Leids Universitair
Medisch Centrum

Counterfactual prediction for personalized healthcare using observational data

STRATOS mini-symposium 2023

Nan van Geloven[1] (TG7 Causal Inference)
Ewout Steyerberg (TG6 Diagnostic tests and prediction models)
Junfeng Wang
Vanessa Didelez (TG7 Causal Inference)
Ruth Keogh (TG4 Measurement error and misclassification + Steering Group)
All other participants Lorentz workshop

[1]Dept of Biomedical Data Sciences, LUMC, Leiden, the Netherlands
n.van_geloven@lumc.nl

Aims

Introduction to `Counterfactual prediction'

Feedback from Lorentz workshop *`Counterfactual prediction for personalized medicine'*

Discuss potential connections to Stratos topic groups

Share and make follow-up plans

Introduction to counterfactual prediction

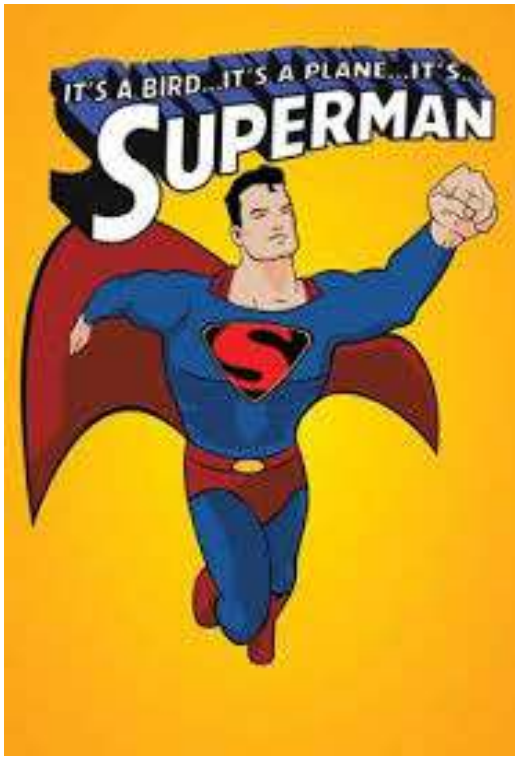
What's in a name?

Counterfactual prediction

Causal prediction

Prediction under hypothetical interventions

Prediction under interventions



Prediction

$E(Y \mid X = x)$ risk of outcome conditional on X

Causal inference

$E(Y^1 - Y^0)$ average treatment effect (ATE)

$E(Y^1 - Y^0 \mid M = m)$ conditional average treatment effect (CATE)

Prediction under interventions

$E(Y^1 \mid V = v)$ risk of outcome conditional on V if treatment would be 1

$E(Y^0 \mid V = v)$ risk of outcome conditional on V if treatment would be 0

Prediction

$E(Y \mid X = x)$ risk of outcome conditional on X

X may include anything: no need to worry about confounding, mediation, colliders etc.

Causal inference

$E(Y^1 - Y^0)$ average treatment effect (ATE)

$E(Y^1 - Y^0 \mid M = m)$ conditional average treatment effect (CATE)

Prediction under interventions

$E(Y^1 \mid V = v)$ risk of outcome conditional on V if treatment would be 1

$E(Y^0 \mid V = v)$ risk of outcome conditional on V if treatment would be 0

Prediction

$E(Y \mid X = x)$ risk of outcome conditional on X

X may include anything: no need to worry about confounding, mediation, colliders etc.

Causal inference

$E(Y^1 - Y^0)$ average treatment effect (ATE)

$E(Y^1 - Y^0 \mid M = m)$ conditional average treatment effect (CATE)

M effect modifiers; need to account for confounding and other potential biases

Prediction under interventions

$E(Y^1 \mid V = v)$ risk of outcome conditional on V if treatment would be 1

$E(Y^0 \mid V = v)$ risk of outcome conditional on V if treatment would be 0

Prediction

$E(Y \mid X = x)$ risk of outcome conditional on X

X may include anything: no need to worry about confounding, mediation, colliders etc.

Causal inference

$E(Y^1 - Y^0)$ average treatment effect (ATE)

$E(Y^1 - Y^0 \mid M = m)$ conditional average treatment effect (CATE)

M effect modifiers; need to account for confounding and other potential biases

Prediction under interventions

$E(Y^1 \mid V = v)$ risk of outcome conditional on V if treatment would be 1

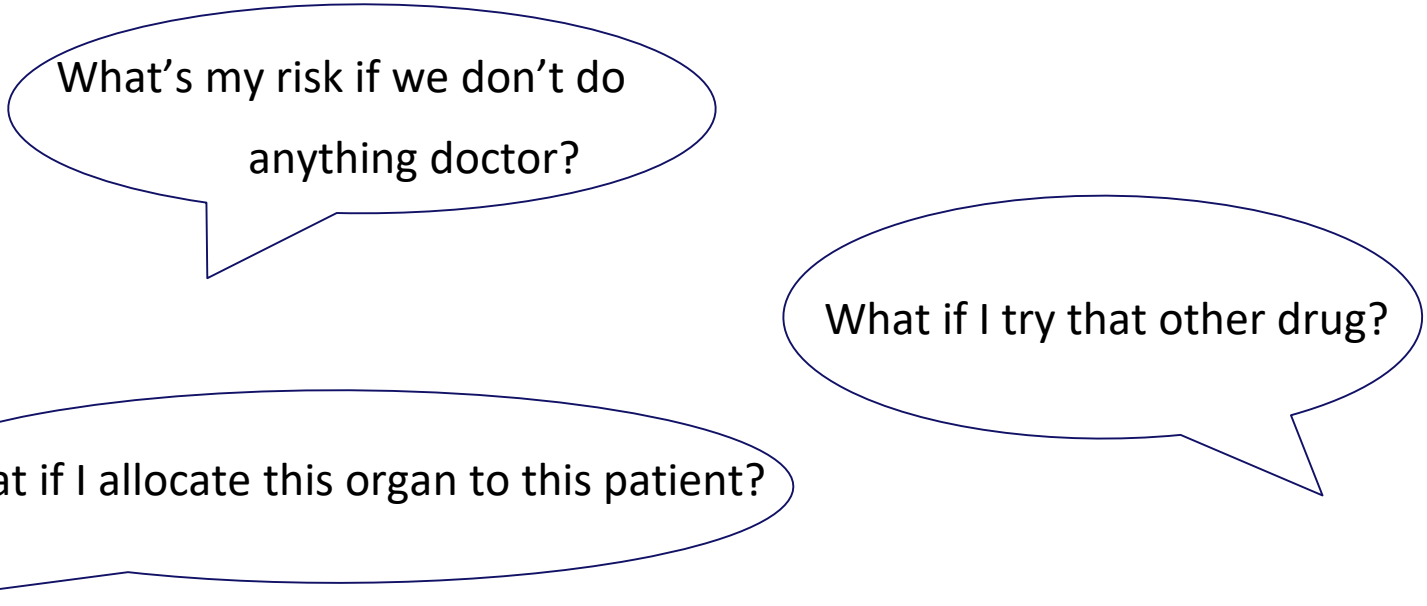
$E(Y^0 \mid V = v)$ risk of outcome conditional on V if treatment would be 0

V may include prognostic factors and effect modifiers; need to account for confounding and other potential biases

What are predictions under interventions for?

Absolute risks under certain treatment choices can inform individual treatment decisions

- individualize risks for a particular patient
- weigh their risks and benefits of different treatment options
- inform allocation of treatments that are subject to resource constraints



Feedback from Lorentz workshop *Counterfactual prediction for personalized medicine*


Lorentz center
Workshop @Snellius
Counterfactual Prediction for Personalized Healthcare
5 - 9 December 2022, Leiden, the Netherlands
Scientific Organizers
Daniala Weir, Utrecht University
Nan van Geloven, Leiden University Medical Centre
Ruth Keogh, London School of Hygiene and Tropical Medicine
Topics
Counterfactual Prediction Algorithms
Performance Metrics for Counterfactual Predictions
Bridging Statistics and AI
Causal Inference, Explainability and Transportability
Clinical Case-Study in Diabetes
The Lorentz Center organizes international workshops for researchers in all scientific disciplines. Its aim is to create an atmosphere that fosters collaborative work, discussions and interactions.
For registration see: www.lorentzcenter.nl
How to analyze data to make individual prognoses that inform optimal treatment decisions
Original photo by Jon Tyson. Poster design: Superflow Studios BV.
Universiteit Leiden
Leiden University Medical Center
Lorentz center
www.lorentzcenter.nl

Participants

London: **Ruth Keogh**, Karla Diaz-Ordaz

Manchester: Matthew Sperrin, Niels Peek

Ghent: Pawel Morzywolek

Bremen: **Vanessa Didelez**

Oslo: Jon Michael Gran

Montreal: Gabrielle Simoneau

Utrecht: Daniala Weir, Timo Brakenhoff, David Liang, Vera Deneer, Thijs van Ommen, Junfeng Wang, Wouter van Amsterdam

Leiden: **Nan van Geloven**, **Hein Putter, Saskia le Cessie, Ewout Steyerberg**, Ilaria Prosepe, Doranne Thomassen

Amsterdam: Sara Magliacane, Giovanni Cina, Joanna Klopotowska, Izak Yasrebi

Delft: Jesse Krijthe, Rickard Karlsson

bold: Stratos members

underlined: organizers

Case study type 2 diabetes

- Patients with multimorbidity[1] & polypharmacy are historically excluded from clinical trials
- High risk of adverse drug events
- Heterogeneous treatment effects on adverse events: response to medications varies between patient subgroups

[1]e.g. additional psychiatric disorders, hypertension, arthritis, kidney disease

Treatment for Type 2 Diabetes

Mr. Koopman is living with type 2 diabetes, as well as hypertension, dyslipidemia and history of pancreatitis. His HbA1c level is 9.5.

Mr. Koopman is using a medication called metformin which is used to lower his blood glucose levels (most common first line therapy)

Adding a second diabetes medication

Dr. Bos explains that his blood glucose levels are still too high and that they should think about starting a second diabetes medication. Four options: SU, DPP-4 i's, GLP-1 RA, SGLT2-I

How should the choice of add-on therapy be individualized for Mr. Koopman?

Case study questions

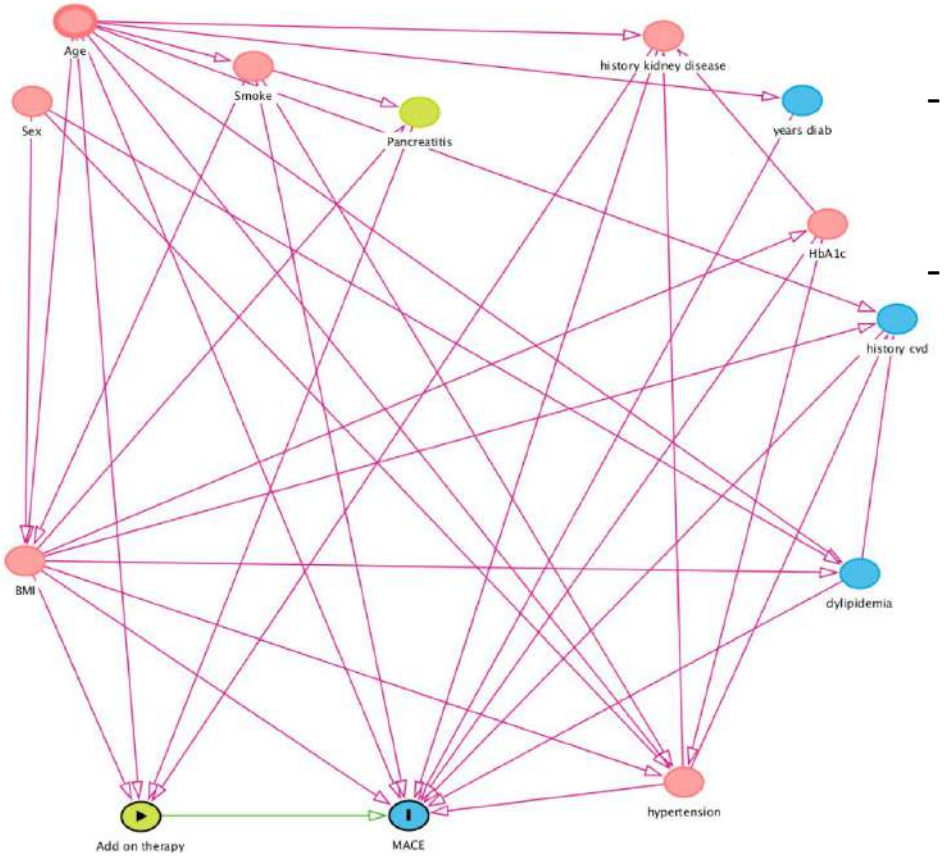
For patients like Mr. Koopman, calculate the absolute risk of 5-year major adverse cardiovascular events when

a) adding no therapy (i.e. 'No add-on') at the index date (time=0)

b) adding a second line agent at the index date (time=0):
GLP1-RA, SU, DPP-4i or SGLT-2

c) adding no treatment during the full 5 years

Synthetic data generation



- Relationships, distributions etc. from literature and from real datasets

- Simulated a synthetic longitudinal dataset including
  - time-fixed and longitudinal covariates
  - time-to-event outcomes
  - both treatment and outcome depend on (longitudinal) covariates

Case study: five working groups

1. Offset method (observational data + treatment effects from trials)
2. Censoring and weighting / MSM
3. G-formula
4. ‘Direct’ doubly robust –learner
5. Counterfactual recurrent network

Public lecture: *`Can predicting the future help us to make better decisions about our health?'*

Lecture to a lay audience in museum Boerhaave (Stratos level 0?)

Focus on whether health calculators / apps give useful advice for lifestyle changes

Introducing confounding and causal inference from observational data along the way

Fun to do!

Computer geeft matige gezondheidsadviezen

Tomas Dooley

Leiden ■ Met kunstmatige intelligentie kun je enorme hoeveelheden data verwerken, maar gezondheidsadviezen geven kun je voorlopig nog niet aan computers overlaten. Dinsdagavond gaat een Lorentzlezing in Museum Boerhaave over dit onderwerp.

„Artificial intelligence (AI) kan veel data verwerken en daar patronen in zien, maar is nog niet voldoende in staat om causale verbanden te leggen", zegt Nan van Geloven, universitair docent biostatistiek aan het Leids Universitair Medisch Centrum. Ze is een van de sprekers tijdens de Lorentzlezing. „Juist die causale verbanden zijn belangrijk als je de juiste medische adviezen wil geven."

Apps over gezondheid bestaan er genoeg, bijvoorbeeld om te waarschuwen tegen hart- en vaatziekten of osteoporose. In dergelijke apps maak je een profiel met daarin zaken zoals je lengte, gewicht, consumptiegewoontes en geslacht. De app gebruikt vervolgens AI om adviezen te geven, zoals: 'Stop nu met roken en verleng je leven met 1,8 procent'.

Te simpels cijfers om complexe gezondheidskwesties te ontvouwen, zegt van Geloven.

Als voorbeeld noemt ze 'gele vingers.' Een AI-algoritme kan duizenden personen naast elkaar leggen en concluderen dat mensen met gele vingers jonger sterven, maar dat betekent niet dat iemand die de vingers in gele verf doopt plotseling korter leeft – ondanks dat AI zou zeggen van wel. De werkelijke doodsoorzaak is namelijk roken, aangezien nicotine gele vingers kan veroorzaken.

„Zelfs als het algoritme zoekgedrag meeneemt, gaat stoppen met roken niet plotseling een leven met exact zoveel procent verlengen. Iemands gezondheid komt voort uit veel meer factoren. Een roker heeft bijvoorbeeld vaker andere ongezonde gewoontes, of is juist meer sociaal actief. Ook die dingen beïnvloeden de gezondheid."

Toekomst

Dit is exact het probleem waar Van Geloven aan werkt, samen met onderzoekers vanuit de hele wereld, waaronder ook andere vakgebieden zoals informatica. „In de toekomst denk ik dat we AI kunnen ontwikkelen die oorzaak-gevolgrelaties beter begrijpt, maar voor nu werken we er nog hard aan. Het einddoel is om AI te hebben die de juiste behandeling weet te adviseren voor iedere persoon."

Nan van Geloven. „Het einddoel is om AI te hebben die de juiste behandeling weet te adviseren voor iedere persoon" MARC DE HAAN

AI legt moeizaam verband tussen oorzaak en gevolg

De Lorentzlezing in Museum Boerhaave begint om 17:30 en is toegankelijk voor iedereen met een museumkaartje. Naast Van Geloven spreken Danilo Welz, universitair docent farmacutische epidemiologie, en Ruth Keogh, univenitair docent biostatistiek. Museum Boerhaave heeft momenteel ook een expositie over AI, genaamd ItrAInpower.

Follow up plans from the Lorentz workshop

1. Place knowledge gaps identified at workshop in ‘Learning Health System’-framework
2. Causal blind spots in risk-based decision making
3. When prediction models become harmful
4. Estimands for sequential prediction under interventions
5. Benchmarking dataset for causal inference
6. Work out estimation methods applied during workshop

PRECOG –reporting guideline for counterfactual prediction

Open access Protocol

BMJ Open

Protocol for the development of a reporting guideline for causal and counterfactual prediction models in biomedicine

Jie Xu ,[1] Yi Guo ,[1] Fei Wang,[2] Hua Xu,[3] Robert Lucero ,[4] Jiang Bian ,[1] Mattia Prosperi[5]

To cite: Xu J, Guo Y, Wang F, *et al*. Protocol for the development of a reporting guideline for causal and counterfactual prediction models in biomedicine. *BMJ Open* 2022;**12**:e059715. doi:10.1136/bmjopen-2021-059715

► Prepublication history for this paper is available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2021-059715).

Received 01 December 2021
Accepted 07 June 2022

ABSTRACT

Introduction While there are guidelines for reporting on observational studies (eg, Strengthening the Reporting of Observational Studies in Epidemiology, Reporting of Studies Conducted Using Observational Routinely Collected Health Data Statement), estimation of causal effects from both observational data and randomised experiments (eg, A Guideline for Reporting Mediation Analyses of Randomised Trials and Observational Studies, Consolidated Standards of Reporting Trials, PATH) and on prediction modelling (eg, Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis), none is purposely made for deriving and validating models from observational data to predict counterfactuals for individuals on one or more possible interventions, on the basis of given (or inferred) causal structures. This paper describes methods and processes that will be used to develop a Reporting Guideline for Causal and Counterfactual Prediction Models (PRECOG).

Methods and analysis PRECOG will be developed following published guidance from the Enhancing the

Strengths and limitations of this study

- ⇒ There are no guidelines for the reporting of data-learnt prediction models that have the specific intent to calculate alternative scenarios (counterfactuals) and identify individualised effects of interventions.
- ⇒ Prediction of Counterfactuals Guideline (PRECOG) will fill a gap in reporting standards for counterfactual prediction modelling and will capitalise on the systematisation and quality of the Enhancing the Quality and Transparency of Health Research network.
- ⇒ PRECOG will be built on diverse (clinical researchers, computer scientists, epidemiologists, statisticians) expertise consensus across multiple development stages.
- ⇒ Even with rigorous study design, execution and reporting standard, causal claims made on observational data analyses might be still mistaken by wrong assumptions or unmeasured, hidden bias.

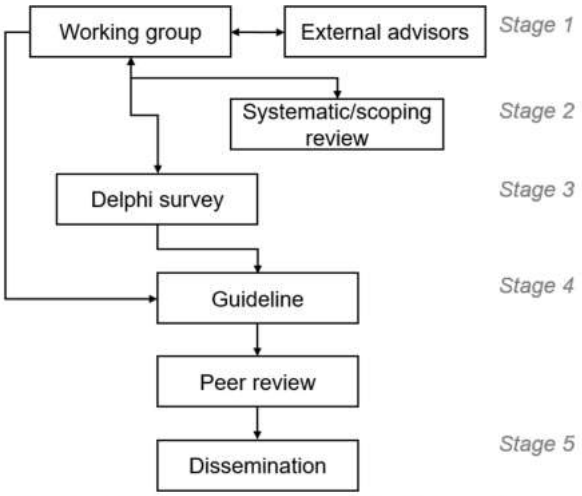


Figure 1 Flow chart of the development of the reporting guideline for causal and counterfactual prediction models.

Discuss potential connections to Stratos topic groups

Evaluating diagnostic tests and prediction models (TG6)

Prediction

- discrimination/calibration/R-squared/…

Causal inference

- typically no data-driven performance assessment
- focus on sensitivity analyses under different assumptions

Prediction under interventions

Initial proposals for assessing predictive performance:

- binary outcomes (Pajouheshnia et al 2017, Coston et al 2020, Boyer 2023)
- time-to-event outcomes (Keogh and van Geloven 2023)

Selection of variables and functional forms in multivariable analysis (TG2)

Prediction

- goal is minimizing prediction error
- penalization / cross-validation / bootstrapping / … (Heinze et al 2018, Sauerbrei et al 2020)

Causal inference

- goal is estimation of causal effect with low bias and high precision
- domain knowledge is key

Prediction under interventions

Mix of the above?

Missing data (TG1)

Prediction

- bias in parameters not a concern
- missing pattern itself could improve prediction (eg, Sperrin et al 2020)
- Ongoing area of research

Causal inference

- aiming for unbiased estimation of causal effects
- causal diagrams support analysis choices (eg, Lee et al. 2021)

Prediction under interventions

Mix of the above?

Measurement error and misclassification (TG4)

Prediction

Predictors: if error used in training set is the same as in deployment setting, ok

Causal inference

In exposure: issue
Confounders: if error seen by historical decision makers setting is the same as in training data, ok

Prediction under interventions

Mix of the above?

Summary and recommendations

Models for predictions under interventions contain a causal part and a non-causal part

This may require mixed strategies for

- performance evaluation
- variable selection
- missing data
- measurement error
- …

Stratos papers should make clear whether advice applies to descriptive, predictive, or causal research

In some areas methodological expansions (fusions) are needed to cater for predictions under interventions

Request for participation in a short survey

Thank you

n.van_geloven@lumc.nl

Obtaining predictions under interventions from data

Individual patient data from RCTs: subgroup analyses / PATH statement – often challenged by limited sample size

Combining observational data with treatment effects from published RCT's, e.g., Predict breast cancer – does not allow treatment heterogeneity

Observational data – challenges in addressing confounding