DeepMind

Functional Causal Bayesian Optimization

UNIVERSITY OF OXFORD

Limor Gultchin

The Alan Turing Institute

Work done with Virginia Aglietti, Alexis Bellot, Silvia Chiappa

uai2023

Functional Causal Bayesian Optimization (fCBO)

What

fCBO is an extension of the causal Bayesian optimization (CBO) method. While the latter considers *hard interventions* optimizing an outcome of interest this new work also considers *soft interventions*

Why

Hard interventions are not always optimal. In this project, we demonstrate why, suggest an alternative method, and provide empirical evidence

Causal Bayesian Optimization (CBO)

Example of causal graph where CBO is applied

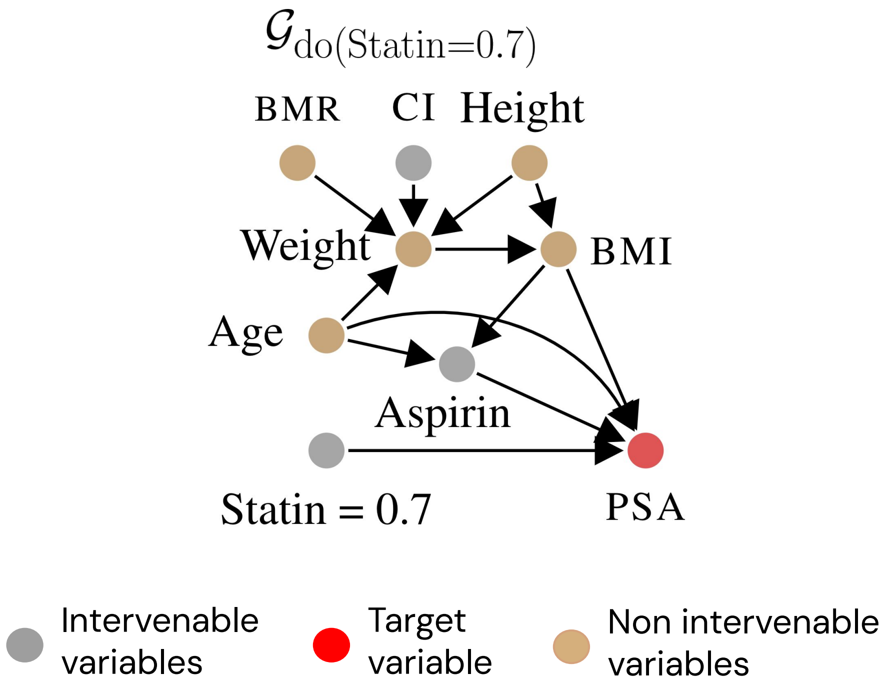
Hard Intervention do($\boldsymbol{X}$=$\boldsymbol{x}$)

Set $\boldsymbol{X}$ to value $\boldsymbol{x}$

Replace existing causal mechanism $p(\boldsymbol{X} \mid pa_{\mathcal{G}}(\boldsymbol{X}))$ with Dirac delta distribution centered at $\boldsymbol{x}$, $\delta_{\boldsymbol{X}}(\boldsymbol{x})$

E.g. do(Statin=0.7)

Replace $p(\text{Statin} \mid \text{Age, BMI})$ with $\delta_{\text{Statin}}(0.7)$



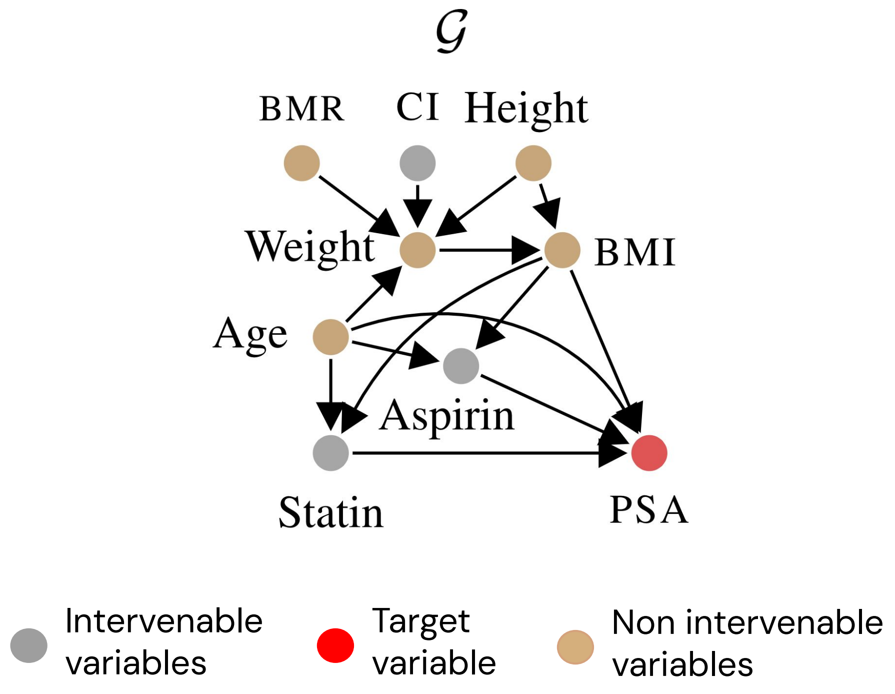
Causal Bayesian Optimization (CBO)

Example of causal graph where CBO is applied

Goal

Find subset $\boldsymbol{X}$ of {CI, Statin, Aspirin} and values $\boldsymbol{x}$ that minimize causal effect on PSA levels (prostate-specific antigen)

$\mathbb{E}[\mathrm{PSA} \mid \mathrm{do}(\boldsymbol{X} = \boldsymbol{x})]$



CBO Problem Formulation

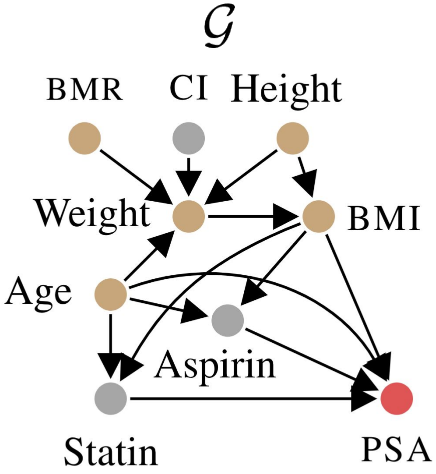
$$\boldsymbol{X}^*, \boldsymbol{x}^* = \arg \min_{\boldsymbol{X} \in \mathcal{P}_{\boldsymbol{I}}, \boldsymbol{x} \in D(\boldsymbol{X})} \mu^Y_{\mathrm{do}(\boldsymbol{X}=\boldsymbol{x})}$$

- $\boldsymbol{X}^*$ → Optimal intervention **set**
- $\boldsymbol{x}^*$ → Optimal intervention **values**
- $\mathcal{P}_{\boldsymbol{I}}$ → **Power set** of intervenable variables
- $D(\boldsymbol{X})$ → **Domain** of possible intervention values
- $\mu^Y_{\mathrm{do}(\boldsymbol{X}=\boldsymbol{x})}$ → **Target effect** $\mathbb{E}[Y|\mathrm{do}(\boldsymbol{X} = \boldsymbol{x})]$

CBO Problem Formulation

$$\boldsymbol{X}^*, \boldsymbol{x}^* = \arg\min_{\boldsymbol{X} \in \mathcal{P}_{\boldsymbol{I}}, \boldsymbol{x} \in D(\boldsymbol{X})} \mu^Y_{\mathrm{do}(\boldsymbol{X}=\boldsymbol{x})}$$

- $\boldsymbol{X}^*$: Optimal intervention **set**
- $\boldsymbol{x}^*$: Optimal intervention **values**
- $\mathcal{P}_{\boldsymbol{I}}$: **Power set** of intervenable variables
- $D(\boldsymbol{X})$: **Domain** of possible intervention values
- $\mu^Y_{\mathrm{do}(\boldsymbol{X}=\boldsymbol{x})}$: **Target effect** $\mathbb{E}[Y|\mathrm{do}(\boldsymbol{X} = \boldsymbol{x})]$



❖ Target variable Y = PSA

❖ Intervenable variables $\boldsymbol{I}$ = {CI, Statin, Aspirin}

❖ $\mathcal{P}_{\boldsymbol{I}}$ = {∅ ,{CI}, {Statin}, {Aspirin}, {CI, Statin}, {CI, Aspirin}, {Statin, Aspirin}, {CI, Statin, Aspirin}}

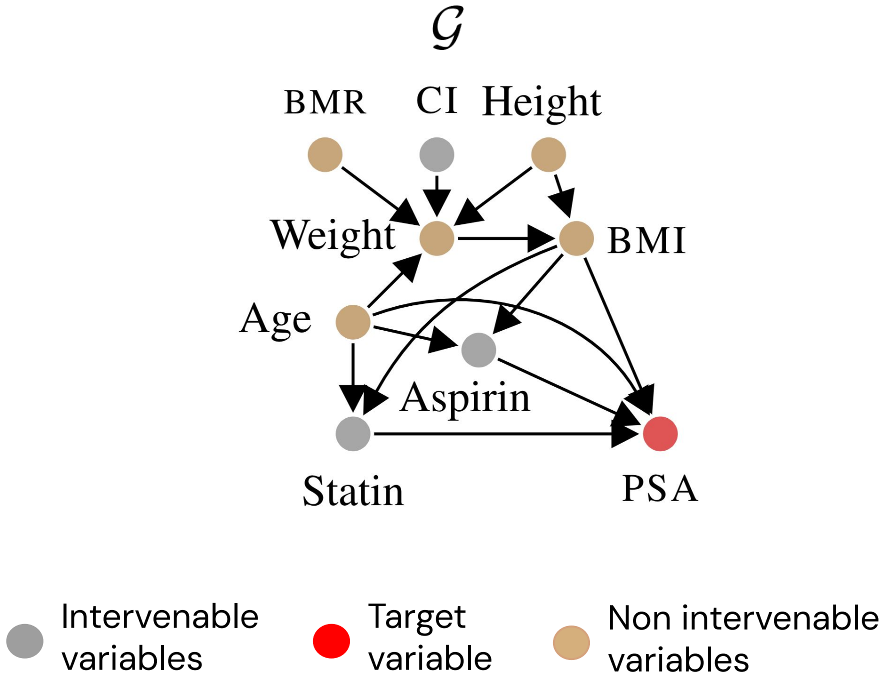
E.g. $\boldsymbol{X}^*$ = {CI, Statin, Aspirin}
$\boldsymbol{x}^*$= (CI=1, Statin=1, Aspirin=0)

Why only Hard Interventions?

Soft Intervention

Often the decision maker has the ability to perform a **conditional/contextual** replacement of the existing causal mechanism, i.e. replace $p(\boldsymbol{X} \mid pa_{\mathcal{G}}(\boldsymbol{X}))$ with another conditional distribution $\pi_{\boldsymbol{X} \mid \boldsymbol{C}_X}$

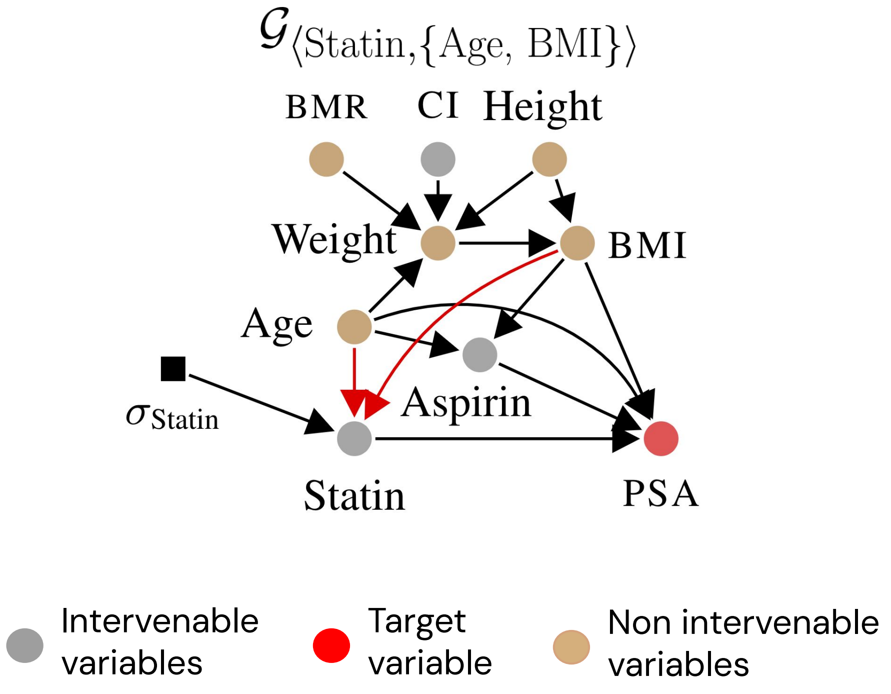
new parents called contexts



Why only Hard Interventions?

Soft Intervention

Often the decision maker has the ability to perform a **conditional/contextual** replacement of the existing causal mechanism, i.e. replace $p(\boldsymbol{X} \mid pa_{\mathcal{G}}(\boldsymbol{X}))$ with another conditional distribution $\pi_{\boldsymbol{X} \mid \boldsymbol{C}_X}$



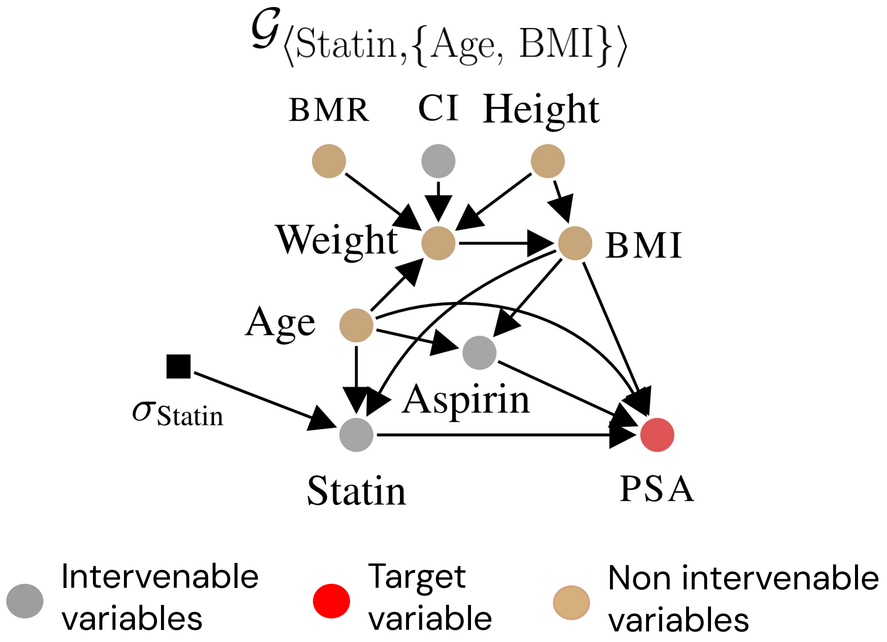
E.g.

When finding an optimal value for Statin, we would likely want to take Age and BMI levels into account, as those hold information about the outcome node

Replace $p(\text{Statin} \mid \text{Age, BMI})$ with $\pi_{\text{Statin} \mid \text{Age,BMI}}$

Functional Causal Bayesian Optimization (fCBO)

- Targeted, more personalized treatment
- Subgroup optimality
- Lower cost treatments, focusing on most needed/promising contexts
- Hard interventions are special cases of soft, so no loss by considering soft



CBO and fCBO Problem Formulation

CBO

$$\boldsymbol{X}^*, \boldsymbol{x}^* = \arg \min_{\boldsymbol{X} \in \mathcal{P}_{\boldsymbol{I}}, \boldsymbol{x} \in D(\boldsymbol{X})} \mu^{Y}_{\mathrm{do}(\boldsymbol{X}=\boldsymbol{x})}$$

- $\boldsymbol{X}^*$: Optimal intervention **set**
- $\boldsymbol{x}^*$: Optimal intervention **values**
- $\mathcal{P}_{\boldsymbol{I}}$: **Power set**
- $D(\boldsymbol{X})$: **Domain**
- $\mu^{Y}_{\mathrm{do}(\boldsymbol{X}=\boldsymbol{x})}$: **Target effect** $\mathbb{E}[Y|\mathrm{do}(\boldsymbol{X} = \boldsymbol{x})]$

fCBO

$$\mathcal{S}^*, \pi^*_{\mathcal{S}^*} = \arg \min_{\mathcal{S} \in \Sigma, \pi_{\mathcal{S}} \in \Pi_{\mathcal{S}}} \mu^{Y}_{\pi_{\mathcal{S}}}$$

- $\mathcal{S}^*$: Optimal **mixed policy scope (MPS)**
- $\pi^*_{\mathcal{S}^*}$: Optimal **MPS realization** (collection of functions)
- Σ: Set of **all MPSs**
- $\pi_{\mathcal{S}}$: Set of **all possible MPS realizations** (interventions)
- $\mu^{Y}_{\pi_{\mathcal{S}}}$: **Target effect** $\mathbb{E}_s[Y]$

Mixed Policy Scope

$$\mathcal{S}^*, \pi^*_{\mathcal{S}^*} = \arg \min_{\mathcal{S} \in \Sigma, \pi_{\mathcal{S}} \in \Pi_{\mathcal{S}}} \mu^Y_{\pi_{\mathcal{S}}}$$

- $\mathcal{S}^*$: Optimal **mixed policy scope** (**MPS**)
- $\pi^*_{\mathcal{S}^*}$: Optimal **MPS realization** (collection of functions)
- Σ: Set of **all MPSs**
- $\pi_{\mathcal{S}}$: Set of **all possible MPS realizations** (interventions)

Need to reason about possible interventions variables and associated contexts.
Requires defining a space made of different <intervention variables, contexts> pairs

Mixed Policy Scope (MPS) $\mathcal{S}$

Collection of tuples $\langle X, \mathbf{C}_X \rangle$ where

- X is an intervenable node $X \in \boldsymbol{I}$
- $\mathbf{C}_X$ is associated set of contexts for intervention $\pi_{X \mid \mathbf{C}_X}$
- $\langle X, \mathbf{C}_X \rangle$ do not introduce cycles in the graph

➔ Characterizing optimal mixed policies: Where to intervene and what to observe. S. Lee, E. Bareinboim, 2020.

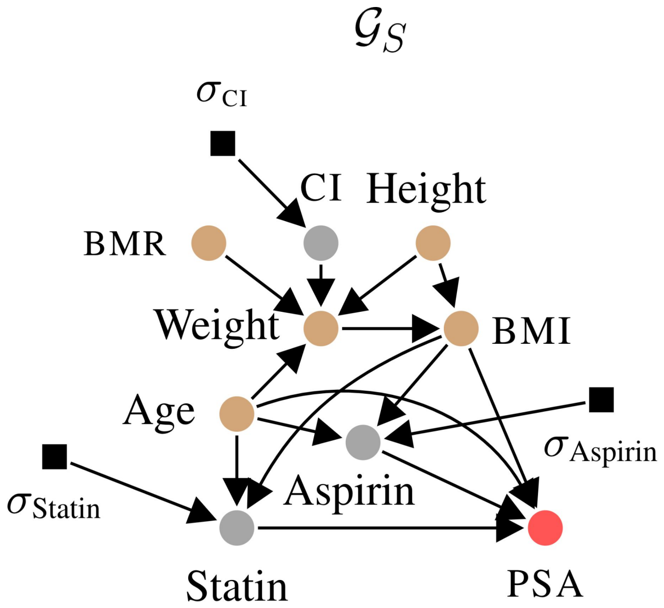
Mixed Policy Scope for the Healthcare Example

Example of MPS

$$\mathcal{S} = \{\langle \text{Statin}, \{\text{Age, BMI}\}\rangle, \langle \text{Aspirin}, \{\text{Age, BMI}\}\rangle, \langle \text{CI}, \emptyset\rangle\}$$

Possible Instantiation of MPS

$$\pi_{\mathcal{S}} = \{\pi_{\text{Statin} \,|\, \text{Age,BMI}}, \pi_{\text{Aspirin} \,|\, \text{Age,BMI}}, \pi_{\text{CI}}\}$$



$$\pi_{\text{Statin|Age, BMI}} = \delta_{\text{Statin}}(\alpha * \text{Age} + \beta * \text{BMI})$$

$$\pi_{\text{CI}} = \delta_{\text{CI}}(x) \text{ (i.e. } \text{do}(\text{CI} = x))$$

$\pi_{\text{Statin|Age, BMI}}$ can also be stochastic, e.g. $\mathcal{N}(\text{Age} + \text{BMI}, \sigma)$, but in this work we focus on deterministic soft interventions

Main Contributions & Challenges

1

Conceptual/ Theoretical Development

Intuitively, **soft interventions should matter**

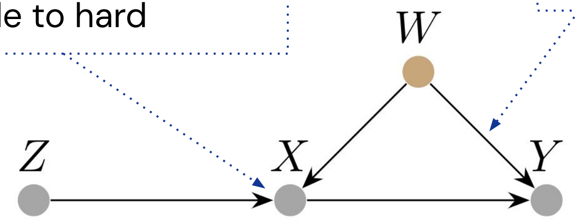
But we needed to show **how and when**

When do Soft Interventions Matter

$1_{\text{Conceptual}}$

A soft intervention on X preferable to hard

Context to target open path



Optimality of Soft Interventions

- We show theoretically conditions under which hard interventions are suboptimal compared to soft

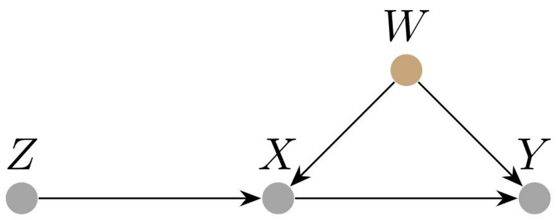
Eq. (1)

$$\mathcal{S}^*, \pi^*_{\mathcal{S}^*} = \arg \min_{\mathcal{S} \in \Sigma, \pi_{\mathcal{S}} \in \Pi_{\mathcal{S}}} \mu^Y_{\pi_{\mathcal{S}}}$$

Proposition 3.2 (Sub-optimality of hard interventions). *Let $\mathcal{G}$ satisfy condition (i) $\exists C \in pa_{\mathcal{G}}(Y)$ with $C \notin \boldsymbol{I}$; or (ii) $\exists C \in sp_{\mathcal{G}}(Y)$. If there exists $X \in an_{\mathcal{G}}(Y) \cap \boldsymbol{I}$ such that $\{\langle X, C \rangle\}$ is an* MPS*, then there exists at least one* SCM *compatible with $\mathcal{G}$ for which restricting the search space in f*CGO *from Σ to Σ_{hard} would lead to a higher target effect.*

When do Soft Interventions Matter

1 Conceptual



Optimality of Soft Interventions for conditional target effect

- We show that performing hard interventions, which are constant across the population, may lead to suboptimal conditional target effect when optimizing for the overall target effect

- A soft intervention that can take subpopulation into account can avoid such suboptimality

Eq. (1)

$$\mathcal{S}^*, \pi^*_{\mathcal{S}^*} = \arg \min_{\mathcal{S} \in \Sigma, \pi_{\mathcal{S}} \in \Pi_{\mathcal{S}}} \mu^Y_{\pi_{\mathcal{S}}}$$

Proposition 3.4 (Optimizing conditional target effects). *If* $\mathcal{S}^*, \pi^*_{\mathcal{S}^*} = \arg\min_{\mathcal{S} \in \Sigma, \pi_{\mathcal{S}} \in \Pi_{\mathcal{S}}} \mu^Y_{\pi_{\mathcal{S}}}$ *then* $\mathcal{S}^*, \pi^*_{\mathcal{S}^*} = \arg\min_{\mathcal{S} \in \Sigma, \pi_{\mathcal{S}} \in \Pi_{\mathcal{S}}} \mu^Y_{\pi_{\mathcal{S}}, \boldsymbol{C}=\boldsymbol{c}}$ $\forall \boldsymbol{C} \subset \boldsymbol{V} \setminus Y$ *such that* $\boldsymbol{C} \cap de_{\mathcal{G}}(\boldsymbol{I}) = \emptyset$ *and* $\forall \boldsymbol{c} \in \mathcal{R}_{\boldsymbol{C}}$.

Main Contributions & Challenges

1

Conceptual/ Theoretical

Intuitively, **soft interventions should matter**

But we needed to show **how and when**

2

Technical

Optimizing soft interventions requires **new methodology and code**, including a **functional GP approach**, and a **kernel measuring distance between soft, and possible hard/soft mixed interventions**

CBO

Search Space

Reduce the search space by leveraging invariances of the target effects

From rule 3 of do-calculus (action deletion)

GP Surrogate Models

Model each target effect using a Gaussian process (GP)

Acquisition Function

Acquisition function that accounts for all target effects

- ➔ Structural Causal Bandits Where to Intervene?. S. Lee, E. Bareinboim, 2018.
- ➔ Causal Bayesian Optimisation. V. Aglietti, X. Lu, A. Paleyes, J. Gonzalez 2020.

fCBO

2 Technical

Search Space

Reduce the search space by leveraging invariances of the target effect w.r.t intervention nodes and contexts

do-calculus (+rule 2) applied to MPSs

GP Surrogate Models

Model each MPS target effect using a functional Gaussian process (GP)

Computing distances across MPS interventions requires a specialized kernel construction

Acquisition Functional

Acquisition functional that accounts for all MPS effects

- ➔ Characterizing optimal mixed policies: Where to intervene and what to observe. S. Lee, and E. Bareinboim, 2020.
- ➔ Bayesian Functional Optimization. N. A. Vien, H. Zimmermann, M. Toussaint, 2018.

Functional Causal Bayesian Optimization: GP Construction

2 Technical

$$g_{\mathcal{S}}(\pi) \sim \mathcal{GP}(m_{\mathcal{S}}(\pi), K^{\theta}_{\mathcal{S}}(\pi, \pi'))$$

Surrogate model(s) over $\mu^{Y}_{\mathcal{S}}$

Target effect under possible interventions on MPS $\mathcal{S}$, $\mathbb{E}_{\mathcal{S}=\cdot}[Y]$

- $m_{\mathcal{S}}(\pi)$ Prior mean functional, initialized at 0
- $K^{\theta}_{\mathcal{S}}(\pi, \pi')$ Prior covariance functional, RBF kernel with hyperparameters $\boldsymbol{\theta}$
- $g_{\mathcal{S}} : \Pi_{\mathcal{S}} \to \mathbb{R}$ Functional objective from the space $\Pi_{\mathcal{S}}$ of all bounded (vector-valued) functions on $\boldsymbol{C}_{\mathcal{S}} = \bigcup_{\langle X, \boldsymbol{C}_X \rangle \in \mathcal{S}} \boldsymbol{C}_X$ to the reals

Functional Causal Bayesian Optimization: The Algorithm

Algorithm 1 fCBO

Inputs: $\mathcal{G}, \boldsymbol{I}, Y, \{\mathcal{D}^I_{\mathcal{S}}\}_{\mathcal{S}\in\Sigma}, T, S$

$\mathbb{M}_\Sigma \leftarrow \text{ReduceSearchSpace}(\mathcal{G}, \boldsymbol{I}, Y)$ — Reduce space over MPSs

Initialise GPs $g_{\mathcal{S}}(\pi_{\mathcal{S}})\ \forall \mathcal{S} \in \mathbb{M}_\Sigma$ with $\mathcal{D}^I_{\mathcal{S}}$

for $t = 1, \ldots, T$ **do**

1. Select MPS $\mathcal{S}_t$ and DMP $\pi^t_{\mathcal{S}_t}$ via the fEI — Max acquisition functional
2. Obtain $\mathcal{D}^I_{\pi^t_{\mathcal{S}_t}} = \{y^{(s)}\}_{s=1}^S$ from $p_{\pi^t_{\mathcal{S}_t}}(Y)$
3. Compute sample mean estimates $\hat{\mu}^Y_{\pi^t_{\mathcal{S}_t}}$ using $\mathcal{D}^I_{\pi^t_{\mathcal{S}_t}}$ — Obtain associated target effect
4. $\mathcal{D}^I_{\mathcal{S}_t} \leftarrow \mathcal{D}^I_{\mathcal{S}_t} \cup (\pi^t_{\mathcal{S}_t}, \hat{\mu}^Y_{\pi^t_{\mathcal{S}_t}})$ — Update interventional dataset
5. Update $\tau(g_{\mathcal{S}_t} | \mathcal{D}^I_{\mathcal{S}_t})$ — Posterior update

end

Output: $(\mathcal{S}^*, \pi^*_{\mathcal{S}^*})$ with min $\hat{\mu}^Y_{\pi^*_{\mathcal{S}^*}}$ over $\{\mathcal{D}^I_{\mathcal{S}}\}_{\mathcal{S}\in\mathbb{M}_\Sigma}$

Main Contributions & Challenges

1

Conceptual/ Theoretical

Intuitively, **soft interventions should matter**

But we needed to show **how and when**

2

Technical

Optimizing soft interventions requires **new methodology and code**, including a **functional GP approach**, and a **kernel measuring distance between soft, and possible hard/soft mixed interventions**

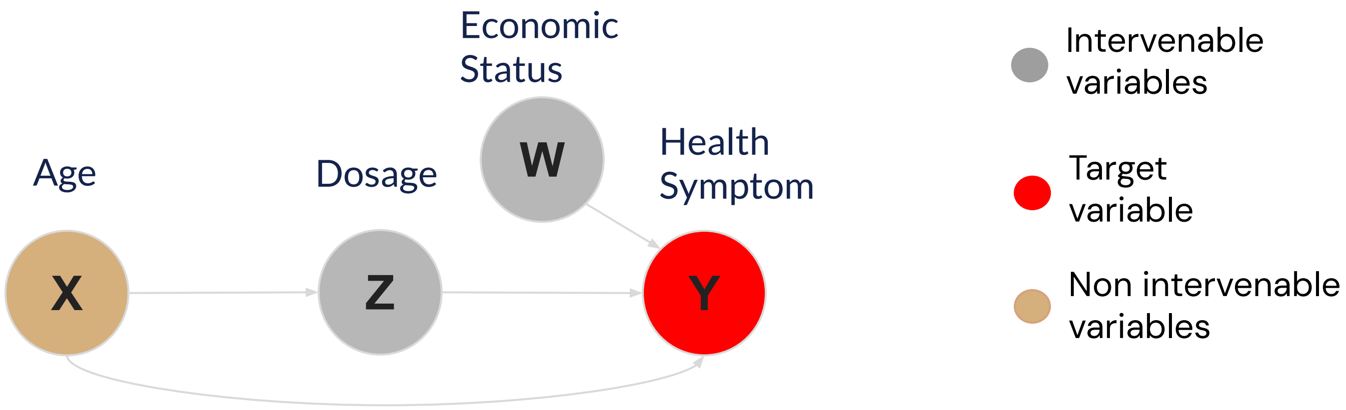
3

Experiments

Defining a **simulated setting** showcasing soft interventions, including **subgroup optimality**

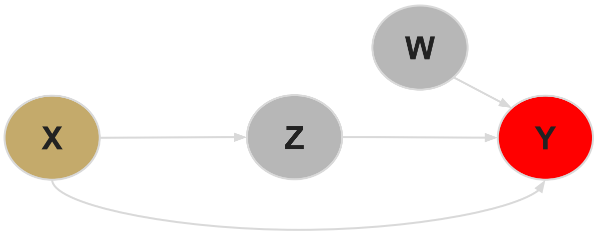
Experimenting in the **healthcare setting,** showcasing soft interventions and their cost implications

Experiments: Augmented Chain Setting



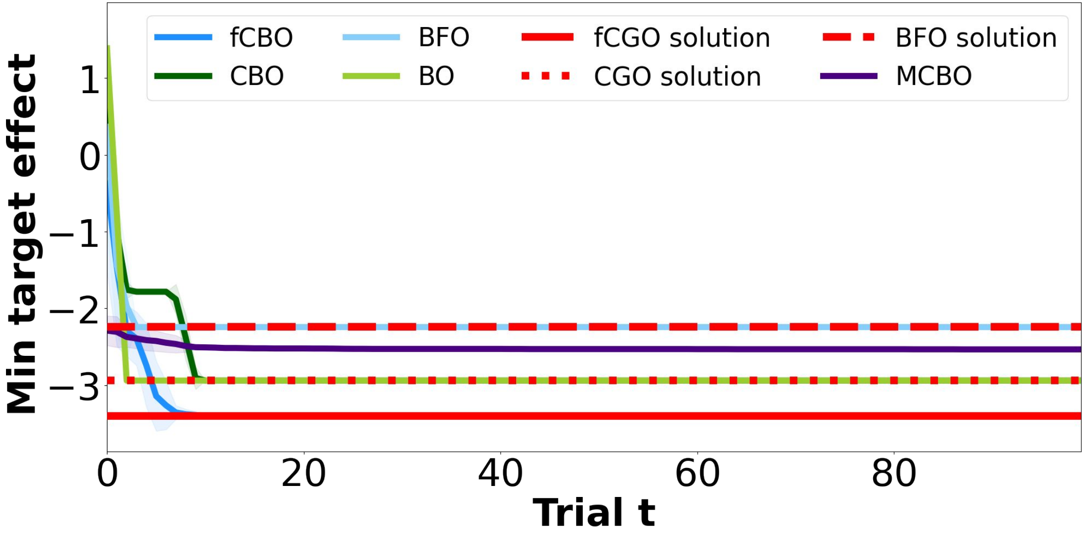
$$x = u_x$$
$$z = -0.5x + u_z$$
$$w = u_w$$
$$y = -w - 3zx + u_y$$

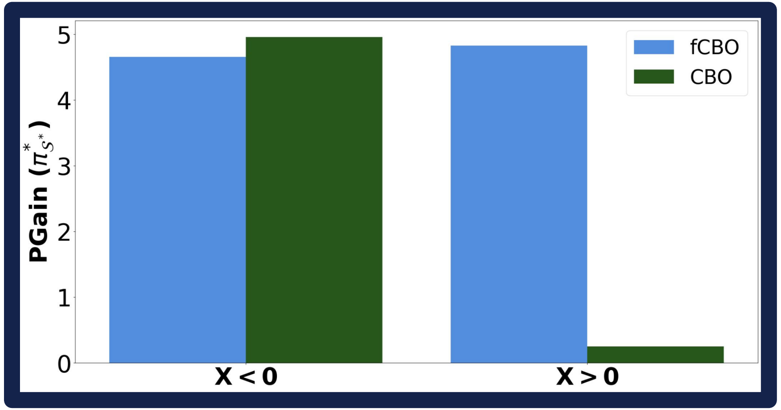
Showcasing Subgroup Optimality



CBO $\boldsymbol{X}^* = \{Z, W\}\ \boldsymbol{x}^* = (-1, 1)$

fCBO $\mathcal{S}^* = \{\langle Z, X\rangle\ \langle W, \emptyset\rangle\}$



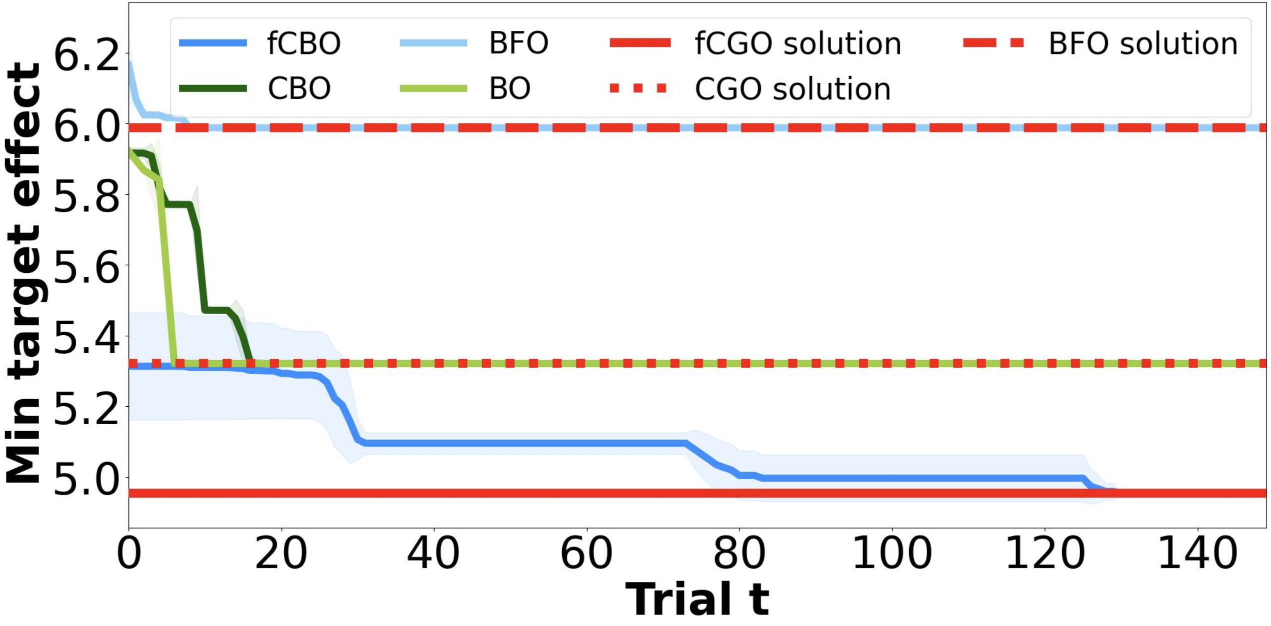


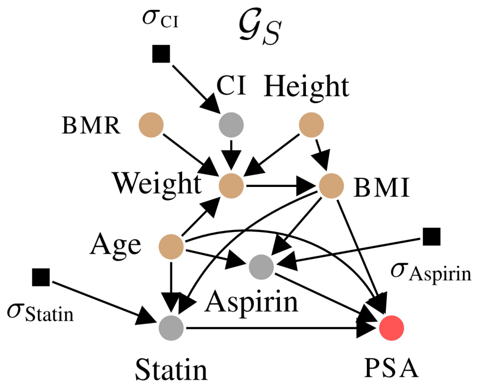
$\mathrm{PGain}(\pi_{\mathcal{S}}, \boldsymbol{C} = \boldsymbol{c}) = \hat{\mu}^{Y}_{\boldsymbol{C}=\boldsymbol{c}} - \hat{\mu}^{Y}_{\pi_{\mathcal{S}}, \boldsymbol{C}=\boldsymbol{c}}$

Experiments: Healthcare Example

CBO $\mathbb{M}_{\mathcal{G}} = \{\{\text{CI}\}, \{\text{Statin}\}, \{\text{Aspirin}\}, \{\text{CI, Statin}\}, \{\text{CI, Aspirin}\}, \{\text{Statin, Aspirin}\}, \{\text{CI, Statin, Aspirin}\}\}$

fCBO $\mathbb{M}_{\Sigma} = \{\{\langle \text{CI}, \emptyset \rangle\}, \{\langle \text{Statin}, \emptyset \rangle\}, \{\langle \text{Aspirin}, \emptyset \rangle\}, \ldots, \{\langle \text{CI}, \emptyset \rangle, \langle \text{Statin}, \{\text{Age, BMI}\} \rangle, \langle \text{Aspirin}, \{\text{Age, BMI}\} \rangle\}\}$

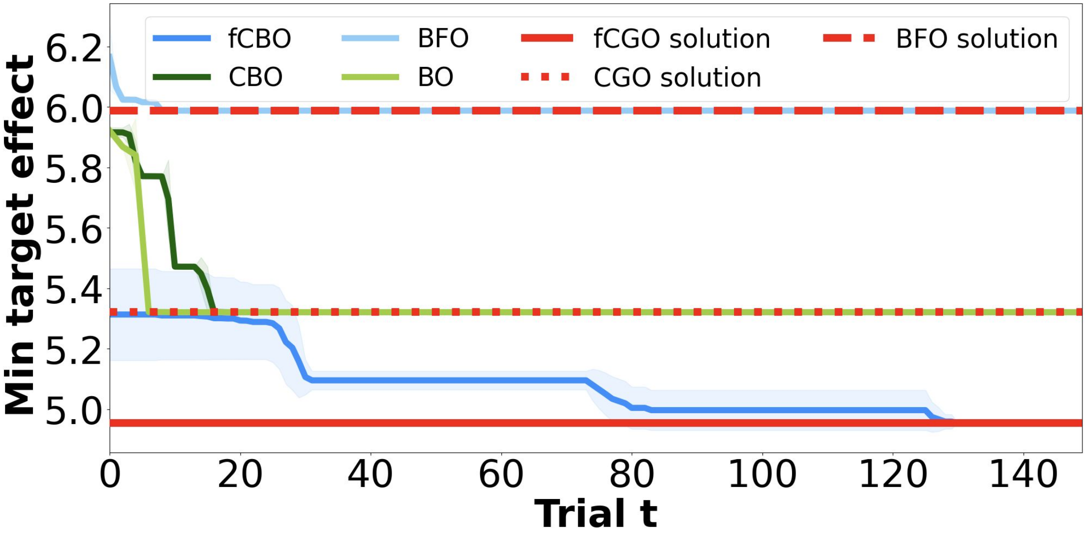


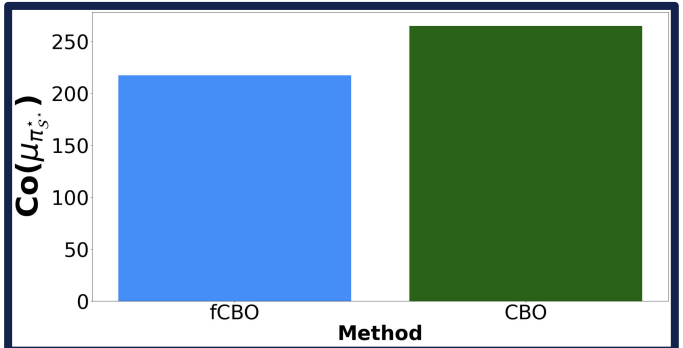


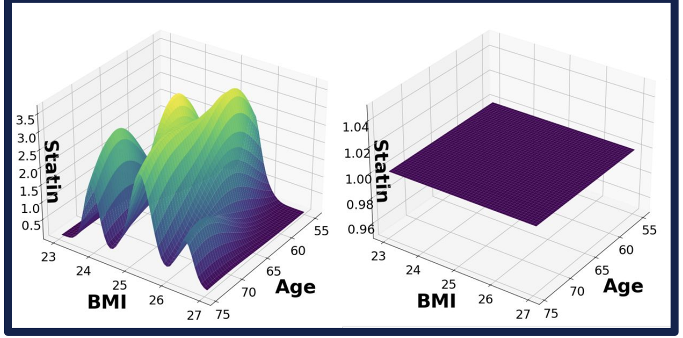
Showcasing cost and targeted allocation

CBO $\boldsymbol{X}^* = \{\text{CI}, \text{Statin}, \text{Aspirin}\}$ $\boldsymbol{x}^* = (1, 1, 0)$

fCBO $\boldsymbol{S}^* = \{\langle \text{CI}, \emptyset \rangle, \langle \text{Statin}, \{\text{Age, BMI}\} \rangle, \langle \text{Aspirin}, \emptyset \rangle\}$







Functional Causal Bayesian Optimization (fCBO)

- Targeted, more personalized treatment
- Subgroup optimality
- Lower cost treatments, focusing on most needed/promising contexts
- Hard interventions are special cases of soft, so no loss by considering soft