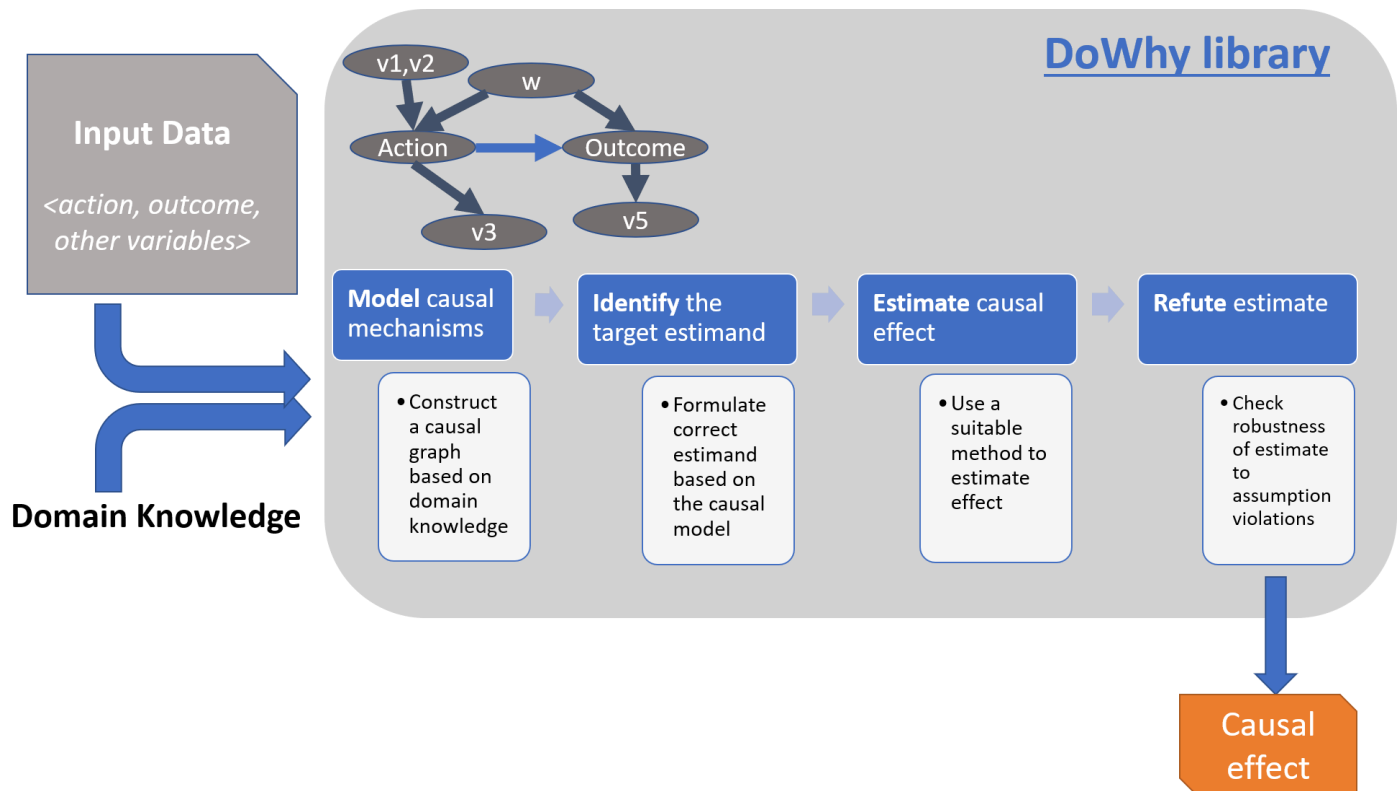


# Estimating Causal Effects



The causal effect of a variable  $A$  on  $Y$  is defined as the expected change in  $Y$  due to a change in  $A$ . Using the do-calculus notation, the average causal effect can be written as,  $E[Y|do(A)]$ . Sometimes, we are interested in the causal effect only on a subpopulation or want to compare the causal effect across sub-populations. In that case, we can estimate the *conditional* average causal effect (CACE) given a set of covariates  $X$ ,  $E[Y|do(A), X]$ .

Estimating the causal effect requires four steps:

1. Model a causal inference problem using assumptions.
2. Identify an expression for the causal effect under these assumptions ("causal estimand").
3. Estimate the expression using statistical methods such as matching or instrumental variables.
4. Finally, verify the validity of the estimate using a variety of robustness checks.

This workflow is captured by four key verbs in DoWhy:

[Skip to main content](#)

- `model` (*CausalModel* or *graph*)
- `identify` (*identify\_effect*)
- `estimate` (*estimate\_effect*)
- `refute` (*refute\_estimate*)

Using these verbs, DoWhy implements a causal effect estimation API that supports a variety of methods. *model* encodes prior knowledge as a formal causal graph, *identify* uses graph-based methods to identify the causal effect, *estimate* uses statistical methods for estimating the identified estimand, and finally *refute* tries to refute the obtained estimate by testing robustness to assumptions. Therefore, after building the causal graph, the next step to estimate causal effect is to identify whether the effect can be estimated from available data. In other words, before considering an estimation algorithm, it is important to determine an identification strategy. DoWhy supports the following identification algorithms:

- Backdoor
- Frontdoor
- Instrumental variable
- ID algorithm

Once a causal effect is identified, we can choose an estimation method compatible with the identification strategy. For estimating the average causal effect, DoWhy supports the following methods.

- **Methods based on matching confounders' values:**
  - Distance-based matching ( `DistanceMatchingEstimator` )
- **Methods based on estimating the treatment assignment**
  - Propensity-based Stratification ( `PropensityScoreStratificationEstimator` )
  - Propensity Score Matching ( `PropensityScoreMatchingEstimator` )
  - Inverse Propensity Weighting ( `PropensityScoreWeightingEstimator` )
- **Methods based on estimating the outcome model**
  - Linear Regression ( `LinearRegressionEstimator` )
  - Generalized Linear Models, including logistic regression ( `GeneralizedLinearModelEstimator` )
- **Methods based on the instrumental variable equation**

[Skip to main content](#)

- Binary Instrument/Wald Estimator ( `InstrumentalVariableEstimator` )
- Regression discontinuity ( `RegressionDiscontinuityEstimator` )
- **Methods for front-door criterion and mediation analysis**
  - Two-stage linear regression ( `TwoStageRegressionEstimator` )

For estimating the conditional average causal effect, DoWhy supports calling EconML methods. For more details on EconML, check out their [documentation](#). If the data-generating process for the outcome Y can be approximated as a linear function, you may also use the linear regression method for CACE estimation.

For related notebooks, see [Example notebooks](#)

[Identifying causal effect](#)

[Estimating average causal effect using backdoor](#)

[Estimating average causal effect with natural experiments](#)

[Estimating conditional average causal effect](#)

[Estimating average causal effect using GCM](#)

< [Previous](#)  
[Performing Causal Tasks](#)

[Identifying causal effect](#) > [Next](#)

© Copyright 2022, PyWhy contributors.

Created using [Sphinx](#) 7.1.2.

Built with the [PyData Sphinx Theme](#) 0.14.4.