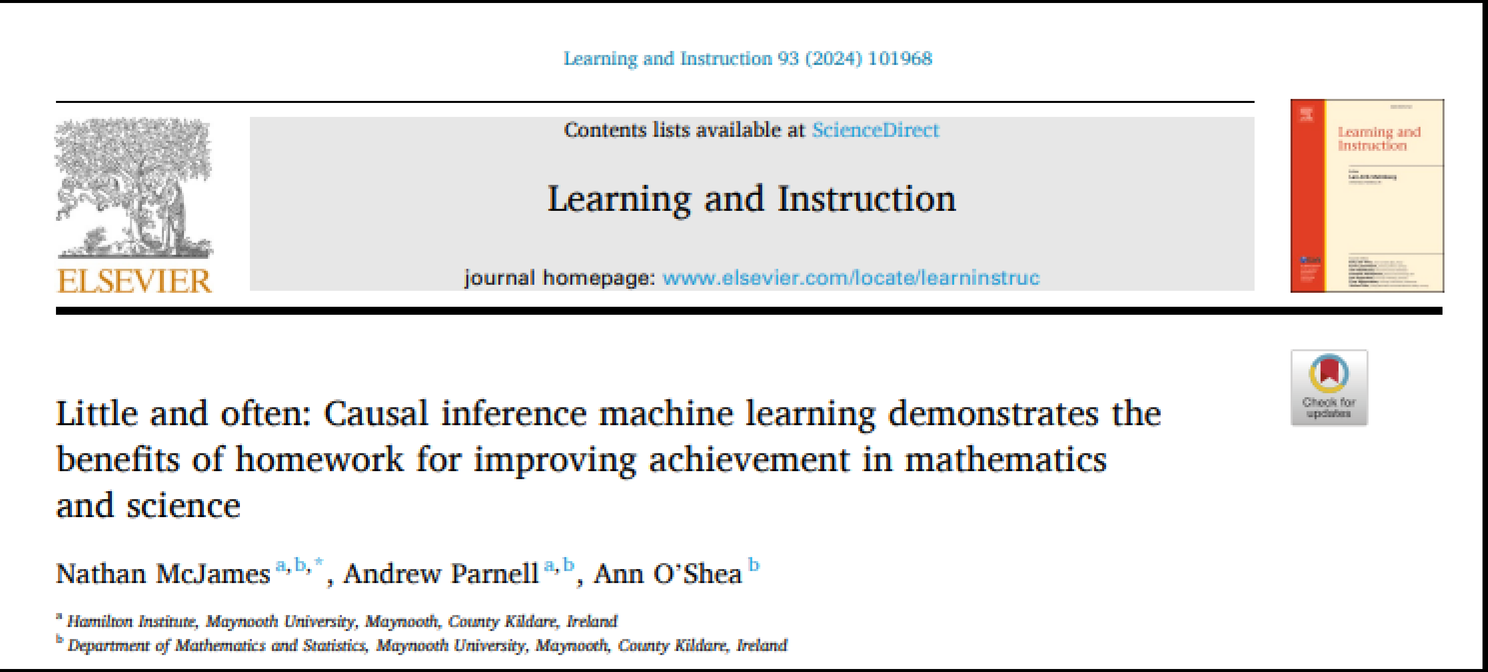
Causal Inference in Educational Research

Peter Edelsbrunner

bit.ly/PeterE_presentations



Learning and Instruction 93 (2024) 101968

Contents lists available at ScienceDirect

Learning and Instruction

journal homepage: www.elsevier.com/locate/learninstruc

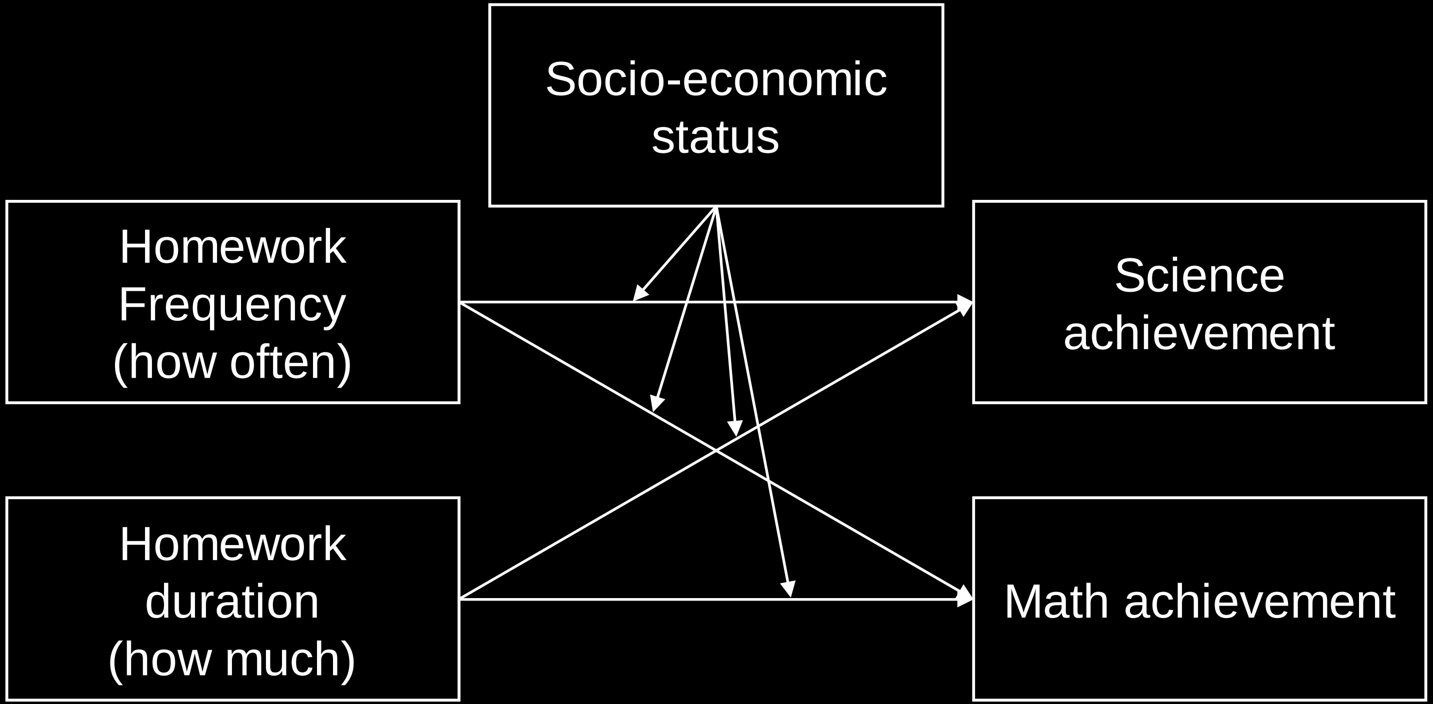
ELSEVIER

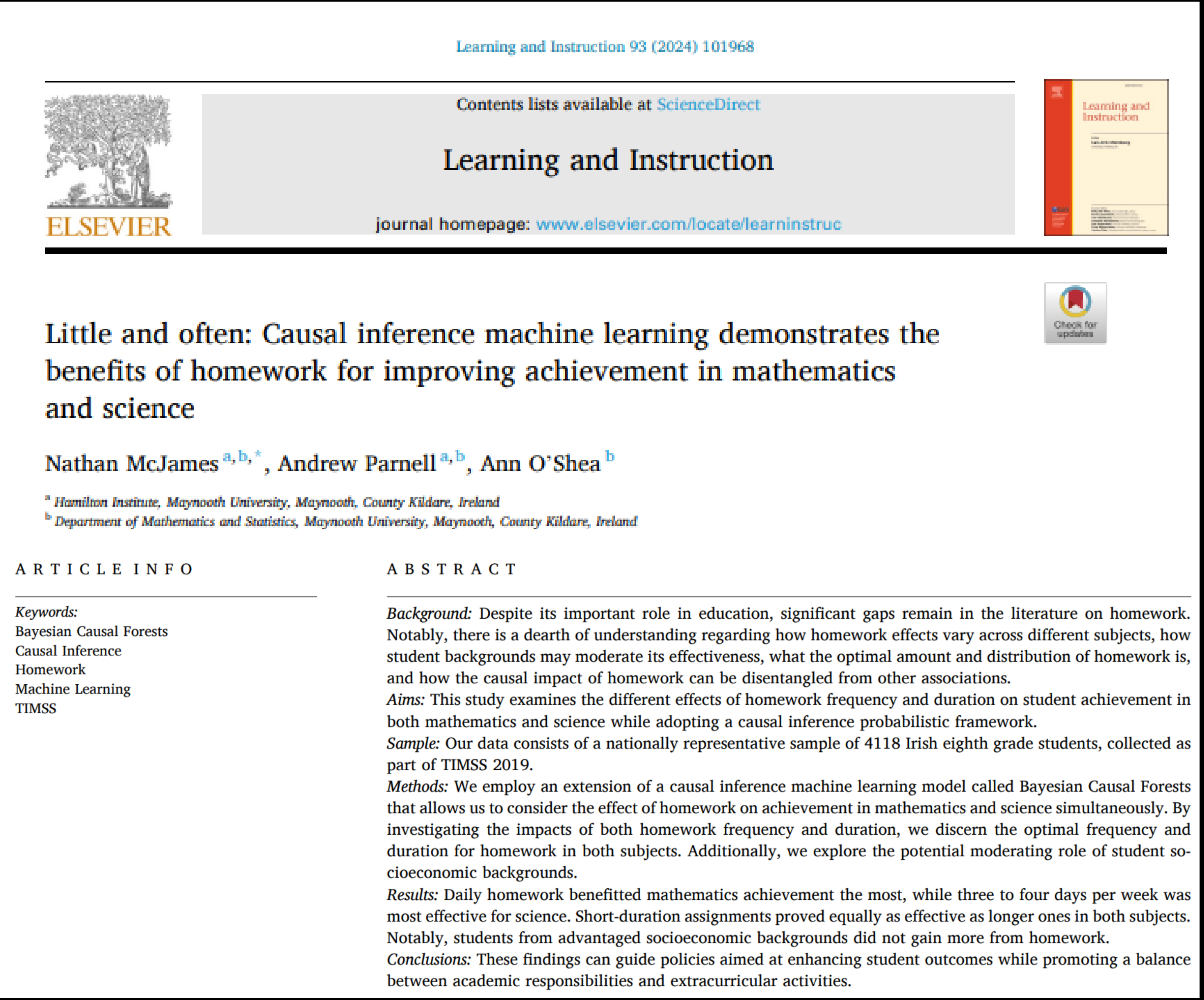
Little and often: Causal inference machine learning demonstrates the benefits of homework for improving achievement in mathematics and science

Nathan McJames[a,b,*], Andrew Parnell[a,b], Ann O'Shea[b]

[a] Hamilton Institute, Maynooth University, Maynooth, County Kildare, Ireland
[b] Department of Mathematics and Statistics, Maynooth University, Maynooth, County Kildare, Ireland

N = 4118 Irish 8th grade students (TIMSS)




Learning and Instruction 93 (2024) 101968

Contents lists available at ScienceDirect

Learning and Instruction

journal homepage: www.elsevier.com/locate/learninstruc

Little and often: Causal inference machine learning demonstrates the benefits of homework for improving achievement in mathematics and science

Nathan McJames [a,b,*], Andrew Parnell [a,b], Ann O'Shea [b]

[a] Hamilton Institute, Maynooth University, Maynooth, County Kildare, Ireland
[b] Department of Mathematics and Statistics, Maynooth University, Maynooth, County Kildare, Ireland

ARTICLE INFO

Keywords:
Bayesian Causal Forests
Causal Inference
Homework
Machine Learning
TIMSS

ABSTRACT

Background: Despite its important role in education, significant gaps remain in the literature on homework. Notably, there is a dearth of understanding regarding how homework effects vary across different subjects, how student backgrounds may moderate its effectiveness, what the optimal amount and distribution of homework is, and how the causal impact of homework can be disentangled from other associations.
Aims: This study examines the different effects of homework frequency and duration on student achievement in both mathematics and science while adopting a causal inference probabilistic framework.
Sample: Our data consists of a nationally representative sample of 4118 Irish eighth grade students, collected as part of TIMSS 2019.
Methods: We employ an extension of a causal inference machine learning model called Bayesian Causal Forests that allows us to consider the effect of homework on achievement in mathematics and science simultaneously. By investigating the impacts of both homework frequency and duration, we discern the optimal frequency and duration for homework in both subjects. Additionally, we explore the potential moderating role of student socioeconomic backgrounds.
Results: Daily homework benefitted mathematics achievement the most, while three to four days per week was most effective for science. Short-duration assignments proved equally as effective as longer ones in both subjects. Notably, students from advantaged socioeconomic backgrounds did not gain more from homework.
Conclusions: These findings can guide policies aimed at enhancing student outcomes while promoting a balance between academic responsibilities and extracurricular activities.

Article:

Little and often: Causal inference machine learning demonstrates the benefits of homework for improving achievement in mathematics and science - ScienceDirect

What do *you* think?

https://forms.gle/k7oGRDHqQbyh88rZ6

Machine learning that „is causal“

Unveils a causal effect

Learning and Instruction 93 (2024) 101968

Contents lists available at ScienceDirect

Learning and Instruction

journal homepage: www.elsevier.com/locate/learninstruc

ELSEVIER

Check for updates

Little and often: Causal inference machine learning demonstrates the benefits of homework for improving achievement in mathematics and science

Nathan McJames[a,b,*], Andrew Parnell[a,b], Ann O'Shea[b]

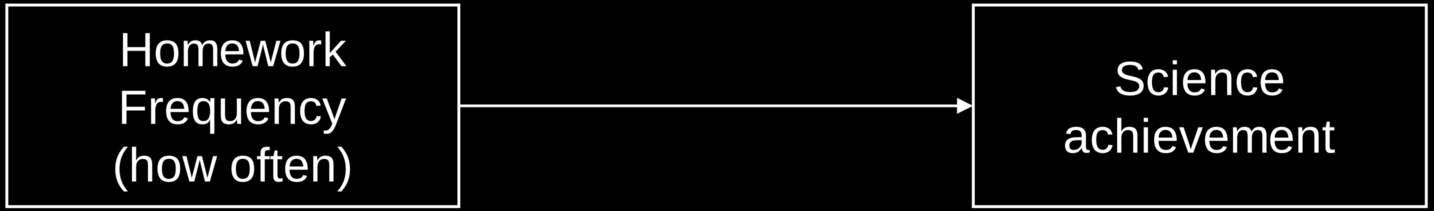
[a] Hamilton Institute, Maynooth University, Maynooth, County Kildare, Ireland
[b] Department of Mathematics and Statistics, Maynooth University, Maynooth, County Kildare, Ireland

A R T I C L E I N F O

Keywords:
Bayesian Causal Forests
Causal Inference
Homework
Machine Learning
TIMSS

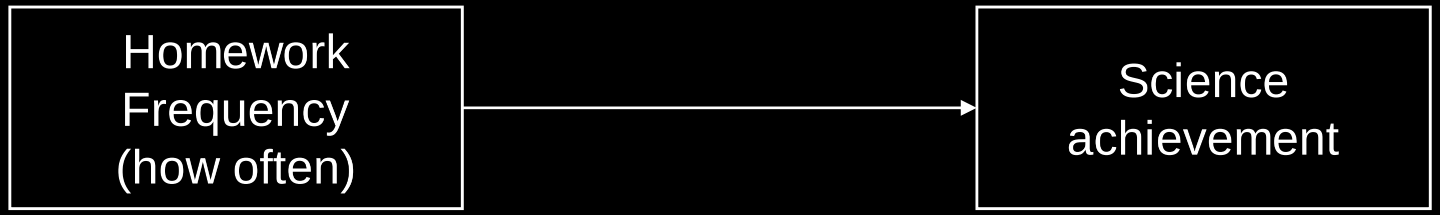
A B S T R A C T

Background: Despite its important role in education, significant gaps remain in the literature on homework. Notably, there is a dearth of understanding regarding how homework effects vary across different subjects, how student backgrounds may moderate its effectiveness, what the optimal amount and distribution of homework is, and how the causal impact of homework can be disentangled from other associations.
Aims: This study examines the different effects of homework frequency and duration on student achievement in both mathematics and science while adopting a causal inference probabilistic framework.
Sample: Our data consists of a nationally representative sample of 4118 Irish eighth grade students, collected as part of TIMSS 2019.
Methods: We employ an extension of a causal inference machine learning model called Bayesian Causal Forests that allows us to consider the effect of homework on achievement in mathematics and science simultaneously. By investigating the impacts of both homework frequency and duration, we discern the optimal frequency and duration for homework in both subjects. Additionally, we explore the potential moderating role of student socioeconomic backgrounds.
Results: Daily homework benefitted mathematics achievement the most, while three to four days per week was most effective for science. Short-duration assignments proved equally as effective as longer ones in both subjects. Notably, students from advantaged socioeconomic backgrounds did not gain more from homework.
Conclusions: These findings can guide policies aimed at enhancing student outcomes while promoting a balance between academic responsibilities and extracurricular activities.



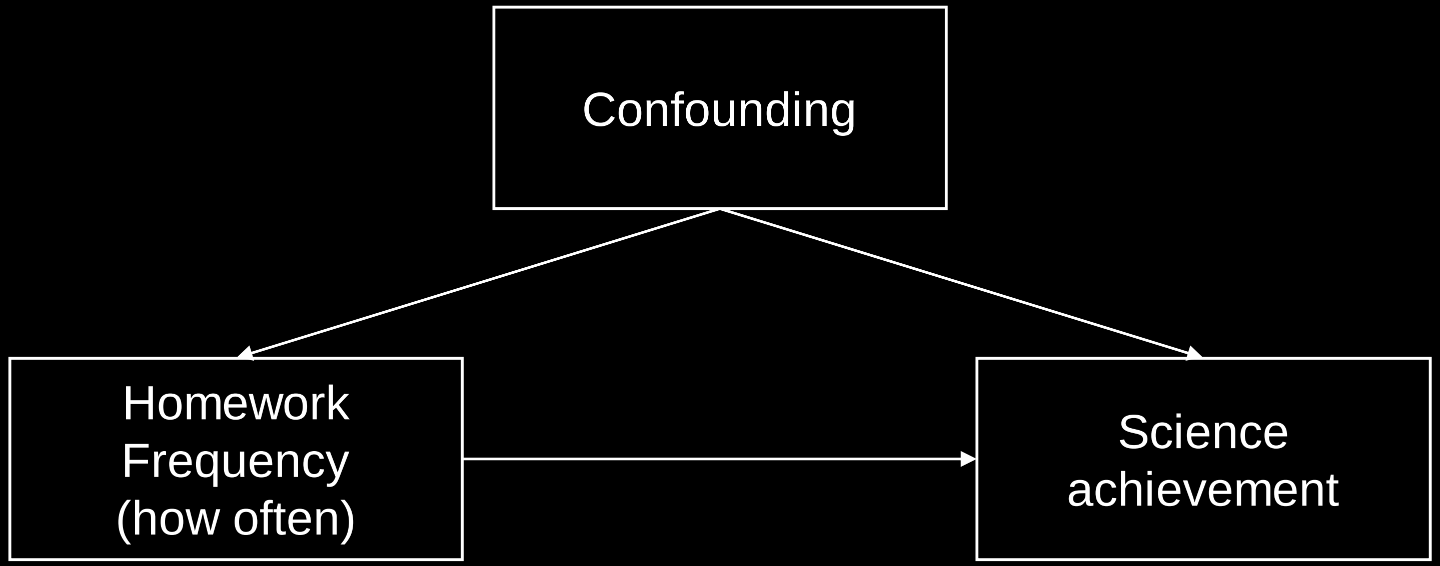
Option 1: Experiment
Random allocation of learners under controlled conditions (i.e., only one thing varies)

Option 2: Observational (i.e., correlational) data



Conditions for causal effect:
1) X must come before Y
2) X and Y must correlate
3) There may be no alternative explanations for the correlation

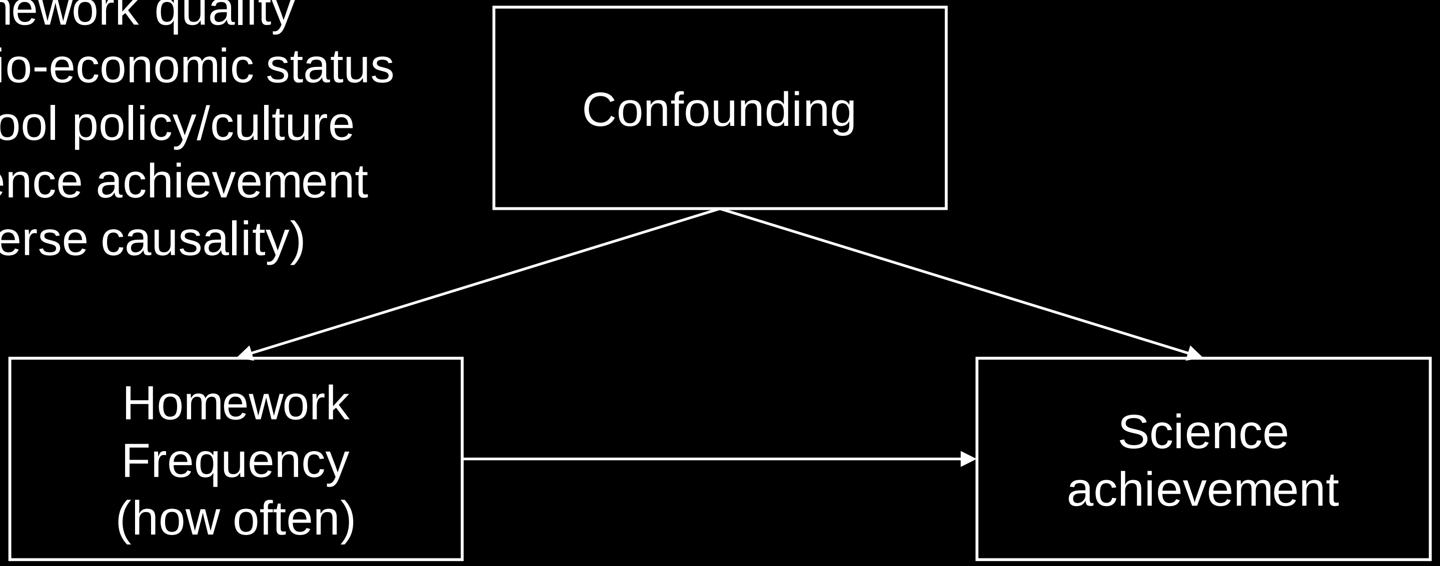
Option 2: Observational (i.e., correlational) data



Unobserved variables that affect both X and Y

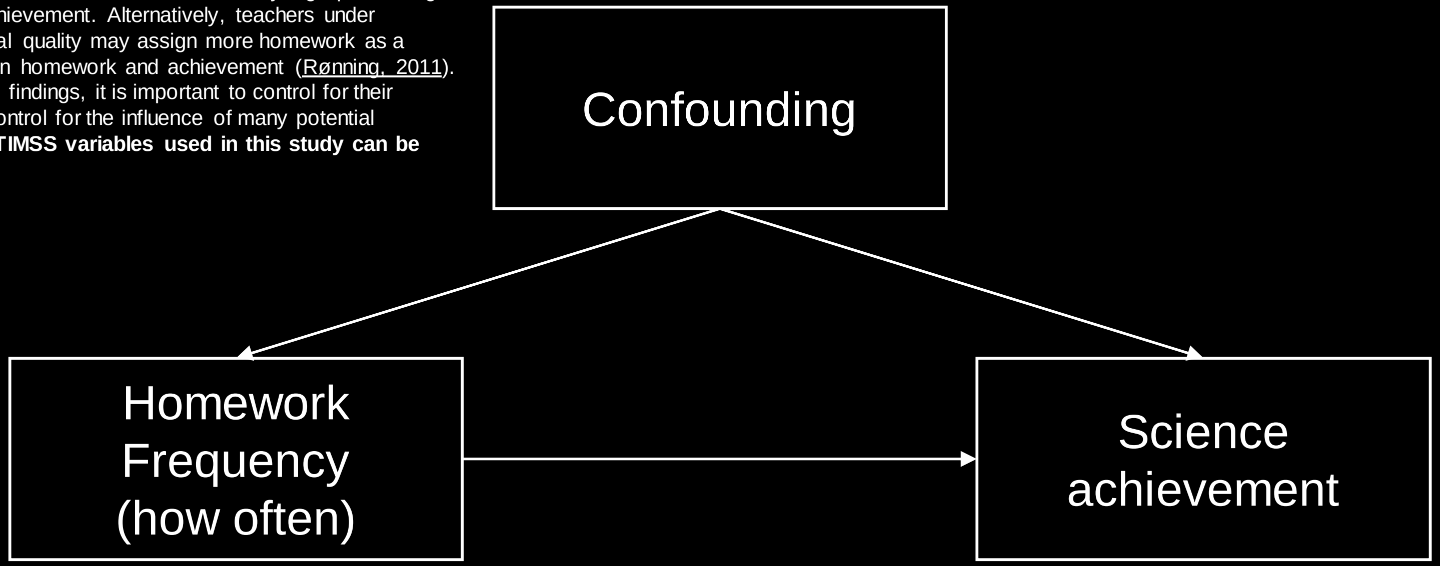
Option 2: Observational (i.e., correlational) data

- Teacher competency
- Teacher enthusiasms
- Homework quality
- Socio-economic status
- School policy/culture
- Science achievement (reverse causality)



Option 2: Observational (i.e., correlational) data

TIMSS collects data on many other important factors such as **socioeconomic status**, **teacher experience**, and **school emphasis on academic success**. These factors can act as confounding variables, simultaneously influencing both the likelihood of receiving treatment and the outcome of interest, potentially biasing results (Greenland et al., 1999). For instance, teachers in academically focused schools with already high performing students may assign more homework, believing it enhances achievement. Alternatively, teachers under pressure to improve student performance due to low instructional quality may assign more homework as a means of compensation, obscuring the true relationship between homework and achievement (Rønning, 2011). Given the ability of confounding variables to lead to unexpected findings, it is important to control for their effects wherever possible. Therefore, we have taken steps to control for the influence of many potential confounding variables from the TIMSS study. A **full list of the TIMSS variables used in this study can be found in the supplementary materials**.

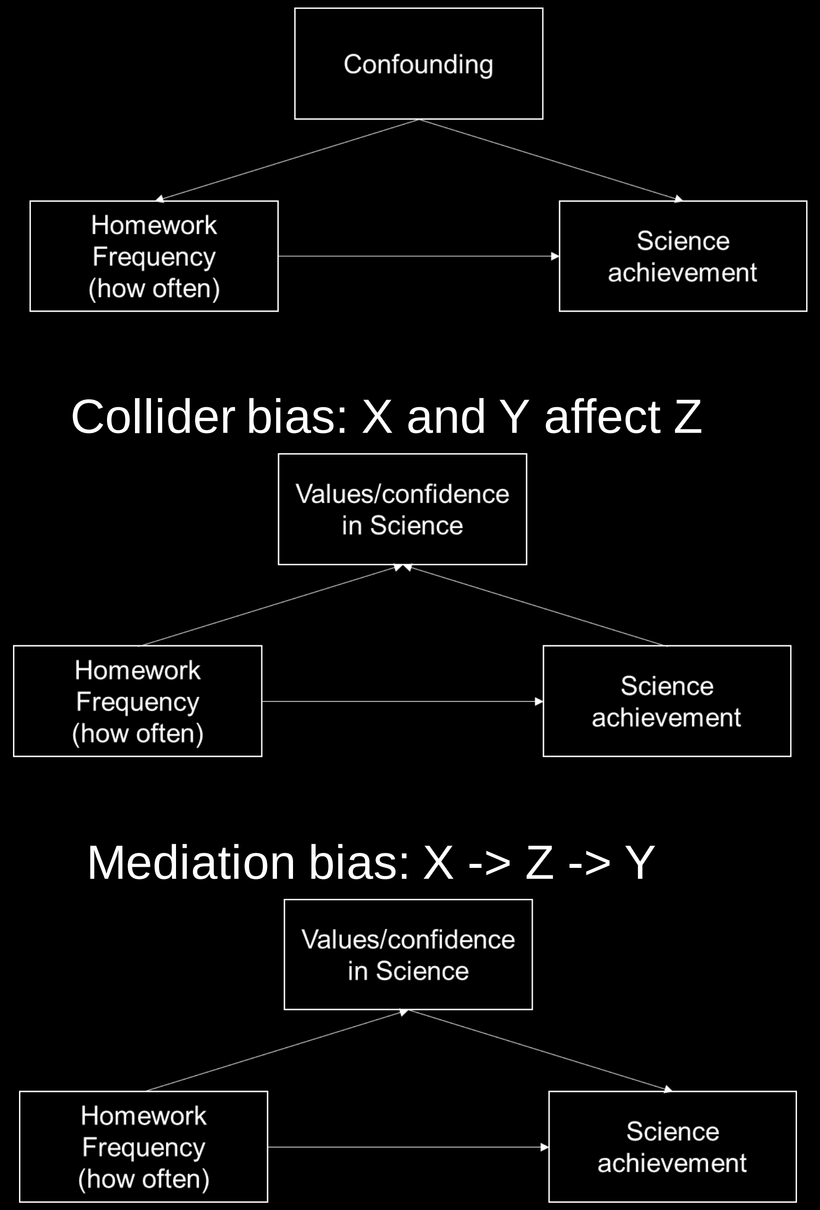


Option 2: Observational (i.e., correlational) data

TIMSS Variables Used		
Variable Code	Obtained From	Description
BSDAGE	Student Questionnaire	Student Age
BSBG01	Student Questionnaire	Student Gender
BSBG03	Student Questionnaire	How often student speaks English at home
BSBG04	Student Questionnaire	Number of books at home
BSBG07	Student Questionnaire	How far in education student expects to go
BSBG08A	Student Questionnaire	Was parent/guardian A born in Ireland
BSBG08B	Student Questionnaire	Was parent/guardian B born in Ireland
BSBG09A	Student Questionnaire	Was student born in Ireland
BSBG10	Student Questionnaire	How often student is absent Collider
BSBG11A	Student Questionnaire	How often student feels hungry when arriving at school
BSBG11B	Student Questionnaire	How often student feels tired when arriving at school
BSDGEDUP	Student Questionnaire	Parent's highest education level
BSBGHER	Student Questionnaire	Number of home educational resources
BSBGSSB	Student Questionnaire	Sense of school belonging Collider
BSBGSB	Student Questionnaire	School bullying
BSBGSCM/BSBGSCS	Student Questionnaire	Confidence in mathematics/science Collider
BSBGSVM/BSBGSVS	Student Questionnaire	Student values mathematics/science Collider
BSBGICM/BSBGICS	Student Questionnaire	Instructional clarity in mathematics/science
BSBG05A	Student Questionnaire	Has computer/tablet at home
BSBG05B	Student Questionnaire	Has study desk at home
BSBG05C	Student Questionnaire	Has own bedroom
BSBG05D	Student Questionnaire	Has home internet connection
BSBG05E	Student Questionnaire	Has own mobile phone
BSBG05F	Student Questionnaire	Has gaming system
BSBG05G	Student Questionnaire	Home TV has "premium" TV channels
BTBG01	Teacher Questionnaire	Number of years teaching Confounding
BTBG02	Teacher Questionnaire	Teacher gender
BTBG03	Teacher Questionnaire	Teacher age
BTBG10	Teacher Questionnaire	Number of students in class
BTBM14/BTBS14	Teacher Questionnaire	Instructional time with class per week
BTBGTJS	Teacher Questionnaire	Teacher job satisfaction
BTBGSOS	Teacher Questionnaire	Safe and orderly school
BTBGLSN	Teacher Questionnaire	Teaching is limited by students not ready for instruction
BTBGEAS	Teacher Questionnaire	Emphasis on academic success
BTDMME	Teacher Questionnaire	Type of degree
BCBGDAS	Principal Questionnaire	School discipline
BCBGEAS	Principal Questionnaire	Emphasis on academic success Self-selection
BCBGMRS/BCBGSRS	Principal Questionnaire	Resource shortages in mathematics/science
BCDGSBC	Principal Questionnaire	School average socioeconomic background

Table A1

Variable codes of potential confounders controlled for as part of the study. All variables listed were used in both the μ and τ parts of the multivariate BCF model.

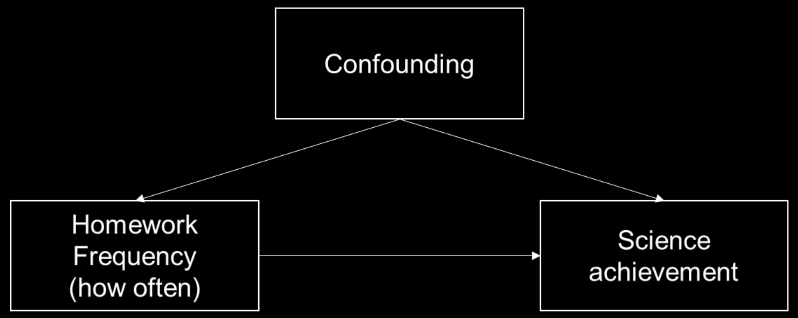


Correct approach for causal inference:

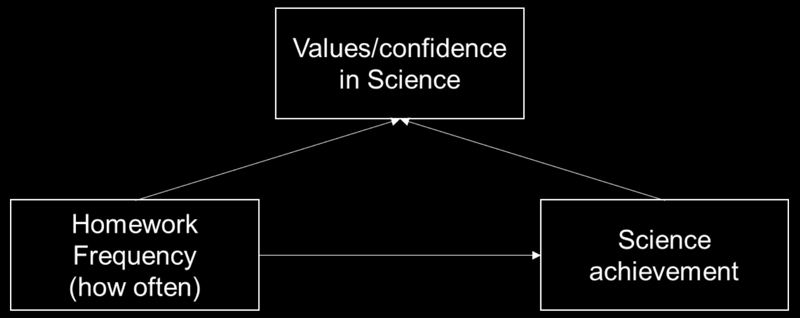
1) Make graphical model

2) Include confounders, exclude collider/mediation variables

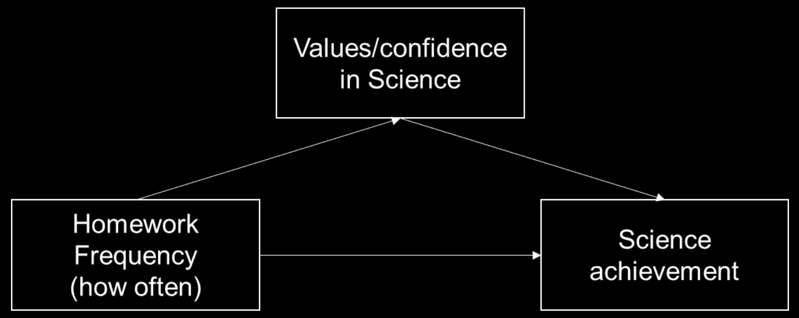
3) Show that the theoretical model appears plausible (i.e., I have all important confounders) and, given specified model, effect of interest is identifiable



Collider bias: X and Y affect Z



Mediation bias: X -> Z -> Y



McJames et al. (2024). *Throw all variables in as confounders*

Supplementary Materials B
Technical Details

Mathematical Description of BART, BCF, and MVBCF

BART

Given an outcome variable y of length n, and a covariate matrix X consisting of n observations of d variables, the BART model (Chipman et al., 2010) can be written as follows:

$$y_i = \sum_{j=1}^{J} g(T_j, M_j, x_i) + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma^2)$$

In the equation above, the function $g()$ calculates the individual contribution of each tree, T_j, out of a total of J trees. The parameters M_j represent the terminal nodes associated with the j-th tree, T_j. The residuals, ϵ_i, are assumed to follow a normal distribution with a mean of 0 and a variance of σ^2. Since the BART model is Bayesian, appropriate priors need to be specified for T_j, M_j, and σ^2.

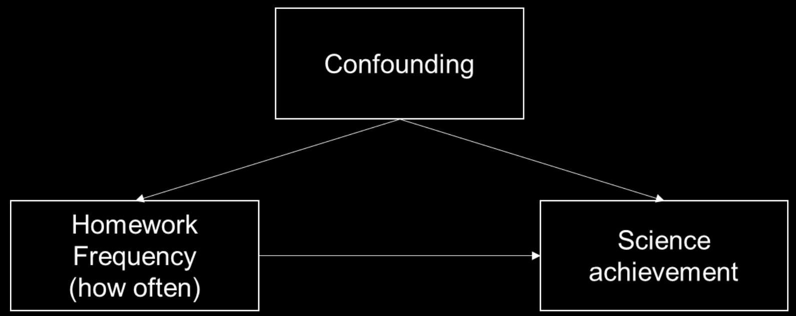
BCF

BCF (Hahn et al., 2020) expresses the outcome y as:

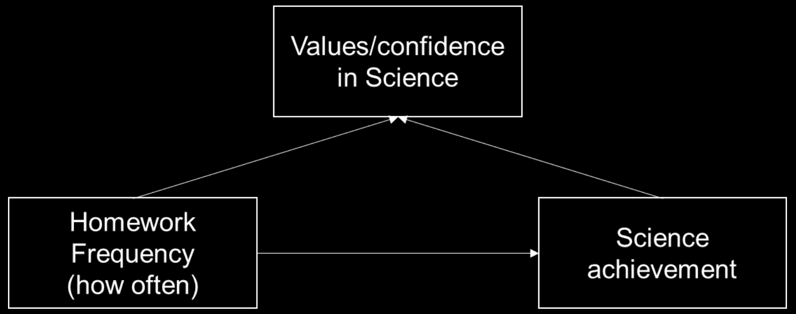
$$y_i = \mu(x_i, \boxed{\hat{\pi}_i}) + \tau(x_i)Z_i + \epsilon_i$$

Propensity score

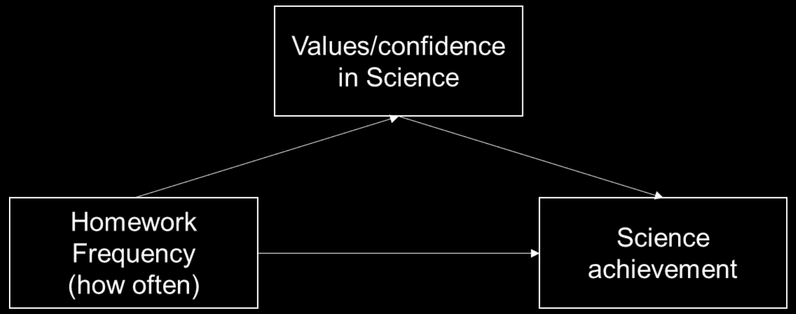
In the equation above, $\mu()$ and $\tau()$ are both BART ensembles that work together to estimate two distinct components of y: a prognostic effect, μ, which represents the expected outcome under control when the treatment variable Z is coded as 1 for treatment and 0 for control, and a treatment effect, τ, which indicates the impact on y resulting from receiving the treatment. The additional covariate, $\hat{\pi}_i$, included in the $\mu()$ part of the model, is called the propensity score. This propensity score, denoted as $\hat{\pi}_i = P(Z_i = 1)$, represents the estimated probability of individual i receiving treatment. The propensity score can be estimated using logistic regression, BART, or any other appropriate classification model capable of providing estimated probabilities.



Collider bias: X and Y affect Z



Mediation bias: X -> Z -> Y



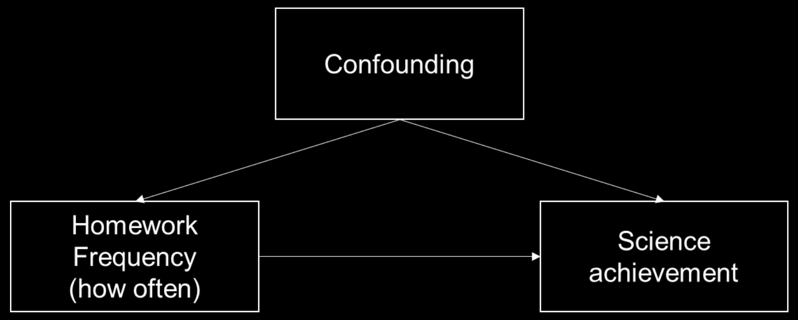
McJames et al. (2024). Propensity score used in BCF

Assumes we know and have measured all variables affecting treatment propensity

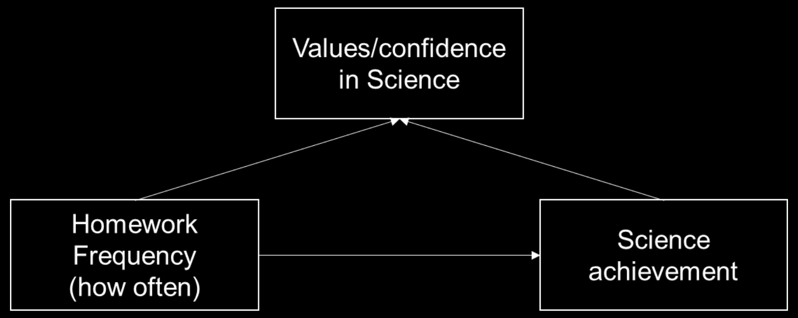
Assumes we have data even from unlikely treatment conditions given covariates

BART (Bayesian additive regression tree):
Relaxes linearity assumptions etc.

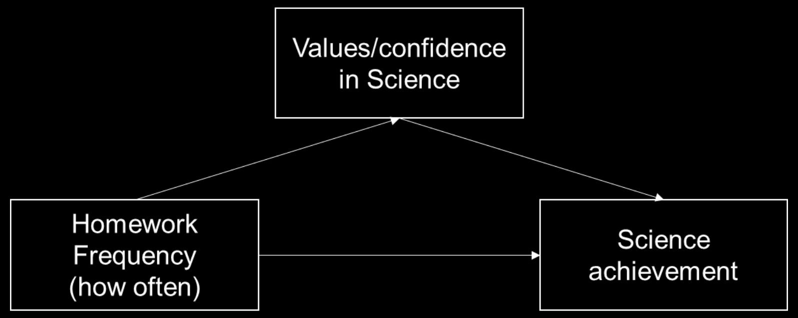
All further assumptions same as in any „causal model“



Collider bias: X and Y affect Z



Mediation bias: X -> Z -> Y



My conclusion:

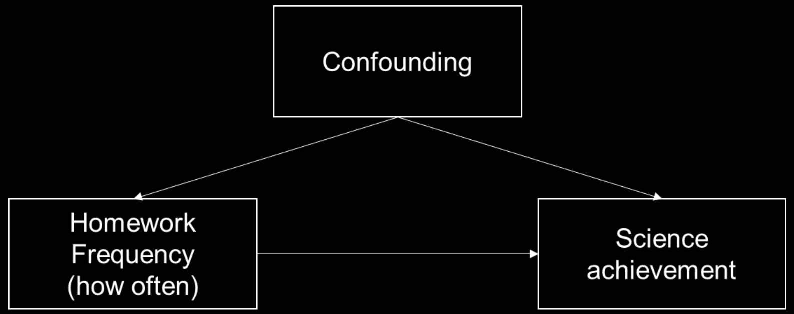
- Causal Machine Learning makes the same assumptions regarding prerequirements for causal inference
    - *Include confounders*
    - *Beware of colliders and mediators*
    - *Make sure to include all relevant confounders*

- Only advantage: Non-parametric/non-linear

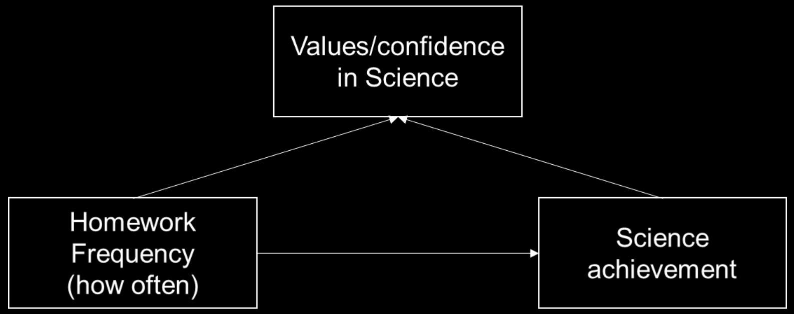
- McJames et al. (2024) is questionable

- „causal“ ML does not „learn causality“

- Literature:
- http://journals.sagepub.com/doi/abs/10.1177/2515245917745629
  Causal inference on human behaviour | Nature Human Behaviour
  [2206.15475] Causal Machine Learning: A Survey and Open Problems



Collider bias: X and Y affect Z



Mediation bias: X -> Z -> Y