Disease area	Research question	Challenges	Contribution to research	Development of understanding aided by TMLE?	
Reilly ME et al.	2021	Foot Ankle Int	Symptomatic hallux valgus, clinical and radiographic outcomes, lapidus procedure vs. scarf osteotomy	To compare clinical and radiographic outcomes between patients with symptomatic hallux valgus treated with the modified Lapidus procedure vs. scarf osteotomy	Positivity violations if not appropriately accounted for	Although the modified Lapidus procedure led to a higher probability of achieving a normal intermetatarsal angle, both procedures yielded similar improvements in 1-y patient-reported outcome measures	None specified	
Torres LK et al.	2021	Thorax	Acute respiratory distress syndrome, mortality	To estimate the attributable mortality, if any, of acute respiratory distress syndrome (ARDS)	Adequately adjusted for confounders, and utilization of statistical methodology to estimate causal effects	Acute respiratory distress syndrome has a direct causal link with mortality	None specified	
Mozafar Saadati H et al.	2021	Diabetes Metab Syndr	Obesity, cardiovascular disease	To elucidate the effect modification of general and central obesity by sex and age on the risk of cardiovascular events	Weakness and misspecification of statistcal models. Mediator effect of some biologic factors	Among males and age 54, waist-to-hip ratio index was associated with a higher risk of coronary heart disease and heart failure while body mass index was so for females and age > 54	None specified	
Almasi-Hashiani A et al.	2021	BMC Public Health	Reproductive factors, breast cancer	To estimate the causal effect of reproductive factors on breast cancer risk in a case–control study	Misspecification of regression models may cause extreme bias in treatment effect estimates	Postmenopausal women, and women with a higher age at first marriage, shorter duration of breastfeeding, and history of oral contraceptive use are at the higher risk of breat cancer	The contradiction of our study findings with other studies may be justified by different confounders adjusted for in analysis and statistical methods used; TMLE method, along with the super learner approach, have been identified more efficient for controlling confounding	Acknowledge strengths of methodology, expanded
Dadi AF et al.	2021	BMC Pregnancy Childbirth	Perinatal depression, infant diarrhea, acute respiratory infection, malnutrition	To estimate associations between perinatal depression and infant diarrhea, Acute Respiratory Infection (ARI), and malnutrition in Gondar Town, Ethiopia	Differences in findings might potentially be due to bias as a result of unobserved confounding, which cannot be overcome using standard regression techniques Failure to account for unobserved confounding can weaken the quality of evidence derived from such studies	There was no evidence for an association between perinatal depression and the risk of infant diarrhea, acute respiratory infection, and malnutrition amongst women in Gondar Town	None specified	
Zou R et al.	2021	Clin Nutr	Maternal folate levels, pregnancy, neuropsychiatic disorders	To estimate the association between prenatal exposure to folate and brain development (neuropsychiatric disorders) in late childhood has been rarely investigated	Appropriately control for confounding	Low maternal folate levels during pregnancy are associated with altered offspring brain development in childhood, suggesting the importance of essential folate concentrations in early pregnancy	None specified	

(*continued on next page*)

Table A.2 (*continued*)

Authors	Year	Journal	Disease area	Research question	Challenges	Contribution to research	Development of understanding aided by TMLE?	
Goel AR et al.	2021	J Am Acad Audiol	Hearing aids, audiometric outcomes	To examine how hearing aids affect standard audiometric outcomes over long-term periods of follow-up	Reduce bias due to study designs that have found differing evidence	Our analysis revealed discernible effects of 5 y of hearing aid use on hearing ability, specifically as measured by the three-frequency pure-tone average (PTA3-Freq), novel extended pure-tone average (PTAExt), and word recognition sound (WRS), suggesting a greater decline in hearing ability in patients using hearing aids	None specified	
Beydoun HA et al.	2021	Sci Rep	Brain metastases, stereotactic radiosurgery	To compare hospitalization outcomes among US inpatients with brain metastases who received stereotactic radiosurgery (SRS) and/or non-SRS radiation therapies without neurosurgical intervention	Model misspecification	Stereotactic radiosurgery (SRS) alone or in combination with non-SRS therapies may reduce the risks of prolonged hospitalization and nonroutine discharge among hospitalized US patients with brain metastases who underwent radiation therapy without neurosurgical intervention	To our knowledge, this study is the first to apply SuperLearner algorithms while estimating average treatment effect (ATE) using TMLE among hospitalized US patients who underwent SRS and/or non-SRS therapies for brain metastases. It has already been established that SRS is two to sixfold more expensive than non-SRS therapies. However, additional research is needed to elucidate shorter hospital stays and fewer non-routine discharges among patients who underwent SRS with or without non-SRS therapies, although fewer neurological complications may be partly responsible for improved clinical outcomes among individuals treated with SRS	Acknowledge strengths of methodology, expanded
Chavda MP et al.	2022	J Crit Care	Obesity, mortality, cardiac arrest	To estimate the conditional and causal effects of obesity on mortality in cardiac arrest patients using the Australian and New Zealand Intensive Care Society (ANZICS) Adult Patient Database (APD)	Methodological issues leading to conflicting results	After adjustment, there was no association between obesity and outcomes in cardiac arrest patients admitted to intensive care unit	Recently, using TMLE approach has alleviated finding of the obesity paradox in critically ill patients which was present with traditional regression analysis, however, this trial did not examine cardiac arrest patients	Acknowledge strengths of methodology, confirmed

(*continued on next page*)

Table A.2 (*continued*)

Authors	Year	Journal	Disease area	Research question	Challenges	Contribution to research	Development of understanding aided by TMLE?	
Crowner JR et al.	2022	Ann Vasc Surg	Chronic limb threatening ischemia, nonoperative management	To assess whether chronic limb threatening ischemia (CLTI) objective performance goals (OPGs) could be attained with nonoperative management alone among patients with CLTI		A comprehensive set treatment goals and expected amputation free survival outcomes can guide revascularization, but also assure that appropriate outcomes are achieved for patients treated without revascularization	None specified	
Akosile M et al.	2018	Int J Clin Biostat Biom	Survival, right heart catheterization (RHC)	Investigated differences in survival among patients with and without right heart catheterization using data from the Study to Understand Prognoses and Preferences for Outcomes and Risks and Treatments (SUPPORT)	Existence of unadjusted bias in the original analysis	Critically ill patients who received a right heart catheterization had a significantly decreased 30- and 60-d survival compared to patients who did not receive one after adjusting for a variety of potential confounder selection strategies	This paper used an innovative alternative to propensity score matching, TMLE, to confirm that patients with RHC had a significantly decreased 30- and 60-d survival in comparison to patients without RHC during initial care. This paper will advance the understanding of TMLE for analysis of observational studies, and promote the application of TMLE in the critical care studies	Acknowledge strengths of methodology, confirmed
Bruun-Rasmussen et al. (2022)	2022	EClinicalMedicine	Transfusion	Determining the causal effect of donor sex on the risk of death after red blood cell (RBC) transfusion in male and female patients	Model misspecification, time-varying confounding	Treating male patients with RBC units exclusively from male donors increases the 28-d survival compared with the current practice. Further, transfusing female patients with RBC units exclusively from donors of either sex increases patient survival compared with the current practice where patients can receive a mix of female and male donated RBC units. If a sex-matched transfusion policy was implemented across all blood banks in Denmark, where ≈ 40.000 patients are transfused annually, our estimates suggest that, annually, 732 (95% CI: 668–800) males and 248 (196–300) females could be saved within 28-d of the first transfusion	The findings from previous observational studies have been conflicting. Our findings suggest beneficial effects of a sex-matched transfusion policy	Expanded

(*continued on next page*)

Table A.2 (*continued*)

Authors	Year	Journal	Disease area	Research question	Challenges	Contribution to research	Development of understanding aided by TMLE?	
Jafarzadeh et al. (2022)	2022	Arthritis Rheumatol	Osteoarthritis (OA)	Examine whether a strategy that reduced pain when the knee pain of participants reached a certain threshold could reduce the risk of a Knee replacement (KR)	Time-dependent confounding and selection bias due to informative censoring because of loss to follow-up or death in the Osteoarthritis Initiative cohort. Time-dependent confounding adjustment avoided the bias of adjusting for intermediate factors	The absolute long-term risk of a KR decreased from 6.3% to ≥5.8% when pain interventions that actually reduced pain wereapplied as knee pain reached ≥5 on the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) pain subscale (translating into 6 KRs avoided per 1000 painful knees). Developing long-term intervention strategies to successfully address chronic knee pain will have significant but modest public health and economic benefits	Our findings suggest that treatments with even modest reductions in pain commensurate with current treatments would substantially decrease knee replacement rates. These data provide additional strong evidence that effective treatments for osteoarthritis are critically needed	Confirmed
Møller et al. (2022)	2022	BMC Cardiovasc Disord	Miocardial infarction (MI)	We investigated whether MI patients without chest pain could be expected to benefit from increased emergency ambulance dispatch and prehospital acetylsalicylic acid (ASA) treatment	Causal framework required for a policy intervention for myocardial infarction	We found no improvement in 30-d survival when hypothetically increasing chance of receiving emergency ambulance dispatch to all nonchest pain MI patients. Increasing prehospital administration of ASA to emergency ambulance transported nonchest pain MI patients was found to reduce 30-d mortality by 3.3% CI 95% [1.4%;5.2%] to 5.3% CI 95% [1.7%;9%] depending on the intervention	Contrary to our hypothesis, the hypothetical intervention on the probability of receiving an emergency ambulance did not change the risk of 30-d mortality and 1-y combined outcome. We found a relatively large reduction in the risk of 30-d mortality among non-chest pain MI patients when hypothetically increasing prehospital ASA assignment	Expanded
Rajan et al. (2022)	2022	Foot Ankle Surg	Foot surgery	Association between concomitant hammertoe correction and surgical outcomes of hallux valgus rection using validated patient-reported outcome measures		Patients in both cohorts demonstrated significant improvements in physical function, pain interference, pain intensity, and global physical health as measured by the Patient-Reported Outcomes Measurement Information System (PROMIS) domains. However, patients who underwent hallux valgus correction with concomitant hammertoe correction were found to have less improvements in the pain interference and pain intensity domains, along with overall higher postoperative pain interference scores, indicating less improvement in pain-related outcomes	In cases where hallux valgus patients exhibit risk factors for developing hammertoe or show early signs of hammertoe formation, our findings can aid surgeons in counseling patients on surgical outcomes if they proceed to develop this lesser toe pathology	Expanded

(*continued on next page*)

Table A.2 (*continued*)

Authors	Year	Journal	Disease area	Research question	Challenges	Contribution to research	Development of understanding aided by TMLE?	
Infectious disease epidemiology								
Schnitzer ME et al.	2014	Biometrics	Hepatitis C virus, liver disease	Despite modern effective HIV treatment, hepatitis C virus (HCV) coinfection is associated with a high risk of progression to end-stage liver disease (ESLD) which has emerged as the primary cause of death in this population	Identifying and adjusting for variables (baseline or time-varying) that affect both HCV clearance and ESLD. Missing data	We found a clinically but not statistically significant protective effect of the clearance of hepatitis C virus on end-stage liver disease, adjusting for time in the model	A protective effect of HCV clearance on ESLD is consistent with studies that have shown curative HCV therapy greatly reduces progression to ESLD, hepatic decompensation, transplantation, hospitalization and death	Confirmed
Davis FM et al.	2017	J Vasc Surg	Surgical site infection, graft failure	Surgical site infection after open lower extremity bypass, leading to increased rate of graft failure	Structural or process-of-care characteristics of the hospitals where the procedures were performed	Surgical site infection after lower extremity bypass is associated with an increase in rate of amputation and reoperation	None specified	
Vauchel T et al.	2019	Am J Infect Control	Imipenem-resistant acinetobacter baumannii (IR-AB), renal outcomes	To explore the impact of an outbreak of imipenem-resistant acinetobacter baumannii (IR-AB) on renal outcomes		The episode of imipenem-resistant acinetobacter baumannii (IR-AB) outbreak was associated with an increased risk of kidney events, which appears to be driven by the use of colistin	Both the pathophysiological background of kidney toxicity of colistin and the robust statistical analysis, using machine learning, strongly suggest that such a causal relationship exists. Also, our protocols were not modified between the 2 periods. Finally, performing a randomized controlled trial in this setting is not feasible	Acknowledge strengths of methodology, expanded
Kempker RR et al.	2020	Clin Infect Dis	Drugs for multidrug-resistant tuberculosis	Bedaquiline and delamanid are newly available drugs for treating multidrug-resistant tuberculosis (MDR-TB); however, there are limited data guiding their use and no comparison studies	Limited data on the clinical outcomes of patients treated with bedaquiline and delamanid under programmatic conditions	Among patients with multidrug-resistant tuberculosis, bedaquiline-based regimens were associated with higher rates of sputum culture conversion, more favorable outcomes, and a lower rate of acquired drug resistance vs. delamanid-based regimens	In the absence of existing data from randomized controlled trials of bedaquiline versus delamanid, the results from our study help inform clinicians and national tuberculosis programs on the relative efficacy of bedaquiline versus delamanid.	Expanded
Westling T et al.	2020	Int J Infect Dis	Sepsis, aqueous chlorhexidine (CHG) used to reduce risk of bloodstream infections (BSI)	Assess the impact of bathing of neonates with 2% chlorhexidine solution on bloodstream infecitons, suspected sepsis, and mortality in a low-income country neonatal care unit	Causal effects from observational data ncluding time-varying confounders	Aqueous chlorhexidine bathing at admission was associated with a reduced risk of bloodstream infections due to a pathogenic organism after adjusting for potential confounding	None specified	

(*continued on next page*)

Table A.2 (*continued*)

Authors	Year	Journal	Disease area	Research question	Challenges	Contribution to research	Development of understanding aided by TMLE?	
Aiemjoy K et al.	2020	PLoS Negl Trop Dis	Distance from water source, *Chlamydia trachomatis*, antibody reponses	Children living further from a water source would have higher exposure to *Chlamydia trachomatis* and enteric pathogens as determined by antibody responses	Flexible modeling approach to capture the relationship between seroprevalence and age	Children living further from a water source had higher seropreva- lence of *Salmonella enterica* and *Giardia intestinalis* indicating that improving access to water in the Ethiopia's Amhara region may reduce exposure to these enteropathogens in young children	None specified	
Figueroa S et al.	2020	Environ Res	Immune stimulation, acute lymphoblastic leukemia	We utilized targeted machine learning and traditional statistical methods to investigate the association of multiple measures of early immune stimulation with acute lymphoblastic leukemia (ALL) in Costa Rican children	Complex biological processes underlying immune dysregulation and leukemogenesis	Exposure to pets and farm animals was inversely associated with acute lymphobalstic leukemia risk, whereas having a fever longer than 1 wk (a putative proxy of severe infection) was associated with an increased risk	These null findings from California Childhood Leukemia Study (CCLS) are somewhat consistent with ours, suggesting that these specific characteristics (i.e., daycare and birth order) may not be the best surrogate measures of early immune stimulation in Hispanic children	Confirmed
Amusa L et al.	2021	BMC Public Health	Male circumcision, sexually transmitted infections	Protective effect of male circumcision against some sexually transmitted infections (STIs)	Mimic an RCT	We present further evidence of a protective association of medical male circumcision against human immunodeficiency virus and herpes simplex virus type-2 in this hyperendemic South African setting	Specifically, we found that medical male circumcision (MMC) has a protective association with HIV and HSV-2. Though the utilization of TMLE did not indicate a null effect nor alter the direction of the association, we found evidence of more precise effects	Expanded
Kerschberger B et al.	2021	Am J Epidemiol	Antiretroviral therapy, unfavorable health outcome	Same-day ART, effect on composite unfavorable treatment outcome	Assess real-world effectiveness	There was a reduced risk of composite unfavorable income amongst those with early antiretroviral therapy initiation comapred to same-day antiretroviral therapy	A strength of this study is that we applied different analytical approaches, including state-of-the-art methods (e.g., TMLE), all of which concurred in their main findings	Confirmed
Akhtar S et al.	2022	Mult Scler Relat Disord	Hepatitis B vaccine (HBV), multiple sclerosis (MS)	HBV vaccine and multiple sclerosis risk	Marginal causal effect over a population	The results suggest a significant nonspecific protective effect of recombinant HBV against multiple sclerosis risk	The effect estimates from these 3 analyses showed a consistent direction of the relationship as a significant nonspecific protective effect of HBV vaccination against MS risk. However, the doubly robust estimates obtained by TMLE were the least biased compared with the estimates obtained in the propensity score matched analysis and conditional logistic regression methods	Acknowledge strengths of methodology, confirmed

(*continued on next page*)

Table A.2 (*continued*)

Authors	Year	Journal	Disease area	Research question	Challenges	Contribution to research	Development of understanding aided by TMLE?	
Chen C et al.	2022	J Nutr Health Aging	Preoperative hs-crp/albumin ratio and postoperative Systemic inflammatory response syndrome (SIRS)	Preoperative high-sensitive C-reactive protein (hs-CRP)/albumin ratio and SIRS	Interactions and stratified analysis	Preoperative hs-CRP/albumin ratio (CAR) was significantly associated with increased risk of postoperative SIRS in elderly patients	None specified	
Isfordink CJ et al.	2022	AIDS	Prevalence hepatitis C virus (HCV)-viremia, direct-acting antivirals	Prevalence of HCV-viremia. Clinical determinants to lack of direct-acting antivirals	Inference to settings with prolonged unrestricted access to treatment	Prevalence of hepatitis C virus-viremic among people with HIV is low in the Netherlands, coinciding with widespread direct-acting antiviral-uptake	None specified	
de la Court et al. (2022)	2022	Epidemiol Infect	Preexposure prophylaxis eligibility criteria, risk of HIV	To reappraise pre-exposure prophylaxis (PrEP) eligibility criteria toward the men who have sex with men (MSM) with highest HIV-risk	These methods, in addition to the use of anal sexually transmitted infection (STI) as HIV proxy, addresses some of the methodological limitations of previous HIV studies with low HIV incidence, allowing for more informative conclusions. Lastly, population attributable fractions provide a valuable addition to TMLE	Including chemsex as an additional PrEP eligibility criterion among MSM could thus further tailor PrEP provision and improve HIV prevention outcomes	Addition of chemsex to the PrEP eligibility criteria in the Netherlands, and in countries with similar epidemics, seems warranted	Confirmed
Occupational epidemiology								
Brown DM et al.	2015	Epidemiology	Particulate matter, heart disease	Incidence of ischemic heart disease (IHD) in relation to accumulated exposure to particulate matter (PM) in a cohort of aluminum workers	Time-varying confounding, a component of the healthy worker survivor effect	The accumulation of exposure to PM2.5 appears to result in higher risks of ischemic heart disease in both aluminum smelter and fabrication work- ers	The TMLE estimate of the risk ratio returned smaller confidence intervals than those from the inverse-probability weighted estimate of the hazard ratio. In 3 of the 4 point estimates, we estimated larger effects using targeted minimum loss-based estimates than unadjusted estimates	Expanded
Izano MA et al.	2019	Environ Epidemiol	Metalworking fluids (MWF), colon cancer	The relation between exposure to straight, soluble, and synthetic MWFs and the incidence of colon cancer in a cohort of automobile manufacturing industry workers	Time-varying confounding affected by prior exposure	Evidence for a causal effect of straight metalworking fluids exposure on colon cancer risk	The reasons for the inconsistencies between our findings and those in the aerospace cohort are not clear but suggest that our adjustment for time-varying confounding affected by prior exposure using TMLE may have allowed us to detect an effect otherwise hidden by healthy worker survivor bias	Acknowledge strengths of methodology, expanded

(*continued on next page*)

Table A.2 (*continued*)

Authors	Year	Journal	Disease area	Research question	Challenges	Contribution to research	Development of understanding aided by TMLE?	
Pharmacoepidemiology								
Sukul D et al.	2017	J Invasive Cardiol	Preprocedural P2Y12 inhibitors, percutaneous coronary intervention (PCI)	Assess the association between preprocedural P2Y12 inhibitor administration and clinically important in-hospital outcomes		Decline in the rate of preprocedural P2Y12 inhibitor administration. No significant differences in outcomes between patients treated with preprocedural P2Y12 inhibitors and those that were not	None specified	
Bell-Gorrod H et al.	2020	Am J Epidemiol	Human immunodeficiency virus-positive patients, mortality	Assessed the impact of delayed switch from first-line ART treatment to second-line ART treatment on mortality in 9 South African treatment programs, a large cohort with long follow-up	Small patient numbers and limited follow-up times in previous studies	Early treatment switch is particularly important for patients with low CD4 counts at failure	Our marginal structural working models were more complex than the marginal structural models (MSMs) in these studies, which makes a more refined interpretation of the dose–response relationship between delay in switching and mortality possible; however, both previous studies [13] and current research [23] suggest that it could be important to allow for even more flexible approaches to model specification and fitting than ours	Acknowledge strengths of methodology and highlight other sources of concern, expanded
Rossides M et al.	2021	Respirology	Methotrexate or azathioprine	To estimate the relative risk of infectious disease at 6 mo associated with the initiation of methotrexate compared to initiation of azathioprine in patients with sarcoidosis		Methotrexate appears to be associated with a lower risk of infection in sarcoidosis than azathioprine, but randomized trials should confirm this finding	Regarding double robustness, we did not observe any advantage of TMLE as point estimates from TMLE and modified Poisson regression models were similar. An exception was the analysis with the lowest power where we restricted to infections resulting in hospitalizations in which TMLE indicated a somewhat larger association (Risk ratio (RR): 0.80 vs. 0.69). TMLE's efficiency manifested in all analyses where 95% CI were considerably narrower than those estimated using Poisson models	Acknowledge strengths of methodology, confirmed

(*continued on next page*)

Table A.2 (*continued*)

Authors	Year	Journal	Disease area	Research question	Challenges	Contribution to research	Development of understanding aided by TMLE?	
Kahkoska AR et al.	2021	Diabetes Care	Treatment methods after positive test of SARS-COV-2 amongst adults	Determine the respective associations of premorbid glucagon-like peptide-1 receptor agonist (GLP1-RA) and sodium-glucose cotransporter 2 inhibitor (SGLT2i) use, compared with premorbid dipeptidyl peptidase 4 inhibitor (DPP4i) use, with severity of outcomes in the setting of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection	Residual confounding	Among SARS-COV-2-positive adults, premorbid GLP1-RA and SGLT2i use, compared with DPP4i use, was associated with lower odds of mortality and other adverse outcomes, although DPP4i users were older and generally sicker	None specified	
Policy								
Tran L et al.	2016	Epidemiol Methods	Human immunodefficiency virus, low-risk express care task-shifting program	Estimating the joint effect of both time to availability of a nurse-based triage system (low risk express care (LREC)) and individual enrollment in the program among HIV patients in East Africa	Point effect of one or more longitudinal exposures, or a series of sequential treatment decisions	Small impact of both availability and enrollment in the low risk express care program on incare survival.	None specified	
Skeem JL et al.	2017	JAMA Psychiatry	Traditional probation, specialty mental health probation	To test whether specialty probation yields better public safety outcomes than traditional probation.	Short follow-up, limited covariate set	Well-implemented specialty probation appears to be effective in reducing general recidivism.	None specified	
Skeem JL et al.	2018	Psychiatr Serv	Costs or traditional and specialty probation	Specialty mental health probation reduces the likelihood of rearrest for people with mental illness	Small sample size	Well-implemented specialty probation yielded substantial savings—and should be considered in justice reform efforts for people with mental illness.	None specified	
You Y et al.	2019	BMC Med Inform Decis Mak	Type 2 diabetes, health program	Assess the performance of a multidisciplinary-team diabetes care program called DIABETIMSS on glycemic control of type 2 diabetes (T2D) patients	Not possible to evaluate a new program by design, impractical to randomize the initiation	DIABETIMSS program had a small, but significant increase in glycemic control	Though the estimates from standard regression were not radically different from those based upon less biased, machine learning methods, they do show enough difference to be important, mainly when the impacts apply to so many patients.	Acknowledge strengths of methodologyConfirmed

(*continued on next page*)

Table A.2 (*continued*)

Authors	Year	Journal	Disease area	Research question	Challenges	Contribution to research	Development of understanding aided by TMLE?	
Mehta B et al.	2021	Arthroplast Today	Primary care physicians (PCPs) density and total knee and total hip arthroplasty outcomes.	To examine the relationship of primary care provider density with total knee arthroplasty and total hip arthroplasty outcomes.	Missing data, zero inflation	No statistically significant association between PCP density and pain, function, or stiffness outcomes at baseline or 2 y	None specified	
Wong AK et al.	2022	Epidemiology	Public masking mandates, COVID-19 cases and deaths	Evaluating the effect of delays in state-level public masking mandates on the relative growth of COVID-19 cases and deaths in the 50 US states from 1 September to 31 October 2020.	Relience on epidemic modeling	Public masking mandates are associated with population-level reductions in COVID-19 spread	Our results are also in line with a regression-based study by Krishnamachari et al., who found that longer delays between the Centers of Disease Control and Prevention guidance and state-level masking mandates were associated with higher cumulative case rates.	Confirmed
Moreno-Betancur M et al.	2022	Int J Epidemiol	Child development	Evaluate effects on child developmental vulnerability after long-term follow up of a family home visiting program.	Mixed evidence from RCTs, target trial to provide suitable evidence	Results did not provide robust evidence of meaningful beneficial or adverse effects of family home visiting program on child development vulnerability.	None specified	
Kagawa et al. (2022)	2022	Prev Med	Building demolitions	Estimate the short- and longer-term effects of demolitions that took place in 2017 on the probability that Detroit Census blocks experienced serious violent crimes.	Time-dependent confounding that are affected by previous treatment. Classical methods would be biased since there are mediators.	Demolition activities in Detroit in 2017 were not associated with the probability of subsequent violent or drug crimes in Census blocks or block groups. At the block group level, demolition was associated with a higher probability of lower level crimes. Null results for drug crimes are similar to previous demolition studies in Detroit, while the null results for violent crimes differ in many, but not all cases	Our results run counter to most previous research on this topic, which tends to show a protective effect of demolition on violent crime. Understanding why our results differ may provide important insights into the types of demolition programs with the greatest potential to reduce violent crime.	Expanded

RCT: Randomized controlled trials; TMLE: targeted maximum likelihood estimation; LTMLE: Longitudinal TMLE.

Table A.3
Articles by developments

Year of publication	Development	Publication's first author	Related developments
2006	Seminal paper (TMLE)	van der Laan MJ	
2009	Small sample size	Moore KL	Gruber, 2010 (sparse data)
	Case–control studies	Rose S	Rose S, 2011 (2-stage sampling, nested case–control), Balzer L, 2015 (adaptive case control design), Balzer L, 2016 (target population different from sample population)
2010	Collaborative TMLE (c-TMLE) time to event data	van der Laan MJ	Pirracchio R, 2018 (variable importance), Ju C, 2019 (c-TMLE with ordering of the covariates to decrease time-complexity of the whole algorithm.), Ju C, 2019 (LASSO for estimation of PS), Ju C, 2019 (positivity-c-TMLE, truncation of PS), Schnitzer ME, 2020 (longitudinal extension of c-TMLE)
	Longitudinal data (LTMLE)	van der Laan MJ	van der Laan, 2012 (TMLE), Schomaker M, 2019 (~ tutorial, ltmle in complex and realistic settings)
2012	Sequential randomized trials	Chaffee PH	
	Mediation	Zheng W	Lendle SD, 2013 (Natural direct effect among untreated), Zheng W, 2017 (time-varying exposure mediated by a time-varying intermediate variable), Zheng W, 2018 (Chapter in Targeted Learning in Data Science), Rudolph KE, 2018 (TMLE for stochastic direct and indirect effect), Rudolph KE, 2020 (Estimators of the complier stochastic direct effect), Diaz I, 2021 (develop asymptotically optimal non-parametric estimators), Benkeser D, 2021 (case-cohort sampling designs), Rudolph KE, 2022 (nonparametric estimators of transported interventional (in)direct effects), Hejazi N, 2022 (causal mediation for stochastic interventional (in)direct effects)
	Nonindependence	van der Laan MJ	van der Laan MJ, 2014 (non iid), Schnitzer M, 2016 (dependent censoring), Sofrygin O, 2017 (TMLE for connected units), Balzer L, 2019 (hierarchical/cluster data structure), Balzer L, 2021 (2-stage TMLE)
2013	Meta-analysis/safety outcomes	Gruber S	Liu Y, 2022
2014	Pooled TMLE	Petersen M	Ferreira Guerra S, 2020 (selecting a timeline discretization for use with pooled longitudinal targeted maximum likelihood estimation), Zheng W, 2016 (pooled TMLE for hazard functions)
	Interval-censored TMLE	Sapp S	
	Genetics	Wang H	Benkeser D, 2019 (vaccine sieve analysis: vaccine and genetic traits), Yang G, 2022 (vaccine sieve)
2015	PS	Lendle SD	
	Cross-validated TMLE, cv-TMLE	van der Laan MJ	
2016	One-step TMLE	van der Laan MJ	Cai W, 2020 (1-step TMLE for counterfactual average survival curve), Zhu J, 2020 (1-step TMLE for heterogeneous treatment effects)
	TMLE for rare outcomes	Balzer L	Benkeser D, 2018
	TMLE for ordinal outcomes	Dıaz I	
2017	TMLE with missing outcome data	Dıaz I	
	Robust TMLE	Rudolph KE	
	Targeted sequential inference of an optimal treatment rule	Chambaz A	
2018	Projected TMLE	Zheng W	
2019	TMLE for cluster-level exposure	Balzer LB	
	Long-format TMLE	Sofrygin O	
2020	Highly-Adaptive least absolute shrinkage and selection operator (LASSO) Targeted Minimum Loss Estimator (HAL-TMLE)	Cai W	
2022	Threshold response function	van der Laan MJ	