

SOFTWARE FOCUS

CausalMetaR: An R package for performing causally interpretable meta-analyses

Guanbo Wang^{1,2,†}, Sean McGrath³  and Yi Lian^{4,†}

¹CAUSALab, Harvard T.H. Chan School of Public Health, Boston, MA, USA

²Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA

³Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA

⁴Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania, Philadelphia, PA, USA

Corresponding author: Guanbo Wang; Email: gwang@hsph.harvard.edu

Received: 12 June 2024; **Revised:** 4 November 2024; **Accepted:** 15 December 2024

Keywords: CausalMetaR; causal inference; heterogeneous treatment effects; meta-analysis; R package; transportability

Abstract

Researchers would often like to leverage data from a collection of sources (e.g., meta-analyses of randomized trials, multi-center trials, pooled analyses of observational cohorts) to estimate causal effects in a target population of interest. However, because different data sources typically represent different underlying populations, traditional meta-analytic methods may not produce causally interpretable estimates that apply to any reasonable target population. In this article, we present the CausalMetaR R package, which implements robust and efficient methods to estimate causal effects in a given internal or external target population using multi-source data. The package includes estimators of average and subgroup treatment effects for the entire target population. To produce efficient and robust estimates of causal effects, the package implements doubly robust and non-parametric efficient estimators and supports using flexible data-adaptive (e.g., machine learning techniques) methods and cross-fitting techniques to estimate the nuisance models (e.g., the treatment model, the outcome model). We briefly review the methods, describe the key features of the package, and demonstrate how to use the package through an example. The package aims to facilitate causal analyses in the context of meta-analysis.

Highlights

What is already known

- Several methods have been recently developed to estimate causal effects in well-defined target populations from multi-source data (e.g., primary studies in a meta-analysis).
- There are no existing software tools implementing these methods.

What is new

- We present the free, open-source CausalMetaR package for estimating causal effects in a well-defined target population from multi-source data.
- CausalMetaR facilitates estimating average and subgroup treatment effects in internal target populations and external target populations.
- CausalMetaR allows users to use machine-learning methods and cross-fitting to estimate the target parameters, and the estimators are doubly/rate robust, non-parametrically efficient, and asymptotically normal.

 This article was awarded Open Data and Open Materials badges for transparent practices. See the Data availability statement for details.

† G.W. and Y.L. equally contributed to this article.

© The Author(s), 2025. Published by Cambridge University Press on behalf of The Society for Research Synthesis Methodology. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<https://creativecommons.org/licenses/by/4.0>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

Potential impact for Research Synthesis Methods readers

- CausalMetaR can help evidence synthesis practitioners perform more rigorous causal analyses

1. Introduction

We consider the setting where one would like to estimate the causal effects of an intervention in a particular target population using multi-source data. While Average Treatment Effects (ATEs) are often of interest in causal analyses, treatment effects may vary based on specific patient characteristics. In such scenarios, estimating Subgroup Treatment Effects (STEs) can help clinicians tailor treatment strategies for patients and lay the foundation for generating future hypotheses.^{1–3}

When data from multiple sources are available, one may naturally consider pooling the data across sources to estimate such causal effects. As a primary example, meta-analyses frequently pool data across several randomized trials. Other examples include analyses of multi-center trials, pooled analyses of observational cohorts, and combined trials and observational cohorts.

Conventional meta-analytic methods pool data across sources by taking a weighted average of the source-specific estimates of the outcome of interest, where the weights are related to the precision of the sources. Typically, meta-analysis practitioners employ random effects models to account for differences in the source populations. However, the pooled estimate from such approaches generally does not have a clear causal interpretation because it does not pertain to a well-defined target population.⁴

Several methods have been recently developed to estimate causal effects for specific target populations when data are available from multiple sources. Broadly, these methods can be categorized as those that estimate causal effects in a so-called *internal target population* (i.e., the population underlying one of the sources contributing covariate, treatment, and outcome data to the analyses) versus an *external target population* (i.e., a population underlying another source where only covariate data can be obtained from the population).

Specifically, Robertson et al.⁵ and Dahabreh et al.⁶ developed methods to estimate ATEs in internal and external target populations, respectively. These methods have subsequently been extended by Wang et al.⁷ to estimate STEs in internal and external target populations, where the subgroups are defined based on a categorical effect modifier.

These methods can be challenging to apply for a few reasons. These methods require fitting a number of models to estimate the so-called *nuisance functions*, such as the conditional outcome mean, the conditional probability of receiving treatment, and the conditional probability of the source index. To handle high-dimensional covariates and help protect against model misspecification, data analysts may want to use flexible machine learning methods to estimate the nuisance functions. Constructing valid confidence intervals for these estimators when using machine learning methods requires sample splitting and cross-fitting, which can be challenging to implement. Consequently, the lack of available software is one barrier to applying these methods in practice. Additionally, implementing these different methods in one software tool can help researchers easily perform secondary analyses, such as estimating treatment effects in different target populations and subgroups.

In this article, we present the CausalMetaR R package⁸ for performing causally interpretable meta-analyses. The package can be used to estimate ATEs and STEs in internal and external target populations. Users of the package can apply a wide range of flexible models to estimate the nuisance functions. The package implements sample splitting and cross-fitting procedures, which are crucial for obtaining valid confidence intervals when using flexible models for the nuisance functions. Users can also generate forest plots of the causal effect estimates in the internal populations. The package requires individual participant data from the data sources (e.g., observational studies and/or randomized trials). The package can be downloaded from the Comprehensive R Archive Network (CRAN) at <https://CRAN.R-project.org/package=CausalMetaR>.

In the following section, we describe the statistical methods implemented in CausalMetaR. We describe how to use the package in Section 3 and illustrate an example application in Section 4. We conclude with a discussion in Section 5.

2. Methods

In this section, we present the methods used by the package. The methods to estimate the ATEs were developed by Robertson et al.⁵ (for an internal target population) and Dahabreh et al.⁶ (for the external target population). The methods to estimate the STEs were developed by Wang et al.⁷ (for both types of target populations).

2.1. Target parameters, conditions, and identification

Consider a collection of datasets comprising m distinct datasets, each containing n_j , $j = 1, \dots, m$ random samples drawn from a unique population. The total sample size of these m datasets is thus $n = n_1 + \dots + n_m$. Each dataset in this collection comprises information on the outcome variable Y , the data source index $S \in \mathcal{S} = \{1, \dots, m\}$, the binary treatment assignment A ($a \in \{0, 1\}$), and the potential high-dimensional covariate matrix X . We do not require the source of each data to be consistent. That is, some data can be collected from clinical trials, and some data can be collected from observational cohorts.

Throughout this article, we use a counterfactual framework.^{9–11} Define the counterfactual outcome Y^a as the outcome had the subject received treatment a . When the target population is the internal target population, we define the target parameter as the counterfactual means difference in the target population, denoted by $E(Y^1 - Y^0 | S = s)$, $\forall s \in \mathcal{S}$. See the data structures in the left panel of Figure 1.

However, sometimes investigators are interested in a target population that is not represented by any of the m sources of the dataset, but by an external source, in which the outcome and treatment information may be missing. In such cases, we define the target parameter as $E(Y^1 - Y^0 | S = 0)$, where $S = 0$ indicates the external data. The sample size is $N = n + n_0$, where n_0 is the sample size of external dataset. See the data structure in the right panel of Figure 2.

In both scenarios, we can estimate the STEs. Denote the interested categorical effect modifier by $\tilde{X} \subset X$. When the target population is an internal target population, we define the target parameters as $E(Y^1 - Y^0 | \tilde{X} = \tilde{x}, S = s)$, $\forall s \in \mathcal{S}$, when the target population is the external population, the target parameter is $E(Y^1 - Y^0 | \tilde{X} = \tilde{x}, S = 0)$.

The above parameters of interest are based on counterfactual outcomes, which are not fully observed. We provide sufficient identifying conditions for the target parameters below.

A1. *Consistency*: If $A_i = a$, then $Y_i^a = Y_i$, $\forall a \in \{0, 1\}$ and i .

A2. *Exchangeability over A*: $Y^a \perp\!\!\!\perp A | (X, S = s)$, $\forall a \in \{0, 1\}$, $s \in \mathcal{S}$.

Y	S	A	X
	1		
	2		
	⋮		
	⋮		
	⋮		
	⋮		
	⋮		
	m		

Figure 1. The data structure required when the target population is an internal population in the multi-source data. The shaded dataset represents an example of a target population.

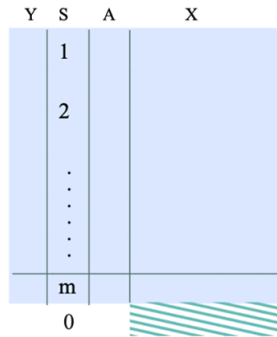


Figure 2. The data structure required when the target population is the external target population. The shaded dataset represents the target population.

Table 1. Identification results for the marginal counterfactual outcomes means in a target population (and subgroup).

	Counterfactual outcome mean	Identified quantity
ATE Internal	$E(Y^a S = s)$	$\psi_{s,a} = E \{ E(Y A = a, X) S = s \}$
ATE External	$E(Y^a S = 0)$	$\psi_a = E \{ E(Y A = a, X) S = 0 \}$
STE Internal	$E(Y^a \tilde{X} = \tilde{x}, S = s)$	$\phi_{s,a}(\tilde{x}) = E \{ E(Y A = a, X) \tilde{X} = \tilde{x}, S = s \}$
STE External	$E(Y^a \tilde{X} = \tilde{x}, S = 0)$	$\phi_a(\tilde{x}) = E \{ E(Y A = a, X) \tilde{X} = \tilde{x}, S = 0 \}$

A3. Positivity of the probability of treatment: If $f(x, S = s) \neq 0$, then $\Pr(A = a|X = x, S = s) > 0, \forall a \in \{0, 1\}, s \in S$.

A4. Exchangeability over S: $\forall a \in \{0, 1\}, Y^a \perp\!\!\!\perp S|X$.

It is important to emphasize that the conditions of exchangeability and treatment positivity are assumed to hold within each individual data source, as opposed to being held to other multi-source datasets. These conditions must be scrutinized in light of substantive knowledge. While it is not possible to empirically test conditions *A2* and *A4*, they do imply a testable condition: $Y \perp\!\!\!\perp S|(X, A = a)$. One may conduct a conditional independence test to evaluate this condition (e.g., using the `ci.test` function in the `bnlearn` package.¹²) For a more comprehensive discussion of these conditions and their implications, as well as potential alternative, less restrictive identifying conditions, interested readers can refer to the papers.^{6,7,13–16}

Through conditions *A1* to *A4*, it can be shown that the counterfactual outcome means in a target population (and subgroup) are identifiable using observed data, as shown Table 1.

Then the target parameters can be identified by the difference of those means. For example, the STE for a internal target population $E(Y^1 - Y^0|\tilde{X} = \tilde{x}, S = s)$ can be identified by $\phi_{s,1}(\tilde{x}) - \phi_{s,0}(\tilde{x})$ (see Table 1).

2.2. Estimation

2.2.1. Influence function-based estimator

While one can estimate the target parameters using the above identification result by replacing $E(Y|A = a, X)$ with corresponding model-based estimators and replacing expectations conditional on S (and $\tilde{X}$) with sample averages, correctly specifying parametric models for the outcome model may be challenging, especially when the covariates are high-dimensional. Incorrect model specification will result in inconsistent estimates for the target parameters. Influence functions can be used to derive non-parametric and efficient estimators of the target parameters. Specifically, one can estimate the target parameters by solving the estimating equations, which equates the empirical mean of the influence functions of the target parameters to zero.

Such estimators can be presented as linear combinations of the following nuisance functions:

1. Outcome model $E(Y|A = a, X)$, the outcome means conditional on a specific treatment and covariates.
2. Source model $\Pr(S = s|X)$, $\forall s \in \mathcal{S}$, the probability of source indices conditional on covariates in the multi-source data.
3. External model $\Pr(S = 0|X)$, the probability of a subject is from the external population conditional on covariates (only applicable when estimating ATEs and STEs for the external target population).
4. Propensity score model $\Pr(A = 1|X, S = s)$, $\forall s \in \mathcal{S}$, the probability of treatment conditional on covariates and source index in the multi-source data.

For example, the estimator for $\phi_{s,a}(\tilde{x})$ is given by

$$\hat{\phi}_{s,a}(\tilde{x}) = \frac{1}{\sum_{i=1}^n M_i} \sum_{i=1}^n \left[M_i \hat{E}(Y_i|A_i = a, X_i) + I(\tilde{X}_i = \tilde{x}_i) \frac{A_i \hat{\Pr}(S_i = s|X_i)}{\hat{\Pr}(A_i = a|X_i)} \{Y_i - \hat{E}(Y_i|A_i = a, X_i)\} \right] \quad (1)$$

where $M_i = I(\tilde{X}_i = \tilde{x}_i, S_i = s)$ and hats denote estimators of the respective parameter. Then $E(Y^1 - Y^0|\tilde{X} = \tilde{x}, S = s)$ can be estimated by $\hat{\phi}_{s,1}(\tilde{x}) - \hat{\phi}_{s,0}(\tilde{x})$. We compute $\hat{\Pr}(A_i = a|X_i)$ by $\sum_{s \in \mathcal{S}} \hat{\Pr}(A_i = 1|S_i = s, X_i = x) \hat{\Pr}(S_i = s|X_i = x)$. These are discussed in the following subsection. Similar estimators apply to the other target parameters.

2.2.2. Nuisance function estimator

The CausalMetaR package leverages the rich capabilities of the SuperLearner package¹⁷ to estimate the nuisance functions described in Section 2.2.1. Doing so gives users a wide array of flexible and robust approaches to estimate the nuisance functions under a unified user interface. Specifically, users can apply one of the following three broad approaches:

1. Users can choose from a range of 41 parametric (e.g., generalized linear models) and nonparametric models (e.g., neural networks, random forests, support vector machines) to estimate the nuisance functions. This can be done by specifying a single algorithm in the library. The model hyperparameters (if applicable) can be fit based on cross-validation.
2. Users can fit multiple models and combine them using the SuperLearner (SL) method. Specifically, SL can either (a) choose the best performing model among the list of candidate models specified by the user, or (b) take an optimally weighted average of the candidate models. Model selection or model averaging is performed based on cross-validation.
3. In the event that users would like to apply a model that is not implemented in SuperLearner (e.g., a Classification And Regression Tree (CART) model), the package allows users to specify their own custom model.

In each of these approaches, SL supports parallel computing.

2.2.3. Stratified sample splitting and cross-fitting

Sample splitting and cross-fitting¹⁸ can be used in estimation to allow users to have a broader choice of data-adaptive (e.g., machine-learning) methods to estimate the nuisance functions and guarantee efficiency.¹⁹

Broadly, this is an iterative procedure. In each replication, the data are split stratified by the data source (and subgroup when estimating STEs), then the nuisance functions and the target parameter are estimated in separate subsamples. The final estimate is the average of the estimates from each subsample. Taking estimating $\phi_s(\tilde{x})$ as an example, Figure 3 illustrates this procedure in one replication.

More specifically, the procedure implemented in CausalMetaR is detailed in Table 2.

When the target population is the external population, we need to estimate one more nuisance function—the external model. In such cases, the data should be split into five sets in Step 1 in Figure 3

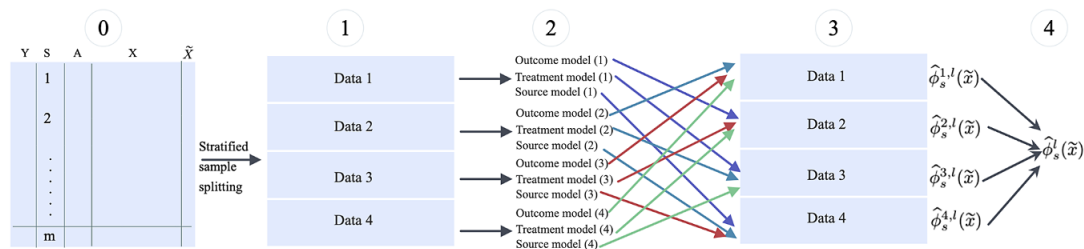


Figure 3. The cross-fitting procedure for estimating $\phi_s(\tilde{x})$ in each of the replication.

Table 2. Procedures of cross-fitting for estimating $\phi_s(\tilde{x})$.

Step	Procedure
0	Data input Repeat Steps 1-4, L times, and for each replication l: (See Figure 3)
1	The data are partitioned into four approximately equal-sized sample splits, denoted by Data k . The partition is stratified by source index (and subgroup when estimating STEs).
2	The outcome, treatment, and source nuisance models are fit in each Data k .
3	The target parameter estimate $\hat{\phi}_s^{k,l}(\tilde{x})$, $\forall k = 1, \dots, 4$ is obtained by (1) using Data k , where the nuisance models are constructed from other data split $\forall k' \neq k$. For example, use Data 1 to obtain $\hat{\phi}_s^{1,l}(\tilde{x})$, where outcome, treatment, and source models are constructed by Data 2, 3, and 4, respectively.
4	The target parameter is calculated from the mean of the predictions across all splits, i.e., $\hat{\phi}_s^l(\tilde{x}) = \sum_{k=1}^4 \hat{\phi}_s^{k,l}(\tilde{x})/4$. The same for the variance estimation.
5	The overall point estimate is $\hat{\phi}_s(\tilde{x}) = \text{median}\{\hat{\phi}_s^l(\tilde{x})\}$, and the overall estimated variance is estimated by $\text{median}[\text{Var}\{\hat{\phi}_s^l(\tilde{x})\} + \{\hat{\phi}_s^l(\tilde{x}) - \hat{\phi}_s(\tilde{x})\}^2]$, $\forall l = 1, \dots, L$

and Table 2, all other procedures are the same. Note that because Step 1 involves stratified sample splitting, the sample size of the multi-source data may not be large enough. For instance, if the data are collected from multiple early-phase trials, sample splitting, and cross-fitting are not recommended.

2.3. Inference

2.3.1. Properties of estimators

The estimators implemented in the package have the following statistical properties. The biases of the estimators depend on how the four nuisance functions described in Section 2.2 are estimated. When these functions are estimated using parametric models, the estimators are consistent if either the outcome model or the collection of the other nuisance function models is correctly specified, which is known as double robustness. When the nuisance functions are estimated non-parametrically, the rates of convergence of the estimators depend on the convergence rates of the non-parametric models. Specifically, if all the nuisance function estimators converge at a rate of $n^{1/4}$ where n is the sample size, estimators converge to the truth at a rate of $n^{1/2}$, which is known as rate robustness.

Additionally, under conditions A1–A4, and when all the nuisance models are correctly specified, the estimators have the smallest variance among the class of doubly robust estimators. Further, the estimators asymptotically follow a normal distribution, allowing us to construct closed-form confidence intervals for them.

2.3.2. Simultaneous confidence intervals

When investigators are interested in estimated STEs in more than one subgroup, using confidence intervals to assess the uncertainty of the target parameters may be misleading. Simultaneous confidence

bands (also known as uniform confidence bands), which take multiple-comparison issues into account, are recommended in such cases to reflect the joint uncertainty of the STE estimates. Our package provides the construction of simultaneous confidence bands when estimating STEs. While simultaneous confidence intervals account for the joint uncertainty in estimating several STEs simultaneously, it does not account for joint uncertainty in estimating treatment effects in multiple target populations simultaneously (which is not the focus of this software).

3. Software functionality

The CausalMetaR package can be downloaded from CRAN and loaded in R as follows:

```
install.packages("CausalMetaR")
library("CausalMetaR")
```

The package has four main functions: (1) `ATE_internal` estimates ATEs in internal target populations, (2) `ATE_external` estimates the ATE in an external target population, (3) `STE_internal` estimates STEs in internal target populations, and (4) `STE_external` estimates STEs in an external target population. The main arguments used in these functions are summarized in Table 3.

The following subsections detail how to specify the data and working models for these functions as well as the output of these functions.

3.1. Data specification

The main functions in CausalMetaR require users to supply R data frames (or vectors) containing data on the outcome, source index, treatment, and covariates. Each of these data frames (or vectors) must

Table 3. Summary of the arguments in the main functions in CausalMetaR.

Argument	Description	Requirements
Data		
Y	Outcome	Vector
S	Source	Categorical vector
A	Treatment	Binary vector
X	Covariates	Data frame
EM ^a	Effect modifier	Categorical vector
X_external ^b	Covariates in external population	Data frame
EM_external ^c	Effect modifier in external population	Categorical vector
Nuisance models		
outcome_model_args	Outcome model arguments	List for SuperLearner
source:model	Source model specification	"MN.glmnet" or "MN.nnet"
source:model_args	Source model arguments	List for SuperLearner
treatment_model_type	Treatment model type	"joint" or "separate"
treatment_model_args	Treatment model arguments	List for SuperLearner
external_model_args ^b	External model arguments	List for SuperLearner
Cross-fitting options		
cross_fitting	Use of cross-fitting	TRUE or FALSE
replications	Number of replications	Integer

^a Only applicable for the `STE_internal` and `STE_external` functions.
^b Only applicable for the `ATE_external` and `STE_external` functions.
^c Only applicable for the `STE_external` function.

contain $n = \sum_{s=1}^m n_s$ rows. The argument *Y* specifies the outcome data, which may be a continuous or binary vector. The data source indicator is specified by the argument *S*, which can be a vector of characters or integers. The argument *A* specifies the treatment data, which must be binary and coded as 0/1. The argument *X* specifies the baseline covariate data, which may contain continuous, binary, and/or categorical covariates.

There are three additional requirements on the input data sets for some of the functions in CausalMetaR. First, when estimating STEs, EM (and EM_external when using STE_external) should be provided as a categorical effect modifier of interest. Note that *X* should *not* include this effect modifier. That is, *X* is defined slightly differently here than the *X* defined in Section 2 (i.e., $X = \{X, EM\}$). Second, when estimating treatment effects in an external target population, users must provide data on the baseline covariates in the target population, specified by the argument *X_external*. The ordering of the columns of *X_external* must be the same as that of *X*.

3.2. Nuisance models specification

Recall that the CausalMetaR package generally uses SL to estimate the working models via the SuperLearner R package.¹⁷ For detailed guidance and discussion on the usage of the package, please refer to the papers.^{20,21} The following subsections describe how to specify the four different working models.

3.2.1. Outcome model

The outcome model is specified by the argument *outcome_model_args*, which is a list that supplies arguments to the SL method (specifically the SuperLearner function in the SuperLearner package). Users should specify the type of model(s) considered by SL through the *SL.library* component in the list. For example, setting *SL.library* to "SL.glm" includes a generalized linear model and setting *SL.library* to "SL.glmnet" includes penalized generalized linear models. Users should generally specify whether the outcome is binary or continuous through the *family* component of the list, where *gaussian()* indicates a continuous outcome and *binomial()* indicates a binary outcome.

For example, we can specify fitting a linear regression model with a Lasso penalty in the *ATE_external* function as follows¹:

```
ATE_external(...,
  outcome_model_args = list(family = gaussian(), SL.library = "SL.glmnet"))
```

3.2.2. Source model

Recall that the source model has a categorical outcome. Since the current version of the SuperLearner package (version 2.0-28) does not fully support multinomial models, the CausalMetaR package has the following two options for specifying the source model. If users set *source:model* to "MN.glmnet", the package will fit a (penalized) multinomial logistic regression model based on the glmnet R package.^{22,23} If users set *source:model* to "MN.nnet" which results in a multinomial log-linear model via neural networks based on the nnet R package.²⁴

3.2.3. Treatment model

The treatment model is specified by the arguments *treatment_model_args* and *treatment_model_type*. Users should supply arguments to the SL method through the argument *treatment_model_type*, similar to the case of the *outcome_model_args* in the outcome model. Users can specify whether to fit a treatment model separately in each data source or whether to fit a single treatment model across all sources through the argument *treatment_model_type*. Setting *treatment_model_type* to "separate" results in fitting $\Pr(A = 1|X, S = s)$ by regressing *A* on *X* in a specific data source *s*. Setting *treatment_model_type* to "joint" results in fitting a single treatment model that includes the source as a categorical predictor, i.e., regressing *A* on *S* and *X* in all the data sources.

For example, we can specify fitting a joint treatment model (across all sources) based on a logistic regression model with a Lasso penalty in the `ATE_external` function as follows:

```
ATE_external(...,
  treatment_model_args = list(family = binomial(),
    SL.library = "SL.glmnet"), treatment_model_type = "joint")
```

3.2.4. External model

Recall that specifying an external model is needed for the `ATE_external` and `STE_external` functions. The argument `external_model_args` specifies the external model in the same manner as the treatment and outcome models.

3.3. Cross-fitting specification

To allow for the use of data-adaptive methods (e.g., machine learning techniques) to model the nuisance functions, sample splitting and cross-fitting are required in the estimation.¹⁹ The use of sample splitting and cross-fitting is specified by the arguments `cross_fitting` and `replicates`. When `cross_fitting` is set to `TRUE`, the package estimates the target parameters through stratified sample splitting and cross-fitting. The argument `replications` specifies the number of cross-fitting replications. To achieve non-sensitive (regarding the sample splitting) results, we recommend users set `replications` to 100L.¹⁸ Because cross-fitting can have high computational cost, we recommend users set `cross_fitting` to `FALSE` when users opt for parametric models to model the nuisance functions or when the sample size of multi-source data is small.

3.4. Output

The four main functions in `CausalMetaR` return lists which include data frames containing the potential outcome mean estimates (`df_A0` and `df_A1`) and treatment effect estimates (`df_dif`). These data frames include the point estimates, standard error (SE) estimates, 95% confidence intervals (CIs), and 95% simultaneous confidence bands (SCBs) when STEs are estimated. Unlike the CIs which reflect a range of values for each (subgroup) treatment effect estimate individually, the SCBs reflect a range of values for all STE estimates simultaneously. Users can also access the fitted working models through the `fit_outcome`, `fit_source`, `fit_treatment`, and `fit_external` components of the output. Though those working models are not accessible in the output when cross-fitting is conducted, users can try fitting models without cross-fitting to check the working models first.

The output of these functions have corresponding `print` and `summary` methods. Moreover, the output of the `ATE_internal` and `STE_internal` functions have corresponding `plot` methods which generate forest plots based on the `metafor` package.²⁵ These functions are illustrated in the complete example in the following section.

4. Example

In this section, we illustrate an application of `CausalMetaR` to an example data set in the package.

The multi-source data set, `dat_multisource`, is a data frame consisting of data from 3 sources with sample sizes of 2,312, 1,147, and 592 respectively. This data frame has columns for a continuous outcome (Y), source indicator (S), binary treatment (A), effect modifier with 5 categories (EM), and nine continuous covariates (X2, ..., X10). The external data set, `dat_external`, is a data frame consisting of the external data in a population with a sample size of 10,083. It has columns for the effect modifier (EM) and the nine covariates (X2, ..., X10).

We estimate the ATE and STEs in the three internal target populations and the external target population. We use SL with (penalized) GLMs and neural networks to estimate the outcome, treatment, and external models. We use separate (source-specific) treatment models, and we use the default multinomial logistic regression model for the source model. We conducted cross-fitting with, for simplicity, 5 replications. Each of the analyses only takes a few minutes to run on a standard laptop computer (8 GB RAM, 1.1 GHz Quad-Core Intel Core i5 processor). Since the example data set was simulated, we obtained the true values of the ATEs and STEs in the internal and external target populations via Monte Carlo integration (see Supplementary Material).

We can apply the CausalMetaR package to perform this analysis as follows. We begin with specifying the working models:

```
outcome_model_args <- list(family = gaussian(),
                           SL.library = c("SL.glmnet", "SL.nnet", "SL.glm"))
treatment_model_args <- list(family = binomial(),
                             SL.library = c("SL.glmnet", "SL.nnet", "SL.glm"))
external_model_args = list(family = binomial(),
                           SL.library = c("SL.glmnet", "SL.nnet", "SL.glm"))
```

Note that we do not explicitly specify the source model here because we will use the default source model.

The following subsections present the R code and output for estimating the ATEs and STEs in the internal and external target populations. In the Supplementary Material, we present additional output from these analyses (e.g., estimates of the potential outcome means).

4.1. Estimating the ATE in the external target population

We apply the ATE_external function to estimate the ATE in the external target population. Since cross-validation is used when fitting the working models (which involves random sampling), we set a random number set for reproducibility.

```
set.seed(1234)
result_ae <- ATE_external(
  Y = dat_multisource$Y,
  S = dat_multisource$S,
  A = dat_multisource$A,
  X = dat_multisource[, 1:10],
  X_external = dat_external[, 1:10],
  outcome_model_args = outcome_model_args,
  treatment_model_args = treatment_model_args,
  external_model_args = external_model_args,
  cross_fitting = TRUE,
  replications = 5)
```

The printed output is given below. It reports an ATE estimate of 6.63 [95% CI: 5.94, 7.32] in the target population.

AVERAGE TREATMENT EFFECT ESTIMATES IN AN EXTERNAL POPULATION

Treatment effect (mean difference) estimates:

Estimate	SE	Lower 95% CI	Upper 95% CI	
6.6294	0.1535	5.8616	7.3972	

4.2. Estimating ATEs in the internal target populations

We can use the `ATE_internal` function to estimate the ATEs in the internal target populations as follows.

```
result_ai <- ATE_internal(
  Y = dat_multisource$Y,
  S = dat_multisource$S,
  A = dat_multisource$A,
  X = dat_multisource[, 1:10],
  outcome_model_args = outcome_model_args,
  treatment_model_args = treatment_model_args,
  cross_fitting = TRUE,
  replications = 5)
```

The printed output is given below. We estimate an ATE of 6.59 [95% CI: 6.31, 6.88] in the population of source A, 7.76 [95% CI: 7.49, 8.03] in the population of source B, and 7.25 [95% CI: 6.99, 7.51] in the population of source C.

AVERAGE TREATMENT EFFECT ESTIMATES IN INTERNAL POPULATIONS

Treatment effect (mean difference) estimates:

```
-----
Source Estimate      SE  Lower 95% CI Upper 95% CI
    A    6.5874 0.1903      6.2145      6.9603
    B    7.7556 0.2577      7.2506      8.2606
    C    7.2916 0.3594      6.5872      7.9960
```

4.3. Estimating STEs in the external target population

Next, we apply the `STE_external` function to estimate the STEs in the external target population as follows.

```
result_se <- STE_external(
  Y = dat_multisource$Y,
  S = dat_multisource$S,
  A = dat_multisource$A,
  X = dat_multisource[, 2:10],
  EM = dat_multisource$EM,
  X_external = dat_external[, 2:10],
  EM_external = dat_external$EM,
  outcome_model_args = outcome_model_args,
  treatment_model_args = treatment_model_args,
  external_model_args = external_model_args,
  cross_fitting = TRUE,
  replications = 5)
```

The printed output is given below. The STE estimates range from 5.41 [95% CI: 4.83, 6.00] (in subgroup e) to 7.58 [95% CI: 6.99, 8.17] (in subgroup c).

SUBGROUP TREATMENT EFFECT ESTIMATES IN AN EXTERNAL POPULATION

Treatment effect (mean difference) estimates:

```
-----
Subgroup Estimate      SE Lower 95% CI Upper 95% CI Lower 95% SCB Upper 95% SCB
    a    7.0787 0.3563      5.9088      8.2485      5.5453      8.6121
    b    5.5207 0.2321      4.5764      6.4650      4.2830      6.7585
```

c	7.5709	0.1805	6.7382	8.4037	6.4794	8.6625
d	6.5748	0.2253	5.6446	7.5051	5.3556	7.7941
e	5.3741	0.3382	4.2343	6.5139	3.8802	6.8681

4.4. Estimating STEs in the internal target populations

Last, we estimate the STEs in the internal target populations by using the STE_internal function.

```
result_si <- STE_internal(  
  Y = dat_multisource$Y,  
  S = dat_multisource$S,  
  A = dat_multisource$A,  
  X = dat_multisource[, 2:10],  
  EM = dat_multisource$EM,  
  outcome_model_args = outcome_model_args,  
  treatment_model_args = treatment_model_args,  
  cross_fitting = TRUE,  
  replications = 5)
```

The printed output for the STE estimates is given below.

SUBGROUP TREATMENT EFFECT ESTIMATES IN INTERNAL POPULATIONS

Treatment effect (mean difference) estimates:

Source	Subgroup	Estimate	SE	Lower	95% CI	Upper	95% CI	Lower	95% SCB	Upper	95% SCB
A	a	6.9197	0.5001		5.9395		7.8999		5.6345		8.2050
	b	5.4340	0.3681		4.7126		6.1555		4.4880		6.3801
	c	7.5452	0.3097		6.9383		8.1522		6.7493		8.3411
	d	6.5053	0.3630		5.7939		7.2168		5.5724		7.4382
	e	5.4595	0.5215		4.4373		6.4816		4.1192		6.7998
B	a	8.2134	0.7554		6.7328		9.6939		6.2720		10.1547
	b	6.8396	0.5381		5.7850		7.8942		5.4567		8.2225
	c	8.7995	0.4232		7.9699		9.6290		7.7118		9.8872
	d	7.7098	0.4529		6.8223		8.5974		6.5460		8.8737
	e	6.4405	0.6975		5.0734		7.8076		4.6479		8.2332
C	a	8.0544	1.2049		5.6929		10.4159		4.9579		11.1509
	b	6.2151	0.6711		4.8998		7.5304		4.4904		7.9398
	c	8.2620	0.6426		7.0026		9.5215		6.6106		9.9135
	d	7.1427	0.6538		5.8612		8.4242		5.4624		8.8231
	e	6.0910	0.8718		4.3823		7.7997		3.8504		8.3316

To generate a forest plot of the STEs in the internal target populations, we can run the command plot(result_si) which produces Figure 4. Users can set the argument use_scb to TRUE to illustrate SCBs instead of CIs in the forest plot. Further options for customizing the forest plot are detailed in the package help file for plot.STE_internal.

5. Discussion

The CausalMetaR package implements state-of-science approaches to estimate ATEs and STEs in internal and external target populations from multi-source data. The estimators implemented in the package are (i) doubly robust in the sense that the estimators can still be consistent even if some of the nuisance function estimators are not correctly specified, (ii) efficient in the sense that they attain the smallest possible asymptotic variance under some mild regularity conditions, and (iii) are asymptotically normally distributed which allows for constructing valid confidence intervals. Users of

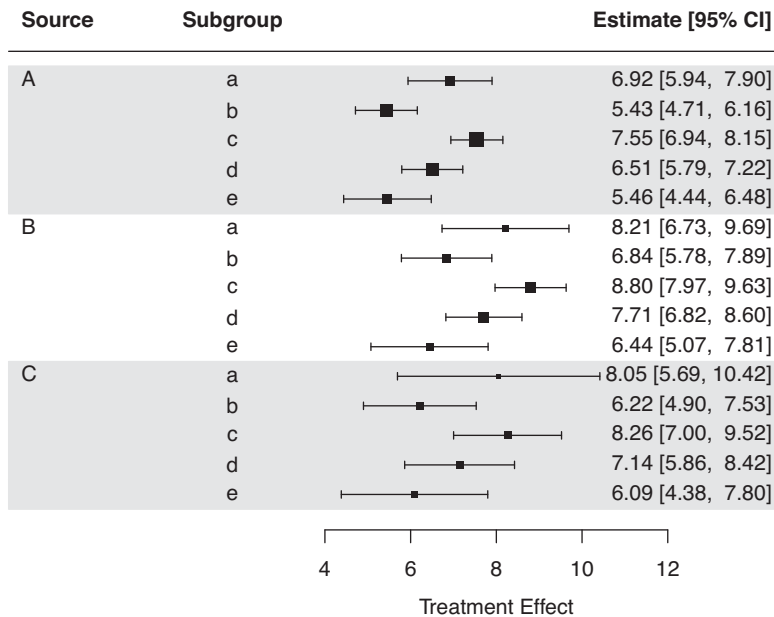


Figure 4. Forest plot of the STEs in the internal target populations.

the package can employ a number of different modelling approaches to estimate the nuisance functions, ranging from parametric approaches such as generalized linear models to highly flexible machine learning approaches such as neural networks. The CausalMetaR package is a complement to R packages that implement conventional meta-analytic approaches (i.e., those that do not estimate causal quantities in well-defined target populations) such as metafor²⁵ and meta.²⁶

The package implements the methods developed by Dahabreh et al.,⁴ Robertson et al.,⁵ and Wang et al.⁷ Such causally-interpretable meta-analyses can be easily adapted to solve many real-world problems, such as estimating causal effects for a study in a meta-analysis and estimating causal effects for a trial when combining data collected from other studies.^{16,27–29} Other approaches for performing causally interpretable meta-analyses have been proposed. Vo et al.^{30,31} developed estimators of average treatment effects in internal source populations for individual participant data (IPD) meta-analyses with case-mix heterogeneity. However, these approaches are not doubly robust and do not consider external target populations. Moreover, IPD network meta-analysis approaches have been developed to estimate a global treatment importance metric and heterogeneous treatment effects in the entire multi-source population.^{32–34} These approaches are not directly comparable to those implemented in CausalMetaR because they do not pertain to the same target populations. When individual participant data of each source are not available to the data analysts due to privacy preservation, federated learning^{35–40} can be used to estimate the causal effects for specific target populations.

In this work, users must designate a categorical effect modifier of interest. This applies to cases where the users have strong a priori knowledge of which variables are the effect modifiers and are interested in certain of them. When the effect modifiers are unknown, one can use data-driven methods to explore the effect modifier and proceed with the analysis^{41–43}; or one can estimate the conditional average treatment effects (conditional on all possible covariates), set subgroups ($\tilde{X}$) by individuals who have similar conditional average treatment effects, and proceed with the analysis.^{3,44} Furthermore, condition A4 is untestable and may be too strong in some scenarios. Such a condition can be replaced by weaker ones, such as the transportability condition of relative effect measures.^{14,15} In addition, when practitioners have prior information about the covariates (such as a variable is an interaction of other main terms), one can integrate this information into the nuisance models. Such integration would potentially improve the estimation accuracy.^{45–48}

The performance of the estimators in this package can be influenced by the sample sizes of the sources. The finite-sample variances of the estimators would be large when the sample sizes of the sources (especially the sources that represents the target population) are small. Though we encourage practitioners to include as many sources as possible to increase the sample size, users should avoid sources that are not transportable to others (such that condition A4 does not hold).

While the current version of CausalMetaR (i.e., version 0.1.2) is stable and complete, we intend to add several features in future versions of the package. In particular, we aim to include methods that estimate STEs in internal and external target populations when the effect modifier ($\tilde{X}$) is continuous.^{49,50} Additionally, we would like to include methods that estimate quantile treatment effects in internal and external target populations.⁵¹ However, methodological work establishing such methods is needed first.

Acknowledgements. We would like to thank Dr. Issa J. Dahabreh for helpful discussions. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the above funding agencies.

Author contributions. All authors have contributed equally to this work. GW, SM, YL: Conceptualization, methodology, investigation, software, validation, visualization, project administration, drafting. All authors approved the final version for publication.

Competing interest statement. The authors declare that no competing interests exist.

Data availability statement. The source code of the software and data presented in this paper are publicly available on GitHub (<https://github.com/ly129/CausalMetaR>).

Funding statement. GW was supported by the Patient-Centered Outcomes Research Institute (PCORI) awards ME-2021C2-22365. S.M. was supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE2140743.

Supplementary material. The supplementary material for this article can be found at <https://doi.org/10.1017/rsm.2025.5>.

Note

- i. Throughout, the inclusion of . . . in the example code denotes the inclusion of all other necessary arguments to complete the call to the function.

References

- [1] Angus DC, Chang C-CH. Heterogeneity of treatment effect: Estimating how the effects of interventions vary across individuals. *JAMA* 2021;326(22): 2312–2313.
- [2] Dahabreh IJ, Kazi DS. Toward personalizing care: assessing heterogeneity of treatment effects in randomized trials. *JAMA* 2023;329(13): 1063–1065.
- [3] Wang G, Heagerty PJ, Dahabreh IJ. Using effect scores to characterize heterogeneity of treatment effects. *JAMA* 2024;331(14): 1225–1226.
- [4] Dahabreh IJ, Petito LC, Robertson SE, Hernán MA, Steingrímsson JA. Towards causally interpretable meta-analysis: transporting inferences from multiple randomized trials to a new target population. *Epidemiol. (Cambridge, Mass.)* 2020;31(3): 334.
- [5] Robertson SE, Steingrímsson JA, Joyce NR, Stuart EA, Dahabreh IJ. Center-specific causal inference with multicenter trials: Reinterpreting trial evidence in the context of each participating center. Preprint, [arXiv:2104.05905](https://arxiv.org/abs/2104.05905), 2021.
- [6] Dahabreh IJ, Robertson SE, Petito LC, Hernán MA, Steingrímsson JA. Efficient and robust methods for causally interpretable meta-analysis: Transporting inferences from multiple randomized trials to a target population. *Biometrics* 2023;79: 1057–1072. <https://doi.org/10.1111/biom.13716>.
- [7] Wang G, Levis A, Steingrímsson J, Dahabreh I. Efficient estimation of subgroup treatment effects using multi-source data. Preprint, [arXiv:2402.02684](https://arxiv.org/abs/2402.02684), 2024. <https://arxiv.org/abs/2402.02684>.
- [8] Lian Y, Wang G, McGrath S, Dahabreh I. CausalMetaR: Causally Interpretable meta-analysis. R package version 0.1.2. 2024. <https://CRAN.R-project.org/package=CausalMetaR>.
- [9] Splawa-Neyman J, Dabrowska DM, Speed TP. On the application of probability theory to agricultural experiments. Essay on principles. Section 9. *Statist. Sci.* 1990;5(4): 465–472.

- [10] Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. *J. Educ. Psychol.* 1974;66(5): 688.
- [11] Robins JM, Greenland S. Causal inference without counterfactuals: Comment. *J. Amer. Stat. Assoc.* 2000;95(450): 431–435.
- [12] Scutari M. Learning Bayesian networks with the bnlearn R package. *J. Stat. Softw.* 2010;35(3): 1–22. <https://doi.org/10.18637/jss.v035.i03>.
- [13] Gao C, Yang S, Shan M, Ye W, Lipkovich I, Faries D. Integrating randomized placebo-controlled trial data with external controls: A semiparametric approach with selective borrowing. Preprint, [arXiv:2306.16642](https://arxiv.org/abs/2306.16642), 2023.
- [14] Dahabreh IJ, Robertson SE, Steingrimsson JA. Learning about treatment effects in a new target population under transportability assumptions for relative effect measures. *Eur. J. Epidemiol.* 2024;39(9): 957–965. <https://doi.org/10.1007/s10654-023-01067-4>.
- [15] Wang G, Levis A, Steingrimsson J, Dahabreh IJ. Causal inference under transportability assumptions for conditional relative effect measures. Preprint, [arXiv:2402.02702](https://arxiv.org/abs/2402.02702), 2024. <https://arxiv.org/abs/2402.02702>.
- [16] Karlsson R, Wang G, Krijthe JH, Dahabreh IJ. Robust integration of external control data in randomized trials. Preprint, [arXiv:2406.17971](https://arxiv.org/abs/2406.17971), 2024.
- [17] Polley E, LeDell E, Kennedy C, Lendle S, van der Laan M. SuperLearner: Super learner prediction. R package version 2.0-28. 2021. <https://CRAN.R-project.org/package=SuperLearner>.
- [18] Chernozhukov V, Chetverikov D, Demirer M, et al. Double/debiased machine learning for treatment and structural parameters. *Econ. J. Jan.* 2018;21(1): C1–C68. <https://doi.org/10.1111/ectj.12097>.
- [19] Zivich PN, Breskin A. Machine learning for causal inference: On the use of cross-fit estimators. *Epidemiol. (Cambridge, Mass.)* 2021;32(3): 393.
- [20] Kennedy C. Guide to superlearner. 2017. Accessed: October 22, 2024. <https://cran.r-project.org/web/packages/SuperLearner/vignettes/Guide-to-SuperLearner.html>.
- [21] Van der Laan MJ, Polley EC, Hubbard AE. Super learner. *Stat. Appl. Genet. Mol. Biol.* 2007;6: 1.
- [22] Friedman J, Tibshirani R, Hastie T. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* 2010;33(1): 1–22. <https://doi.org/10.18637/jss.v033.i01>.
- [23] Kenneth Tay J, Narasimhan B, Hastie T. Elastic net regularization paths for all generalized linear models. *J. Stat. Softw.* 2023;106(1): 1–31. <https://doi.org/10.18637/jss.v106.i01>.
- [24] Venables WN, Ripley BD. *Modern Applied Statistics with S*. 4th ed. New York : Springer; 2002. <https://www.stats.ox.ac.uk/pub/MASS4/>.
- [25] Viechtbauer W. Conducting meta-analyses in R with the metafor package. *J. Stat. Softw.* 2010;36(3): 1–48. <https://doi.org/10.18637/jss.v036.i03>.
- [26] Balduzzi S, Rücker G, Schwarzer G. How to perform a meta-analysis with R: A practical tutorial. *Evid.-Based Mental Health* 2019;22: 153–160.
- [27] Wang G, Poulin-Costello M, Pang H, et al. Evaluating hybrid controls methodology in early-phase oncology trials: A simulation study based on the MORPHEUS-UC trial. *Pharm. Stat.* 2024;23(1): 31–45.
- [28] Ung L, Wang G, Haneuse S, Hernan MA, Dahabreh IJ. Combining an experimental study with external data: study designs and identification strategies. Preprint, 2406.03302 [stat.ME], 2024. <https://arxiv.org/abs/2406.03302v1>.
- [29] Wang G, Liao T-WE, Furfaro D, Anthony Celi L, Sheng-Kai Ma K. Extending inferences from randomized clinical trials to target populations using observational data: A scoping review of transportability methods. Preprint, [arXiv:2402.07236](https://arxiv.org/abs/2402.07236), 2024. <https://arxiv.org/abs/2402.07236>.
- [30] Vo T-T, Porcher R, Chaimani A, Vansteelandt S. A novel approach for identifying and addressing case-mix heterogeneity in individual participant data meta-analysis. *Res. Synth. Methods* 2019;10(4): 582–596.
- [31] Vo T-T, Porcher R, Vansteelandt S. Assessing the impact of case-mix heterogeneity in individual participant data meta-analysis: Novel use of I 2 statistic and prediction interval. *Res. Methods Med. Health Sci.* 2021;2(1): 12–30.
- [32] Wang G, Schnitzer ME, Menzies D, Viiklepp P, Holtz TH, Benedetti A. Estimating treatment importance in multidrug-resistant tuberculosis using targeted learning: An observational individual patient data network meta-analysis. *Biometrics* 2020;76(3): 1007–1016.
- [33] Liu Y, Schnitzer ME, Wang G, et al. Modeling treatment effect modification in multidrug-resistant tuberculosis in an individual patient data meta-analysis. *Stat. Methods Med. Res.* 2022;31(4): 689–705.
- [34] Siddique AA, Schnitzer ME, Bahamyirou A, et al. Causal inference with multiple concurrent medications: A comparison of methods and an application in multidrug-resistant tuberculosis. *Stat. Methods Med. Res.* 2019;28(12): 3534–3549.
- [35] Xiong R, Koenecke A, Powell M, Shen Z, Vogelstein JT, Athey S. Federated causal inference in heterogeneous observational data. *Stat. Med.* 2023;42(24): 4418–4439.
- [36] Han L, Hou J, Cho K, Duan R, Cai T. Federated adaptive causal estimation (face) of target treatment effects. *J. Am. Stat. Assoc.* 2025: 1–25.
- [37] Han L, Shen Z, Zubizarreta J. Multiply robust federated estimation of targeted average treatment effects. *Adv. Neural Inf. Process. Syst.* 2023;36: 70453–70482.
- [38] Guo Z, Li X, Han L, Cai T. Robust inference for federated meta-learning. *J. Am. Stat. Assoc.* 2024: 1–27.
- [39] Vinh Vo T, Lee Y, Hoang TN, Leong T-Y. Bayesian federated estimation of causal effects from observational data. *Uncertain. Artif. Intell.* 2022: 2024–2034.

- [40] Meurisse M, Estupiñán-Romero F, González-Galindo J, et al. Federated causal inference based on real-world observational data sources: Application to a SARS-CoV-2 vaccine effectiveness assessment. *BMC Med. Res. Methodol.* 2023;180(1): 248.
- [41] Bargagli-Stoffi FJ, Cadei R, Lee K, Dominici F. Causal rule ensemble: Interpretable discovery and inference of heterogeneous causal effects. Preprint, [arXiv:2009.09036](https://arxiv.org/abs/2009.09036), 2020.
- [42] Ngufor C, Yao X, Inselman JW, et al. Identifying treatment heterogeneity in atrial fibrillation using a novel causal machine learning method. *Amer. Heart J.* 2023;260: 124–140.
- [43] Jaman A, Wang G, Ertefaie A, et al. Penalized G-estimation for effect modifier selection in the structural nested mean models for repeated outcomes. *Biometrics*, 2025;81(1): ujae165. <https://doi.org/10.1093/biometc/ujae165>.
- [44] Imai K, Li ML. Statistical inference for heterogeneous treatment effects discovered by generic machine learning in randomized experiments. *J. Bus. Econ. Stat.* 2024: 1–29.
- [45] Wang G, Schnitzer M, Chen T, Wang R, Platt RW. A general framework for formulating structured variable selection. *Trans. Mach. Learn. Res.* 2024;43. <https://openreview.net/forum?id=cvOpIhQQMN>.
- [46] Bouchard A, Bourdeau F, Roger J, et al. Predictive factors of detectable viral load in HIV-infected patients. *AIDS Res. Human Retroviruses* 2022;38(7): 552–560.
- [47] Wang G, Perreault S, Platt RW, Wang R, Dorais M, Schnitzer ME. Integrating complex selection rules into the latent overlapping group Lasso for constructing coherent prediction models. Preprint, [arXiv:2206.05337](https://arxiv.org/abs/2206.05337), 2024. <https://arxiv.org/abs/2206.05337>.
- [48] Wang G, Lian Y, Yang AY, et al. Structured learning in time-dependent Cox models. *Stat. Med.* 2024;43(17): 3164–3183. <https://doi.org/10.1002/sim.10116>.
- [49] Künzel SR, Sekhon JS, Bickel PJ, Yu B. Metalearners for estimating heterogeneous treatment effects using machine learning. *Proc. Nat. Acad. Sci.* 2019;116(10): 4156–4165.
- [50] Wang G, Liu S, Yang S. Continuous-time structural failure time model for intermittent treatment. Preprint, [arXiv:2401.15806](https://arxiv.org/abs/2401.15806), 2024. <https://arxiv.org/abs/2401.15806>.
- [51] Firpo S. Efficient semiparametric estimation of quantile treatment effects. *Econometrica* 2007;75(1): 259–276.