Appendix for sharp bounds for generalized causal sensitivity analysis

Anonymous Author(s)
Affiliation
Address
`email`

Contents

Submitted to 37th Conference on Neural Information Processing Systems (NeurIPS 2023). Do not distribute.

A Extended related work

A.1 Partial identification under unobserved confounding

There are various works for partial identification (i.e., bounding causal effects) under unobserved confounding beyond causal sensitivity analysis. In order to achieve informative bounds without restricting the strength of unobserved confounding, these works impose restrictive assumptions on the data-generating process, such availability of additional variables. One example is instrumental variables (IVs), i.e., variables, which only have a direct effect on treatment variables but not on outcomes. Under certain assumptions, IVs render bounds for causal effects informative without assumptions on the underlying confounding structure [1, 13, 21, 28]. Other examples include leaky mediation [1, 28], differential effect [5], noisy proxy settings [14], or discrete canonical SCMs [41].

Note that none of these methods aims at sensitivity analysis, i.e., bounding causal effects under restrictions of the unconfoundedness assumption. Furthermore, none is applicable in the causal inference settings we consider (e.g., continuous treatments or outcomes, no IVs available, etc.).

A.2 Estimation of causal effects under unconfoundedness

Under certain additional assumptions, unconfoundedness makes it possible to point-identify causal effects from the observational data, so that the causal inference problem reduces to a purely statistical problem. Various methods for estimating point-identified causal effects under unconfoundedness have been proposed that make use of machine learning or/and (semiparametric) statistical theory. Examples include methods for conditional average treatment effects [7, 8, 19, 23, 32, 39, 43], average treatment effects [12, 33, 38], instrumental variables [2, 11, 15, 34, 35, 42], time-varying data [3, 24, 25], mediation analysis [37, 10], and distributional effects [6, 20, 26, 27]. Note that all the methods above are biased if the unconfoundedness assumption is violated, which outlines the need for our causal sensitivity analysis.

B Proofs of the GMSM bounds

B.1 Proof of Theorem 1

Proof. We give the proof for continuous $W \in \mathbb{R}$ (see Eq. (4) in the main paper). The derivation for discrete $W \in \mathbb{N}$ (see Eq. (5) in the main paper) follows with the same arguments and the normalization constraint $\sum_w \mathbb{P}^+(w \mid \mathbf{x}, \mathbf{m}_W, \mathbf{a}) = 1$. We prove the equality

$$F_+(w) = \inf_{\mathcal{M} \in \mathcal{C}(\mathcal{S})} F_{\mathcal{M}}(w) \tag{9}$$

by showing both inequalities

$$F_+(w) \geq \inf_{\mathcal{M} \in \mathcal{C}(\mathcal{S})} F_{\mathcal{M}}(w) \quad \text{and} \quad F_+(w) \leq \inf_{\mathcal{M} \in \mathcal{C}(\mathcal{S})} F_{\mathcal{M}}(w). \tag{10}$$

The result for $F_-(w)$ follows analogously.

First inequality ($\geq$): We show that there exists an SCM $\mathcal{M} \in \mathcal{C}(\mathcal{S})$ with induced interventional density $\mathbb{P}_+(w \mid \mathbf{x}, \mathbf{m}_W, \mathbf{a})$ for all w. The construction of $\mathcal{M}$ is similar to that of our motivational toy example in Sec. 4.1 of the main paper. We first define an (interventional) probability density for the unobserved confounder $\mathbf{U}_W \in \mathbb{R}^d$ given $\mathbf{X}$ via

$$\mathbb{P}(\mathbf{u}_W \mid \mathbf{x}, do(\mathbf{A} = \mathbf{a})) = \mathbb{1}(0 \leq u_W^{(1)} \leq c_W^+)\,(1/s_W^+) + \mathbb{1}(1 \geq u_W^{(1)} > c_W^+)\,(1/s_W^-), \tag{11}$$

where $u_W^{(1)}$ denotes the first coordinate of $\mathbf{u}_W$. $\mathbb{P}(\mathbf{u}_W \mid \mathbf{x}, \mathbf{m}_W)$ is a properly normalized density with support $[0,1]^d$ because

$$\int \mathbb{P}(\mathbf{u}_W \mid \mathbf{x}, do(\mathbf{A} = \mathbf{a}))\,\mathrm{d}\mathbf{u}_W = \frac{c_W^+}{s_W^+} + \frac{1 - c_W^+}{s_W^-} = \frac{1 - s_W^-}{s_W^+ - s_W^-} + \frac{s_W^+ - 1}{s_W^+ - s_W^-} = 1, \tag{12}$$

where we used the definition of $c_W^+ = \frac{(1 - s_W^-) s_W^+}{s_W^+ - s_W^-}$.

We now define the probability density for the unobserved confounder $\mathbf{U}_W \in \mathbb{R}^d$ given $\mathbf{X}$, $\mathbf{M}_W$, and treatments $\mathbf{A}$ as the uniform density on $[0,1]^d$, i.e.,

$$\mathbb{P}(\mathbf{u}_W \mid \mathbf{x}, \mathbf{a}) = \mathbb{1}(0 \leq \mathbf{u}_W \leq 1). \tag{13}$$

Note that there always exists an SCM $\mathcal{M}$ which induces the densities in Eq. (11) and Eq. (12) from above (see [30], Proposition 7.1). Furthermore, it holds that

$$s_W^- \leq \frac{\mathbb{P}(\mathbf{U}_W = \mathbf{u}_W \mid \mathbf{x}, \mathbf{a})}{\mathbb{P}(\mathbf{U}_W = \mathbf{u}_W \mid \mathbf{x}, do(\mathbf{A} = \mathbf{a}))} \leq s_W^+ \quad \text{for all} \quad \mathbf{u}_W \in [0,1]^p, \tag{14}$$

so that $\mathcal{M}$ respects the sensitivity constraint of the GMSM $\mathcal{S}$.

Let now $\mathbb{P}(w \mid \mathbf{x}, \mathbf{m}_W, \mathbf{a})$ denote the observational density of W given $\mathbf{X}$, $\mathbf{M}_W$, and $\mathbf{A}$ with corresponding cumulative distribution function (CDF) given by $F(w)$. To complete our construction of $\mathcal{M}$, we define functional assignment $W = f(\mathbf{X}, \mathbf{M}_W, \mathbf{A}, \mathbf{U}_W)$ via the inverse CDF

$$f(\mathbf{x}, \mathbf{m}_W, \mathbf{a}, \mathbf{u}_W) = F^{-1}\left(u_W^{(1)}\right). \tag{15}$$

By denoting $f(\mathbf{x}, \mathbf{m}_W, \mathbf{a}, \cdot)$ as $f_{\mathbf{x}, \mathbf{m}_W, \mathbf{a}}$, we can write the observational distribution under $\mathcal{M}$ as the push forward

$$f_{\mathbf{x}, \mathbf{m}_W, \mathbf{a}\#} \mathbb{P}^{\mathbf{U}_W \mid \mathbf{x}, \mathbf{a}}(w) = \mathbb{P}(w \mid \mathbf{x}, \mathbf{m}_W, \mathbf{a}) \tag{16}$$

due to Eq. (12) and Eq. (15). Note that we used here the assumption of Theorem 1 (main paper) that $\mathbf{U}_W$ is not a parent of $\mathbf{M}_W$. Hence, $\mathcal{M} \in \mathcal{C}(\mathcal{S})$ is compatible with the sensitivity model $\mathcal{S}$. Furthermore, the induced interventional distribution can be written as the push forward

$$f_{\mathbf{x}, \mathbf{m}_W, \mathbf{a}\#} \mathbb{P}^{\mathbf{U}_W \mid \mathbf{x}, do(\mathbf{A} = \mathbf{a})}(w) = \mathbb{P}_+(w \mid \mathbf{x}, \mathbf{m}_W, \mathbf{a}) \tag{17}$$

because of Eq. (11).

Second inequality ($\leq$): We provide proof by contradiction. To do so, we assume that there exists an SCM $\mathcal{M} \in \mathcal{C}(\mathcal{S})$ and $w \in \mathbb{R}$ so that

$$F_+(w) > F_{\mathcal{M}}(w). \tag{18}$$

By the definition of $F_+(w)$, there must exist a set $\mathcal{W}_1 \subseteq \mathbb{R}_{\leq F^{-1}(c_W^+)}$, so that

$$\mathbb{P}_+(w_1 \mid \mathbf{x}, \mathbf{m}_W, \mathbf{a}) > \mathbb{P}(w_1 \mid \mathbf{x}, \mathbf{m}_W, do(\mathbf{A} = \mathbf{a})) \quad \text{for all} \quad w_1 \in \mathcal{W}_1, \tag{19}$$

or a set $\mathcal{W}_2 \subseteq \mathbb{R}_{> F^{-1}(c_W^+)}$, so that

$$\mathbb{P}_+(w_2 \mid \mathbf{x}, \mathbf{m}_W, \mathbf{a}) < \mathbb{P}(w_2 \mid \mathbf{x}, \mathbf{m}_W, do(\mathbf{A} = \mathbf{a})) \quad \text{for all} \quad w_2 \in \mathcal{W}_2, \tag{20}$$

as otherwise $\mathbb{P}(w \mid \mathbf{x}, \mathbf{m}_W, do(\mathbf{A} = \mathbf{a}))$ would not integrate to 1. Let $W = f(\mathbf{X}, \mathbf{M}_W, \mathbf{A}, \mathbf{U}_W)$ be the functional assignment of $\mathcal{M}$ and let $\mathcal{U}_1 = f^{-1}_{\mathbf{x},\mathbf{m}_W,\mathbf{a}}(\mathcal{W}_1) \subseteq \mathbb{R}^d$ and $\mathcal{U}_2 = f^{-1}_{\mathbf{x},\mathbf{m}_W,\mathbf{a}}(\mathcal{W}_2) \subseteq \mathbb{R}^d$ denote the preimages of $\mathcal{W}_1$ and $\mathcal{W}_2$ under $f_{\mathbf{x},\mathbf{m}_W,\mathbf{a}}$ in the confounding space.

We can again write $\mathbb{P}_+(w \mid \mathbf{x}, \mathbf{m}_W, \mathbf{a})$ as a push forward

$$\mathbb{P}^+(w \mid \mathbf{x}, \mathbf{m}_W, \mathbf{a}) = f_{\mathbf{x},\mathbf{m}_W,\mathbf{a}\#}\mathbb{P}_+^{\mathbf{U}_W \mid \mathbf{x},\mathbf{a}}(w) \tag{21}$$

for some density $\mathbb{P}_+(\mathbf{u}_W \mid \mathbf{x}, \mathbf{a})$ on the confounding space. By the definition of $\mathbb{P}_+(w \mid \mathbf{x}, \mathbf{m}_W, \mathbf{a})$ and Eq. (16), we obtain

$$\mathbb{P}_+(\mathbf{u}_1 \mid \mathbf{x}, \mathbf{a}) = \frac{1}{s_W^+}\mathbb{P}(\mathbf{u}_1 \mid \mathbf{x}, \mathbf{a}) \quad \text{and} \quad \mathbb{P}_+(\mathbf{u}_2 \mid \mathbf{x}, \mathbf{a}) = \frac{1}{s_W^-}\mathbb{P}(\mathbf{u}_2 \mid \mathbf{x}, \mathbf{a}) \tag{22}$$

for all $\mathbf{u}_1 \in \mathcal{U}_1$ and $\mathbf{u}_2 \in \mathcal{U}_2$. Due to the definition of $\mathcal{U}_1$ and $\mathcal{U}_2$, it follows that there exist $\mathbf{u}_1 \in \mathcal{U}_1$ and $\mathbf{u}_2 \in \mathcal{U}_2$, so that

$$\frac{\mathbb{P}(\mathbf{u}_1 \mid \mathbf{x}, \mathbf{a})}{\mathbb{P}(\mathbf{u}_1 \mid \mathbf{x}, do(\mathbf{A} = \mathbf{a}))} > \frac{\mathbb{P}(\mathbf{u}_1 \mid \mathbf{x}, \mathbf{a})}{\mathbb{P}_+(\mathbf{u}_1 \mid \mathbf{x}, \mathbf{a})} = s_W^+ \tag{23}$$

and

$$\frac{\mathbb{P}(\mathbf{u}_1 \mid \mathbf{x}, \mathbf{a})}{\mathbb{P}(\mathbf{u}_1 \mid \mathbf{x}, do(\mathbf{A} = \mathbf{a}))} < \frac{\mathbb{P}(\mathbf{u}_1 \mid \mathbf{x}, \mathbf{a})}{\mathbb{P}_+(\mathbf{u}_1 \mid \mathbf{x}, \mathbf{a})} = s_W^-. \tag{24}$$

Both Eq. (23) and Eq. (24) are contradictions to the GMSM constraint Eq. (3) of the main paper. Hence, $\mathcal{M} \notin \mathcal{C}(\mathcal{S})$. □

B.2 Proof of Corollary 1

Here, we formally restate Corollary 1 for *monotone* functionals. For two probability densities $\mathbb{P}(y)$ and $\mathbb{P}'(y)$, we denote $\mathbb{P} \leq \mathbb{P}'$ if $F \geq F'$ holds almost surely for the corresponding CDFs.

Definition 5. A functional $\mathcal{D}$ is called monotone if $\mathcal{D}(\mathbb{P}(\cdot)) \leq \mathcal{D}(\mathbb{P}'(\cdot))$ whenever $\mathbb{P} \leq \mathbb{P}'$.

Intuitively, a monotone functional increases if applied on a distribution that is further right-shifted. Note that both the expectation functional $\mathcal{D}(\mathbb{P}(\cdot)) = \int y \mathbb{P}(y) \, \mathrm{d}y$ and the quantile functionals $\mathcal{D}(\mathbb{P}(\cdot)) = F^{-1}(\alpha)$ for $\alpha \in [0, 1]$ are monotone.

Corollary 1 (Restatement)**.** *If* $\mathbf{M} = \emptyset$ *and* $\mathcal{D}$ *is monotone, we obtain sharp bounds*

$$Q^+(\mathbf{x}, \mathbf{a}, \mathcal{S}) = \mathcal{D}\left(\mathbb{P}_+^Y(\cdot \mid \mathbf{x}, \mathbf{a})\right) \quad \textit{and} \quad Q^-(\mathbf{x}, \mathbf{a}, \mathcal{S}) = \mathcal{D}\left(\mathbb{P}_-^Y(\cdot \mid \mathbf{x}, \mathbf{a})\right). \tag{25}$$

Proof. Follows directly from Theorem 1 for $W = Y$. □

B.3 Proof of Corollary 2

Proof. We derive Algorithm 1 for $Q^+(\mathbf{x}, \bar{\mathbf{a}}, \mathcal{S})$. The case for $Q^-(\mathbf{x}, \bar{\mathbf{a}}, \mathcal{S})$ follows analogously.

Recall that we want to maximize the causal effect

$$Q(\mathbf{x}, \bar{\mathbf{a}}, \mathcal{M}) = \sum_{\mathbf{m}} \mathcal{D}\left(\mathbb{P}^Y(\cdot \mid \mathbf{x}, \mathbf{m}, do(\mathbf{A} = \mathbf{a}_{\ell+1}))\right) \prod_{i=1}^{\ell} \mathbb{P}(m_i \mid \mathbf{x}, \bar{\mathbf{m}}_{i-1}, do(\mathbf{A} = \mathbf{a}_i)), \tag{26}$$

over all possible SCMs $\mathcal{M} \in \mathcal{C}(\mathcal{S})$ that are compatible with the GMSM $\mathcal{S}$. By using the assumption (no unobserved confounding between mediators and outcome), we can write $Q(\mathbf{x}, \bar{\mathbf{a}}, \mathcal{M})$ in terms of functional assignments $f^W_{\mathbf{x},\mathbf{m}_W,\mathbf{a}}$ defined via $W = f^W(\mathbf{X}, \mathbf{M}_W, \mathbf{A}, \mathbf{U}_W)$ and induced (interventional) distributions $\mathbb{P}^{\mathbf{U}_W \mid \mathbf{x}}$ in the following way:

$$Q(\mathbf{x}, \bar{\mathbf{a}}, \mathcal{M}) = \sum_{\mathbf{m}} \mathcal{D}\left(f^Y_{\mathbf{x},\bar{\mathbf{m}}_\ell,\mathbf{a}_{\ell+1}\#}\mathbb{P}^{\mathbf{U}_Y \mid \mathbf{x}}(\cdot)\right) \prod_{i=1}^{\ell} f^{M_i}_{\mathbf{x},\bar{\mathbf{m}}_{i-1},\mathbf{a}_i\#}\mathbb{P}^{\mathbf{U}_{M_i} \mid \mathbf{x}}(m_i). \tag{27}$$

Hence, the optimization problem reduces to maximizing Eq. (27) over all functional assignments $f^W_{\mathbf{x},\mathbf{m}_W,\mathbf{a}}$ and distributions $\mathbb{P}^{\mathbf{U}_W|\mathbf{x}}$ that are compatible with $\mathcal{S}$. Note that the terms in the product do not depend on each other or the term in the sum. Thus, by rearranging the suprema and products, we can equivalently perform the following iterative procedure: First, we initialize

$$Q^+_{\ell+1}(\mathbf{x}, \bar{\mathbf{m}}_\ell, \bar{\mathbf{a}}, \mathcal{S}) = \sup_{\mathcal{M}\in\mathcal{C}(\mathcal{S})} \mathcal{D}\left({f^Y_{\mathbf{x},\bar{\mathbf{m}}_\ell,\mathbf{a}_{\ell+1}}}_{\#}\mathbb{P}^{\mathbf{U}_Y|\mathbf{x}}(\cdot)\right) \tag{28}$$

and then define

$$Q^+_i(\mathbf{x}, \bar{\mathbf{m}}_{i-1}, \bar{\mathbf{a}}, \mathcal{S}) = \sup_{\mathcal{M}\in\mathcal{C}(\mathcal{S})} \sum_{m_i} Q^+_{i+1}(\mathbf{x}, \bar{\mathbf{m}}_{i-1}, m_i, \bar{\mathbf{a}})) \left({f^{M_i}_{\mathbf{x},\bar{\mathbf{m}}_{i-1},\mathbf{a}_i}}_{\#}\mathbb{P}^{\mathbf{U}_Y|\mathbf{x}}(m_i)\right) \tag{29}$$

for all $i \in \{\ell, \dots, 1\}$, which results in the sharp upper bound

$$Q^+(\mathbf{x}, \bar{\mathbf{a}}, \mathcal{S}) = Q^+_1(\mathbf{x}, \bar{\mathbf{a}}, \mathcal{S}). \tag{30}$$

For Eq. (28) and monotone $\mathcal{D}$, we can directly apply Theorem 1 and obtain

$$Q^+_{\ell+1}(\mathbf{x}, \bar{\mathbf{m}}_\ell, \bar{\mathbf{a}}, \mathcal{S}) = \mathcal{D}\left(\mathbb{P}^Y_+(\cdot \mid \mathbf{x}, \bar{\mathbf{m}}_\ell, \mathbf{a}_{\ell+1})\right). \tag{31}$$

For Eq. (29), we need to find an induced distribution ${f^{M_i}_{\mathbf{x},\bar{\mathbf{m}}_{i-1},\mathbf{a}_i}}_{\#}\mathbb{P}^{\mathbf{U}_Y|\mathbf{x}}$ on M_i that is compatible with $\mathcal{S}$ and puts most probability mass on m_i where $Q^+_{i+1}(\mathbf{x}, \bar{\mathbf{m}}_{i-1}, m_i, \bar{\mathbf{a}}))$ is large. Hence, we can apply the discrete version of Theorem 1 with $W = \pi(M_i)$, where $\pi\colon \mathit{supp}(M_i) \to \mathit{supp}(M_i)$ is a permutation map so that $\left(Q^+_{i+1}(\mathbf{x}, \bar{\mathbf{m}}_{i-1}, \pi(m_i)), \bar{\mathbf{a}}\right)_{m_i\in \mathit{supp}(M_i)}$ is ordered in ascending order. The corresponding update step is shown in Algorithm 1.

$\square$

C Special cases of the GMSM

In this section, we prove Lemma 1, i.e., we show that all sensitivity models introduced in Sec. 3.3 of the main paper are special cases of our (weighted) GMSM. Recall that we consider settings with mediators (i.e., $\mathbf{M} = \emptyset$), and write $\Gamma = \Gamma_Y$ for the sensitivity parameter, $q(\mathbf{a}, \mathbf{x}) = q_Y(\mathbf{a}, \mathbf{x})$ for the weight function, and $\mathbf{U} = \mathbf{U}_Y$ for the unobserved confounders. In this case, the weighted GMSM is defined via the confounding restriction

$$\frac{1}{(1-\Gamma)\, q(\mathbf{a}, \mathbf{x}) + \Gamma} \leq \frac{\mathbb{P}(\mathbf{U} = \mathbf{u} \mid \mathbf{x}, \mathbf{a})}{\mathbb{P}(\mathbf{U} = \mathbf{u} \mid \mathbf{x}, do(\mathbf{A} = \mathbf{a}))} \leq \frac{1}{(1-\Gamma^{-1})\, q(\mathbf{a}, \mathbf{x}) + \Gamma^{-1}}. \tag{32}$$

Marginal sensitivity model (MSM): The MSM [36] for binary treatment $\mathbf{A} = A \in \{0, 1\}$ is defined via

$$\frac{1}{\Gamma} \leq \frac{\pi(\mathbf{x})}{1 - \pi(\mathbf{x})} \frac{1 - \pi(\mathbf{x}, \mathbf{u})}{\pi(\mathbf{x}, \mathbf{u})} \leq \Gamma, \tag{33}$$

where $\pi(\mathbf{x}) = \mathbb{P}(A = 1 \mid \mathbf{x})$ denotes the observed propensity score and $\pi(\mathbf{x}, \mathbf{u}) = \mathbb{P}(A = 1 \mid \mathbf{x}, \mathbf{u})$ denotes the full propensity score. By rearranging the terms, we obtain

$$\frac{1}{(1-\Gamma)\mathbb{P}(a \mid x) + \Gamma} \leq \frac{\mathbb{P}(a \mid x, u)}{\mathbb{P}(a \mid x)} \leq \frac{1}{(1-\Gamma^{-1})\mathbb{P}(a \mid x) + \Gamma^{-1}} \tag{34}$$

for $a \in \{0, 1\}$. Furthermore, by Bayes' theorem, it follows that

$$\frac{\mathbb{P}(a \mid x, u)}{\mathbb{P}(a \mid x)} = \frac{\mathbb{P}(u \mid x, a)\mathbb{P}(a \mid x)}{\mathbb{P}(u \mid x)\mathbb{P}(a \mid x)} = \frac{\mathbb{P}(u \mid x, a)}{\mathbb{P}(u \mid x)} = \frac{\mathbb{P}(u \mid x, a)}{\mathbb{P}(u \mid x, do(A = a))}, \tag{35}$$

which implies that the MSM is a weighted GMSM with weight function $q(\mathbf{a}, \mathbf{x}) = \mathbb{P}(\mathbf{a} \mid \mathbf{x})$.

Continuous marginal sensitivity model (CMSM): For continuous treatments $\mathbf{A} \in \mathbb{R}^d$, the continuous marginal sensitivity model (CMSM) [17] is defined via

$$\frac{1}{\Gamma} \leq \frac{\mathbb{P}(\mathbf{a} \mid \mathbf{x}, \mathbf{u})}{\mathbb{P}(\mathbf{a} \mid \mathbf{x})} \leq \Gamma. \tag{36}$$

With the same arguments as in Eq. (35), it follows that the CMSM is a weighted GMSM with weight function $q(\mathbf{a}, \mathbf{x}) = 0$.

Longitudinal marginal sensitivity model (LMSM): For longitudinal settings with time-varying observed confounders $\mathbf{X} = \bar{\mathbf{X}}_T = (\mathbf{X}_1, \ldots, \mathbf{X}_T)$, unobserved confounders $\mathbf{U} = \bar{\mathbf{U}}_T = (\mathbf{U}_1, \ldots, \mathbf{U}_T)$, treatments $\mathbf{A} = \bar{\mathbf{A}}_T = (\mathbf{A}_1, \ldots, \mathbf{A}_T)$, the longitudinal marginal sensitivity model (LMSM) [4] is defined via

$$\frac{1}{\Gamma} \leq \prod_{t=1}^{T} \frac{\mathbb{P}(\mathbf{a}_t \mid \bar{\mathbf{x}}_T, \bar{\mathbf{u}}_t, \bar{\mathbf{a}}_{t-1})}{\mathbb{P}(\mathbf{a}_t \mid \bar{\mathbf{x}}_T, \bar{\mathbf{a}}_{t-1})} \leq \Gamma. \tag{37}$$

It holds that

$$\mathbb{P}(\bar{\mathbf{u}}_T \mid \bar{\mathbf{x}}_T, \bar{\mathbf{a}}_T) = \frac{\prod_{t=1}^{T} \mathbb{P}(\bar{\mathbf{u}}_t \mid \bar{\mathbf{x}}_T, \bar{\mathbf{a}}_t)}{\prod_{t=1}^{T-1} \mathbb{P}(\bar{\mathbf{u}}_t \mid \bar{\mathbf{x}}_T, \bar{\mathbf{a}}_t)} = \prod_{t=1}^{T} \frac{\mathbb{P}(\bar{\mathbf{u}}_t \mid \bar{\mathbf{x}}_T, \bar{\mathbf{a}}_t)}{\mathbb{P}(\bar{\mathbf{u}}_{t-1} \mid \bar{\mathbf{x}}_T, \bar{\mathbf{a}}_{t-1})} \tag{38}$$

$$\overset{(*)}{=} \left(\prod_{t=1}^{T} \frac{\mathbb{P}(\mathbf{a}_t \mid \bar{\mathbf{x}}_T, \bar{\mathbf{u}}_t, \bar{\mathbf{a}}_{t-1})}{\mathbb{P}(\mathbf{a}_t \mid \bar{\mathbf{x}}_T, \bar{\mathbf{a}}_{t-1})}\right) \left(\prod_{t=1}^{T} \frac{\mathbb{P}(\bar{\mathbf{u}}_t \mid \bar{\mathbf{x}}_T, \bar{\mathbf{a}}_{t-1})}{\mathbb{P}(\bar{\mathbf{u}}_{t-1} \mid \bar{\mathbf{x}}_T, \bar{\mathbf{a}}_{t-1})}\right) \tag{39}$$

$$= \left(\prod_{t=1}^{T} \frac{\mathbb{P}(\mathbf{a}_t \mid \bar{\mathbf{x}}_T, \bar{\mathbf{u}}_t, \bar{\mathbf{a}}_{t-1})}{\mathbb{P}(\mathbf{a}_t \mid \bar{\mathbf{x}}_T, \bar{\mathbf{a}}_{t-1})}\right) \left(\prod_{t=1}^{T} \mathbb{P}(\mathbf{u}_t \mid \bar{\mathbf{x}}_T, \bar{\mathbf{a}}_{t-1}, \bar{\mathbf{u}}_{t-1})\right) \tag{40}$$

$$= \left(\prod_{t=1}^{T} \frac{\mathbb{P}(\mathbf{a}_t \mid \bar{\mathbf{x}}_T, \bar{\mathbf{u}}_t, \bar{\mathbf{a}}_{t-1})}{\mathbb{P}(\mathbf{a}_t \mid \bar{\mathbf{x}}_T, \bar{\mathbf{a}}_{t-1})}\right) \mathbb{P}(\bar{\mathbf{u}}_T \mid \bar{\mathbf{x}}_T, do(\bar{\mathbf{A}}_T = \bar{\mathbf{a}}_T)), \tag{41}$$

where $(*)$ follows by applying Bayes' theorem on $\mathbb{P}(\bar{\mathbf{u}}_t \mid \bar{\mathbf{x}}_T, \bar{\mathbf{a}}_t)$. Hence, the LMSM is a weighted GMSM with weight function $q(\mathbf{a}, \mathbf{x}) = 0$.

D Bounds for average causal effects and differences

Here, we show that we can use our sharp bounds to obtain sharp bounds for causal effect averages and differences. We state the results for the upper bound

$$Q^{+}(\mathbf{x}, \bar{\mathbf{a}}, \mathcal{S}) = \sup_{\mathcal{M} \in \mathcal{C}(\mathcal{S})} Q(\mathbf{x}, \bar{\mathbf{a}}, \mathcal{M}). \tag{42}$$

All definitions and bounds for the lower bound $Q^{-}(\mathbf{x}, \bar{\mathbf{a}}, \mathcal{S})$ can be obtained by swapping the signs.

We are interested in the sharp upper bound for the *average causal effect*

$$Q^{+}_{\mathrm{avg}}(\bar{\mathbf{a}}, \mathcal{S}) = \sup_{\mathcal{M} \in \mathcal{C}(\mathcal{S})} \int_{\mathcal{X}} Q(\mathbf{x}, \bar{\mathbf{a}}, \mathcal{M}) \, \mathrm{d}\mathbf{x} \tag{43}$$

and the sharp upper bound for the *causal effect difference*

$$Q^{+}_{\mathrm{diff}}(\mathbf{x}, \bar{\mathbf{a}}_1, \bar{\mathbf{a}}_2, \mathcal{S}) = \sup_{\mathcal{M} \in \mathcal{C}(\mathcal{S})} \left(Q(\mathbf{x}, \bar{\mathbf{a}}_1, \mathcal{M}) - Q(\mathbf{x}, \bar{\mathbf{a}}_2, \mathcal{M}) \right). \tag{44}$$

Lemma 2. *We can compute $Q^{+}_{\mathrm{avg}}(\bar{\mathbf{a}}, \mathcal{S})$ and $Q^{+}_{\mathrm{diff}}(\mathbf{x}, \bar{\mathbf{a}}_1, \bar{\mathbf{a}}_2, \mathcal{S})$ from our sharp bounds $Q^{+}(\mathbf{x}, \bar{\mathbf{a}}, \mathcal{S})$ and $Q^{-}(\mathbf{x}, \bar{\mathbf{a}}, \mathcal{S})$ via*

$$Q^{+}_{\mathrm{avg}}(\bar{\mathbf{a}}, \mathcal{S}) = \int_{\mathcal{X}} Q^{+}(\mathbf{x}, \bar{\mathbf{a}}, \mathcal{S}) \, \mathrm{d}\mathbf{x} \tag{45}$$

and

$$Q^{+}_{\mathrm{diff}}(\mathbf{x}, \bar{\mathbf{a}}_1, \bar{\mathbf{a}}_2, \mathcal{S}) = Q^{+}(\mathbf{x}, \bar{\mathbf{a}}_1, \mathcal{S}) - Q^{-}(\mathbf{x}, \bar{\mathbf{a}}_2, \mathcal{S}). \tag{46}$$

Proof. The result for $Q^{+}_{\mathrm{avg}}(\bar{\mathbf{a}}, \mathcal{S})$ follows directly from interchanging the supremum and integral. For $Q^{+}_{\mathrm{diff}}(\mathbf{x}, \bar{\mathbf{a}}_1, \bar{\mathbf{a}}_2, \mathcal{S})$, we note that

$$\begin{aligned} Q^{+}_{\mathrm{diff}}(\mathbf{x}, \bar{\mathbf{a}}_1, \bar{\mathbf{a}}_2, \mathcal{S}) &\leq \sup_{\mathcal{M}_1 \in \mathcal{C}(\mathcal{S})} Q(\mathbf{x}, \bar{\mathbf{a}}_1, \mathcal{M}_1) - \inf_{\mathcal{M}_2 \in \mathcal{C}(\mathcal{S})} Q(\mathbf{x}, \bar{\mathbf{a}}_2, \mathcal{M}_2) &(47) \\ &= Q^{+}(\mathbf{x}, \bar{\mathbf{a}}_1, \mathcal{S}) - Q^{-}(\mathbf{x}, \bar{\mathbf{a}}_2, \mathcal{S}). &(48) \end{aligned}$$

To show the equality in Eq. 47, we show that, for each pair of SCMs $\mathcal{M}_1, \mathcal{M}_2 \in \mathcal{C}(\mathcal{S})$, we can find an SCM $\mathcal{M} \in \mathcal{C}(\mathcal{S})$ such that

$$Q(\mathbf{x}, \bar{\mathbf{a}}_1, \mathcal{M}_1) - Q(\mathbf{x}, \bar{\mathbf{a}}_2, \mathcal{M}_2) = Q(\mathbf{x}, \bar{\mathbf{a}}_1, \mathcal{M}) - Q(\mathbf{x}, \bar{\mathbf{a}}_2, \mathcal{M}). \tag{49}$$

We can assume w.l.o.g. that all $\mathcal{M} \in \mathcal{C}(\mathcal{S})$ induce the same distributions $\mathbb{P}^{\mathbf{U}_W \mid \mathbf{x}}$ on the confounding space. We denote the functional assignments of $\mathcal{M}_1$ and $\mathcal{M}_2$ as $f^{W}_{\mathcal{M}_1}(\mathbf{x}, \mathbf{m}_W, \mathbf{a}, \mathbf{u}_W)$ and $f^{W}_{\mathcal{M}_2}(\mathbf{x}, \mathbf{m}_W, \mathbf{a}, \mathbf{u}_W)$. We can now define a functional assignment $f^{W}_{\mathcal{M}}(\mathbf{x}, \mathbf{m}_W, \mathbf{a}, \mathbf{u}_W)$ for $\mathcal{M}$ so that

$$f^{W}_{\mathcal{M}}(\cdot, \cdot, \mathbf{a}_1, \cdot) = f^{W}_{\mathcal{M}_1}(\cdot, \cdot, \mathbf{a}_1, \cdot) \quad \text{and} \quad f^{W}_{\mathcal{M}}(\cdot, \cdot, \mathbf{a}_2, \cdot) = f^{W}_{\mathcal{M}_2}(\cdot, \cdot, \mathbf{a}_2, \cdot), \tag{50}$$

which implies Eq. (49).

$\square$

E Importance sampling estimators for finite sample bounds

In this section, we derive estimators for the outcome bound $\mathcal{D}\left(\mathbb{P}_+^Y(\cdot \mid \mathbf{x}, \mathbf{m}, \mathbf{a})\right)$ for continuous $Y \in \mathbb{R}$. We assume that we have already obtained an estimator $\hat{\mathbb{P}}^Y(\cdot \mid \mathbf{x}, \mathbf{m}, \mathbf{a})$ of the observational distribution $\mathbb{P}^Y(\cdot \mid \mathbf{x}, \mathbf{m}, \mathbf{a})$, and that we are able to sample $(y_i)_{i=1}^k \sim \hat{\mathbb{P}}^Y(\cdot \mid \mathbf{x}, \mathbf{m}, \mathbf{a})$ (see Appendix F for implementation details). Note that the outcome bound $\mathcal{D}\left(\mathbb{P}_+^Y(\cdot \mid \mathbf{x}, \mathbf{m}, \mathbf{a})\right)$ depends on the shifted distribution $\mathbb{P}_+^Y(\cdot \mid \mathbf{x}, \mathbf{m}, \mathbf{a})$ and not on the observational distribution $\mathbb{P}^Y(\cdot \mid \mathbf{x}, \mathbf{m}, \mathbf{a})$. Hence, we use an importance sampling approach to derive our estimators, which we outline in the following for the expectation functional and distributional effects. We denote the CDFs corresponding to $\mathbb{P}_+^Y(\cdot \mid \mathbf{x}, \mathbf{m}, \mathbf{a})$ and $\mathbb{P}^Y(\cdot \mid \mathbf{x}, \mathbf{m}, \mathbf{a})$ by $F_{\mathbb{P}_+^{Y \mid \mathbf{x}, \mathbf{m}, \mathbf{a}}}$ and $F_{\mathbb{P}^{Y \mid \mathbf{x}, \mathbf{m}, \mathbf{a}}}$, respectively.

Expectation functional: For the expectation functional, we can rewrite the outcome bound as

$$\begin{aligned}
\mathcal{D}\left(\mathbb{P}_+^Y(\cdot \mid \mathbf{x}, \mathbf{m}, \mathbf{a})\right) &= \mathbb{E}_{Y \sim \mathbb{P}_+^{Y \mid \mathbf{x}, \mathbf{m}, \mathbf{a}}}[Y] && (51)\\
&= \mathbb{E}_{Y \sim \mathbb{P}^{Y \mid \mathbf{x}, \mathbf{m}, \mathbf{a}}}\left[Y \frac{\mathbb{P}_+^Y(Y \mid \mathbf{x}, \mathbf{m}, \mathbf{a})}{\mathbb{P}^Y(Y \mid \mathbf{x}, \mathbf{m}, \mathbf{a})}\right] && (52)\\
&= \mathbb{E}_{Y \sim \mathbb{P}^{Y \mid \mathbf{x}, \mathbf{m}, \mathbf{a}}}\left[\frac{Y}{s_Y^+} \mathbb{1}\left(Y \leq F^{-1}_{\mathbb{P}^{Y \mid \mathbf{x}, \mathbf{m}, \mathbf{a}}}(c_Y^+)\right) + \frac{Y}{s_Y^-} \mathbb{1}\left(Y > F^{-1}_{\mathbb{P}^{Y \mid \mathbf{x}, \mathbf{m}, \mathbf{a}}}(c_Y^+)\right)\right] && (53)
\end{aligned}$$

to obtain the consistent estimator

$$\widehat{\mathcal{D}\left(\mathbb{P}_+^Y(\cdot \mid \mathbf{x}, \mathbf{m}, \mathbf{a})\right)} = \frac{1}{k} \sum_{i=1}^{\lfloor k c_Y^+ \rfloor} \frac{y_i}{\hat{s}_Y^+} + \frac{1}{k} \sum_{i=\lfloor k c_Y^+ \rfloor + 1}^{k} \frac{y_i}{\hat{s}_Y^-}, \tag{54}$$

where $(y_i)_{i=1}^k \sim \hat{\mathbb{P}}^Y(\cdot \mid \mathbf{x}, \mathbf{m}, \mathbf{a})$ are sampled from the estimated observational distribution. This corresponds to Eq. (8) in the main paper.

Distributional effects: We now derive estimators for distributional effects, i.e., for quantile functionals $\mathcal{D}$ of the form

$$\mathcal{D}\left(\mathbb{P}_+^Y(\cdot \mid \mathbf{x}, \mathbf{m}, \mathbf{a})\right) = F^{-1}_{\mathbb{P}_+^{Y \mid \mathbf{x}, \mathbf{m}, \mathbf{a}}}(\alpha) \tag{55}$$

with $\alpha \in (0, 1)$. We again use an importance sampling approach and rewrite

$$\begin{aligned}
F_{\mathbb{P}_+^{Y \mid \mathbf{x}, \mathbf{m}, \mathbf{a}}}(y) &= \mathbb{E}_{Y \sim \mathbb{P}_+^{Y \mid \mathbf{x}, \mathbf{m}, \mathbf{a}}}[\mathbb{1}(Y \leq y)] && (56)\\
&= \mathbb{E}_{Y \sim \mathbb{P}^{Y \mid \mathbf{x}, \mathbf{m}, \mathbf{a}}}\left[\mathbb{1}(Y \leq y) \frac{\mathbb{P}_+^Y(Y \mid \mathbf{x}, \mathbf{m}, \mathbf{a})}{\mathbb{P}^Y(Y \mid \mathbf{x}, \mathbf{m}, \mathbf{a})}\right] && (57)\\
&= \mathbb{E}_{Y \sim \mathbb{P}^{Y \mid \mathbf{x}, \mathbf{m}, \mathbf{a}}}\left[\frac{\mathbb{1}\left(Y \leq \min\{y, F^{-1}_{\mathbb{P}^{Y \mid \mathbf{x}, \mathbf{m}, \mathbf{a}}}(c_Y^+)\}\right)}{s_Y^+} + \frac{\mathbb{1}\left(F^{-1}_{\mathbb{P}^{Y \mid \mathbf{x}, \mathbf{m}, \mathbf{a}}}(c_Y^+) < Y \leq y\right)}{s_Y^-}\right]. && (58)
\end{aligned}$$

Hence, we can sample $(y_i)_{i=1}^k \sim \hat{\mathbb{P}}^Y(\cdot \mid \mathbf{x}, \mathbf{m}, \mathbf{a})$ and obtain the consistent estimator

$$\widehat{\mathcal{D}\left(\mathbb{P}_+^Y(\cdot \mid \mathbf{x}, \mathbf{m}, \mathbf{a})\right)} = \min_{\hat{F}_{\mathbb{P}_+^{Y \mid \mathbf{x}, \mathbf{m}, \mathbf{a}}}(y_i) \geq \alpha} y_i, \tag{59}$$

where

$$\hat{F}_{\mathbb{P}_+^{Y \mid \mathbf{x}, \mathbf{m}, \mathbf{a}}}(y) = \frac{1}{k} \sum_{i=1}^{\lfloor k c_Y^+ \rfloor} \frac{\mathbb{1}(y_i \leq y)}{\hat{s}_Y^+} + \frac{1}{k} \sum_{i=\lfloor k c_Y^+ \rfloor + 1}^{k} \frac{\mathbb{1}(y_i \leq y)}{\hat{s}_Y^-}. \tag{60}$$

F Implementation and hyperparameter tuning details

All our experimental settings feature a continuous outcome $Y \in \mathbb{R}$ and (optionally) discrete mediators $M_i \in \mathbb{N}$. Hence, we need to estimate the conditional outcome density $\mathbb{P}^Y(\cdot \mid \mathbf{x}, \mathbf{m}, \mathbf{a})$ and conditional probability mass functions $\mathbb{P}^{M_i}(\cdot \mid \mathbf{x}, \bar{\mathbf{m}}_{i-1}, \mathbf{a})$ in order to estimate our bounds with Eq. (54), Eq. (59), and Algorithm 1 from the main paper.

Conditional outcome density: We use conditional normalizing flows (CNFs) [40] for estimating the conditional density $\mathbb{P}^Y(\cdot \mid \mathbf{x}, \mathbf{m}, \mathbf{a})$. Normalizing flows (NFs) model a distribution $\mathbb{P}^Y$ of a target variable Y by transforming a simple base distribution $\mathbb{P}^U$ (e.g., standard normal) of a latent variable U through an invertible transformation $Y = f_\theta(U)$, where θ denotes learnable parameters [31]. In order to estimate the *conditional* density $\mathbb{P}^Y(\cdot \mid \mathbf{x}, \mathbf{m}, \mathbf{a})$, we leverage CNFs, that is, we define the parameters θ as an output of a *hyper network* $\theta = g_\eta(\mathbf{x}, \mathbf{m}, \mathbf{a})$ with learnable parameters η. Given a sample $(\mathbf{x}_i, \mathbf{m}_i, \mathbf{a}_i), y_i)_{i=1}^n$, we learn η by maximizing the log-likelihood

$$\ell(\eta) = \sum_{i=1}^{n} \log\left(f_{g_\eta(\mathbf{x}_i, \mathbf{m}_i, \mathbf{a}_i)\,\#}\mathbb{P}^U(y_i)\right) \tag{61}$$

$$\overset{(*)}{=} \sum_{i=1}^{n} \log\left(\mathbb{P}^U\left(f^{-1}_{g_\eta(\mathbf{x}_i, \mathbf{m}_i, \mathbf{a}_i)}(y_i)\right)\right) + \log\left(\left|\frac{\mathrm{d}}{\mathrm{d}y} f^{-1}_{g_\eta(\mathbf{x}_i, \mathbf{m}_i, \mathbf{a}_i)}(y_i)\right|\right), \tag{62}$$

where $f_{g_\eta(\mathbf{x}_i, \mathbf{m}_i, \mathbf{a}_i)\,\#}\mathbb{P}^U(y_i)$ denotes the (push-forward) density induced by $f_{g_\eta(\mathbf{x}_i, \mathbf{m}_i, \mathbf{a}_i)}$ on $\mathbb{R}$ and $(*)$ follows from the change-of-variables theorem for invertible transformations.

In our implementation, we use neural spline flows. That is, we model the invertible transformation f_θ via a spline flow as described in [9]. We use a feed-forward neural network for the hyper network $g_\eta(\mathbf{x}, \mathbf{m}, \mathbf{a})$ with 2 hidden layers, ReLU activation functions, and linear output. We set the latent distribution $\mathbb{P}^U$ to a standard normal distribution $\mathcal{N}(0, 1)$. For training, we use the Adam optimizer [22].

Conditional probability mass functions: The estimation of the conditional probability mass function $\mathbb{P}^{M_i}(\cdot \mid \mathbf{x}, \bar{\mathbf{m}}_{i-1}, \mathbf{a})$ is a standard (multi-class) classification problem. We use feed-forward neural networks with 3 hidden layers, ReLU activation functions, and softmax output. For training, we minimize the standard cross-entropy loss by using the Adam optimizer [22]. We use the same approach to estimate the propensity scores $\mathbb{P}^{\mathbf{A}}(\cdot \mid \mathbf{x})$ for discrete treatments $\mathbf{A}$.

Hyperparameter tuning:

We perform hyperparameter tuning for our experiments on synthetic data using grid search on a validation set. The tunable parameters and search ranges are shown in Table 2. For reproducibility purposes, we report the selected hyperparameters as *.yaml* files.[1]

Table 2: Hyperparameter tuning details.

Model	Tunable parameters	Search range
CNFs	Epochs	50
	Batch size	32, 64, 128
	Learning rate	0.0005, 0.001, 0.005
	Hidden layer size (hyper network)	5, 10, 20, 30
	Number of spline bins	2, 4, 8
Feed forward neural networks	Epochs	30
	Batch size	32, 64, 128
	Learning rate	0.0005, 0.001, 0.005
	Hidden layer size	5, 10, 20, 30
	Dropout probability	0, 0.1

[1] Code is available in the supplementary materials and at https://anonymous.4open.science/r/SharpCausalSensitivity-D87C.

G Experiments using synthetic data

Here we provide details regarding our experiments using synthetic data. This includes data generation, obtaining oracle sensitivity parameters, and details regarding experimental evaluation.

Overall data-generating process: We first describe the overall data-generating process which we use as a basis to generate data for all settings (i)-(iii) and binary/continuous treatments. We construct an SCM following the causal graph in Fig. 1 (right) from the main paper. We have an observed confounder $X \in \mathbb{R}$, a (binary or continuous) treatment A, two binary mediators M_1 and M_2, and a continuous outcome $Y \in \mathbb{R}$. Furthermore, we consider three unobserved confounders: (i) U_{M_1} confounding the A-M_1 relationship, (ii) U_{M_2} confounding the A–M_2 relationship, and (iii) U_Y confounding the A–Y relationship. Our data-generating process is inspired by synthetic experiments from previous works on causal sensitivity analysis [16, 18]. We start the data-generating process by sampling

$$X \sim \text{Uniform}[-1,1], \quad \text{and} \quad U_{M_1}, U_{M_2}, U_Y \overset{(i.i.d)}{\sim} \text{Bernoulli}(p=0.5) \tag{63}$$

Depending on the setting, we either generate binary treatments $A \in \{0,1\}$ via

$$A \sim \text{Bernoulli}(\mathrm{sigmoid}(3x + \gamma_{M_1} u_{M_1} + \gamma_{M_2} u_{M_2} + \gamma_Y u_Y)) \tag{64}$$

or continuous treatments $A \in (0,1)$ via

$$A \sim \text{Beta}(\alpha, \beta) \text{ with } \alpha = \beta = 2 + x + \gamma_{M_1}(u_{M_1} - 0.5) + \gamma_{M_2}(u_{M_2} - 0.5) + \gamma_Y(u_Y - 0.5), \tag{65}$$

where γ_{M_1}, γ_{M_2}, and γ_Y are parameters controlling the strength of unobserved confounding. We then generate the mediators and outcome via functional assignments

$$M_1 = f_{M_1}(X, A, U_{M_1}, \epsilon_{M_1}), \quad M_2 = f_{M_2}(X, A, M_1, U_{M_2}, \epsilon_{M_2}) \tag{66}$$

and

$$Y = f_Y(X, A, M_1, M_2 U_Y, \epsilon_Y), \tag{67}$$

where $\epsilon_{M_1}, \epsilon_{M_2}, \epsilon_Y \sim \mathcal{N}(0,1)$ are standard normal distributed noise variables. The functional assignments are defined as

$$f_{M_1}(x, a, u_{M_1}, \epsilon_{M_1}) = \mathbb{1}\left\{a \sin(x) + (1-a)\sin(4x) + \rho_{M_1}\left((u_{M_1} - 0.5) + \epsilon_{M_1}\right) > 0\right\} \tag{68}$$

for M_1,

$$\begin{aligned}
f_{M_2}(x, a, m_1 u_{M_2}, \epsilon_{M_2}) =& \mathbb{1}\{a\, m_1 \sin(x) + (1-a)\, m_1 \sin(4x) && (69)\\
& - a\,(1-m_1)\sin(x) - (1-a)\,(1-m_1)\sin(4x) && (70)\\
& + \rho_{M_2}\left((u_{M_2} - 0.5) + \epsilon_{M_2}\right) > 0\} && (71)
\end{aligned}$$

for M_2, and

$$\begin{aligned}
f_Y(x, a, m_1, m_2, u_Y, \epsilon_Y) =& a\, m_1\, m_2 \sin(x) + (1-a)\, m_1\, m_2 \sin(4x) && (72)\\
& + a\, m_1\,(1-m_2)\sin(8x) + (1-a)\, m_1\,(1-m_2)\sin(x) && (73)\\
& - a\,(1-m_1)\, m_2 \sin(x) - (1-a)\,(1-m_1)\, m_2 \sin(4x) && (74)\\
& - a\,(1-m_1)\,(1-m_2)\sin(8x) && (75)\\
& - (1-a)\,(1-m_1)\,(1-m_2)\sin(x) && (76)\\
& + \rho_Y\left((u_Y - 0.5) + \epsilon_Y\right) && (77)
\end{aligned}$$

for Y, where ρ_{M_1}, ρ_{M_2}, and ρ_Y are parameters that control the noise level.

Settings (i)-(iii): We define the settings (i)-(iii) in Sec. 5 via specific values of the confounding parameters γ_{M_1}, γ_{M_2}, and γ_Y, and the noise parameters ρ_{M_1}, ρ_{M_2}, and ρ_Y (see Table 3). Note that the settings are defined to mimic the causal graphs in Fig. 1 from the main paper. For example, the only unobserved confounder in setting (i) is U_Y, which means that we can ignore the mediators and use our data to evaluate our bounds for settings without mediators.

Table 3: Definition of settings (i)-(iii).

	γ_{M_1}	γ_{M_2}	γ_Y	ρ_{M_1}	ρ_{M_2}	ρ_Y
Setting (i), binary A	0	0	1.5	0.2	0.2	2
Setting (i), continuous A	0	0	1.5	0.2	0.2	1
Setting (ii)	1.5	0	1.5	1	0.2	1
Setting (iii)	1.5	1.5	1.5	0.2	0.2	1

Obtaining Γ_W^*: We provide details regarding our approach to obtain oracle sensitivity parameters Γ_W^* for all $W \in \{M_1, M_2, Y\}$. By sampling from our previously defined SCM we can obtain Monte Carlo estimates of the GMSM density ratio

$$r(u_W, x, a) = \frac{\mathbb{P}(u_W \mid x, a)}{\mathbb{P}(u_W \mid x, a)} \overset{(*)}{=} \frac{\mathbb{P}(a \mid x, u_W)}{\mathbb{P}(a \mid x)} \tag{78}$$

for all $u_W \in \{0, 1\}$, a, and x, where $(*)$ follows from Bayes' theorem. We then define

$$r_W^+(x, a) = \max_{u_W \in \{0,1\}} r(u_W, x, a) \quad \text{and} \quad r_W^-(x, a) = \min_{u_W \in \{0,1\}} r(u_W, x, a). \tag{79}$$

For binary treatment settings, we define parameters $\Gamma_W^+ = \Gamma_W^+(x, a)$ and $\Gamma_W^- = \Gamma_W^-(x, a)$ that attain the density ratio bounds in the MSM from Eq. (34), i.e.

$$r_W^+(x, a) = \frac{1}{(1 - {\Gamma_W^+}^{-1})\mathbb{P}(a \mid x) + {\Gamma_W^+}^{-1}} \quad \text{and} \quad r_W^-(x, a) = \frac{1}{(1 - \Gamma_W^-)\mathbb{P}(a \mid x) + \Gamma_W^-}. \tag{80}$$

For continuous treatment settings, we define Γ_W^+ and Γ_W^- as the sensitivity parameters that attain the density ratio bounds in the CMSM from Eq. (36), i.e.

$$r_W^+(x, a) = \Gamma_W^+ \quad \text{and} r_W^-(x, a) = \frac{1}{\Gamma_W^-}. \tag{81}$$

Finally, we define Γ_W^* as the parameter corresponding to the maximum possible violation of unconfoundedness, i.e.,

$$\Gamma_W^* = \max\{\Gamma_W^+, \Gamma_W^-\} \tag{82}$$

By definition of Γ_W^*, our bounds should contain the oracle causal effect whenever we choose sensitivity parameters $\Gamma_W \geq \Gamma_W^*$ for all $W \in \{M_1, M_2, Y\}$.

Weighted GMSM experiment (Table 1): For our experiment in Table 1, we modify the treatment assignment from Eq. (65) in setting (i) to

$$A \sim \mathrm{Beta}(\alpha, \beta) \tag{83}$$

with

$$\alpha = \beta = 2 + x + \mathbb{1}(x < 0)\left(\gamma_{M_1}(u_{M_1} - 0.5) + \gamma_{M_2}(u_{M_2} - 0.5) + \gamma_Y(u_Y - 0.5)\right). \tag{84}$$

Hence, unobserved confounding only affects individuals with $x < 0$. We then compare our bounds under the CMSM with our bounds under a weighted CMSM (Def. 5) with weight function $q_Y(x) = \mathbb{1}(x > 0)$.

We also provide results for the bounds from Jesson et al. [17] under the CMSM. We implemented the grid search algorithm from Jesson et al. [17] and used $5,000$ samples for the search space. For a fair comparison, we also used $5,000$ samples for our importance sampling estimators. Note that the method from Jesson et al. [17] requires estimation of both the conditional outcome density $\mathbb{P}^Y(\cdot \mid \mathbf{x}, \mathbf{a})$ and the conditional expectation $\mathbb{E}[Y \mid \mathbf{x}, \mathbf{a}]$. For $\mathbb{P}^Y(\cdot \mid \mathbf{x}, \mathbf{a})$, we use the same (normalizing flow-based) estimator as for our bounds. For $\mathbb{E}[Y \mid \mathbf{x}, \mathbf{a}]$, we train a separate feed-forward neural network with linear output activation for continuous outcomes. Implementation and hyperparameter tuning are done the same way as described in Appendix F for the feed-forward neural networks.

H Experiment using real-world data

Data: We consider a setting from the COVID-19 pandemic where mobility in Switzerland (captured through telephone movement) was monitored to obtain a leading predictor of case growth. In total, $\sim 1,5$ billion trips were monitored from 10 February through 26 April 2020. All data are recorded across 26 different states (cantons). For our analysis, we use an aggregated, de-identified, and pre-processed version of the data provided by Persson, Parie, and Feuerriegel [29]. The preprocessed data is publically available at https://github.com/jopersson/covid19-mobility/blob/main/Data. The code for our analysis is available at https://anonymous.4open.science/r/SharpCausalSensitivity-D87C.

We consider a binary treatment A in the form of a stay-at-home order, which bans gatherings with more than 5 people. We encode mobility as a single binary mediator M, which is 1 if the total number of trips on a specific day is larger than the median number of trips during the entire time horizon, and 0 otherwise. Our outcome is the 10-day-ahead case growth. We include the following observed variables as confounders $\mathbf{X}$: the canton code (swiss member state at a subnational level), the canton population, and whether the weekday is a Monday or not. After removing the first 10 recorded days for each canton (due to spillover effects from other countries) and rows with missing values, we obtain a dataset with $n = 3276$ observations.

Analysis: We perform a causal sensitivity analysis for the natural directed effect (NDE) of the stay-at-home order A on the case growth Y. That is, we are interested in the part of the causal effect of A on Y that is not explained by the path via M (i.e., through the change in mobility). The NDE in an SCM $\mathcal{M}$ is defined as

$$NDE(\mathcal{M}) = \int Q(\mathbf{x}, (a_0 = 0, a_1 = 1), \mathcal{M}) - Q(\mathbf{x}, (a_0 = 0, a_1 = 0), \mathcal{M}) \, \mathrm{d}\mathbf{x}. \tag{85}$$

Fig. 5 (main paper) shows causal sensitivity analysis for violations of the unconfoundedness between treatment A and mediator M. Hence, we consider a GMSM for binary treatments with sensitivity parameters Γ_M and $\Gamma_Y = 0$. For each Γ_M, we estimate our bounds for the expectation functional and the treatment combinations $\bar{\mathbf{a}} = (0, 1)$ and $\bar{\mathbf{a}} = (0, 0)$. We then obtain bounds for the NDE as described in Appendix D.

I Additional experimental results

Here, we provide additional experimental results on synthetic data that extend the results from Sec. 5 in the main paper. We provide (i) results for additional treatment combinations and (ii) results for distributional effects. We follow the same experimental setup described in Sec. 5 (main paper) and Appendix. G.

I.1 Additional treatment combinations

Results for additional treatment combinations are shown in Fig. 5 (binary treatment settings) and Fig. 6 (continuous treatment settings). The results are similar to those in Sec. 5 in the main paper and empirically confirm the validity of our bounds. Hence, our results remain valid independently of the choice of treatment combination.

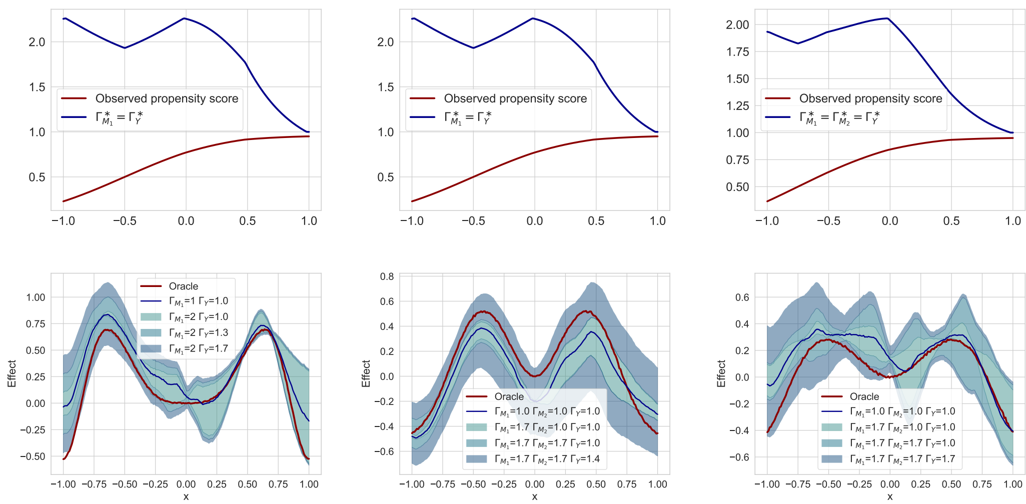

Figure 5: Results for additional treatments in the binary treatment setting. From left to right is shown: setting (ii) with $\bar{\mathbf{a}} = (0, 1)$, setting (iii) with $\bar{\mathbf{a}} = (0, 1, 0)$, and setting (iii) with $\bar{\mathbf{a}} = (0, 0, 1)$. The top row shows the oracle sensitivity parameter Γ^*_W (depending on x), and the bottom row shows the bounds.

I.2 Distributional effects

We also provide results for distributional effects, that is, we choose the α-quantile functional $\mathcal{D}\left(\mathbb{P}^Y_+(\cdot \mid \mathbf{x}, \mathbf{m}, \mathbf{a})\right) = F^{-1}_{\mathbb{P}^{Y \mid \mathbf{x},\mathbf{m},\mathbf{a}}_+}(\alpha)$. Here, we consider three quantiles with $\alpha = 0.7$, $\alpha = 0.5$ (median), and $\alpha = 0.3$. We use our importance sampling estimator derived in Appendix. E (Eq. 59) to estimate our bounds. The results are shown in Fig. 7 (binary treatment) and Fig. 8 (continuous treatment) for settings (i)-(iii) from Fig. 1 in the main paper. Again, our bounds cover the underlying oracle effect in regions where the chosen sensitivity parameters Γ_W are larger than the oracle sensitivity parameters Γ^*_W. This also confirms empirically the validity of our bounds for distributional effects.

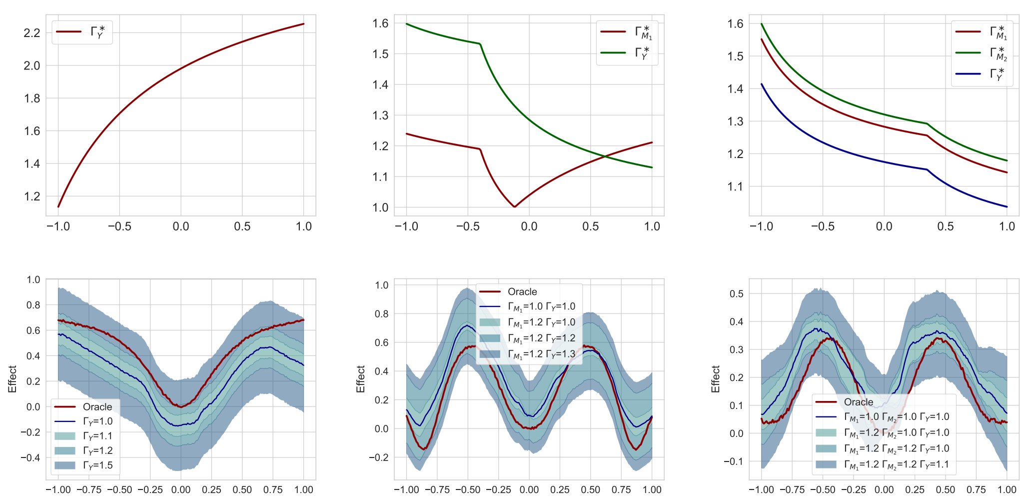

Figure 6: Results for additional treatments in the continuous treatment setting. From left to right is shown: setting (i) with $\bar{\mathbf{a}} = 0.9$, setting (ii) with $\bar{\mathbf{a}} = (0.2, 0.4)$, and setting (iii) with $\bar{\mathbf{a}} = (0.4, 0.5, 0.3)$. The top row shows the oracle sensitivity parameter Γ_W^* (depending on x), and the bottom row shows the bounds.

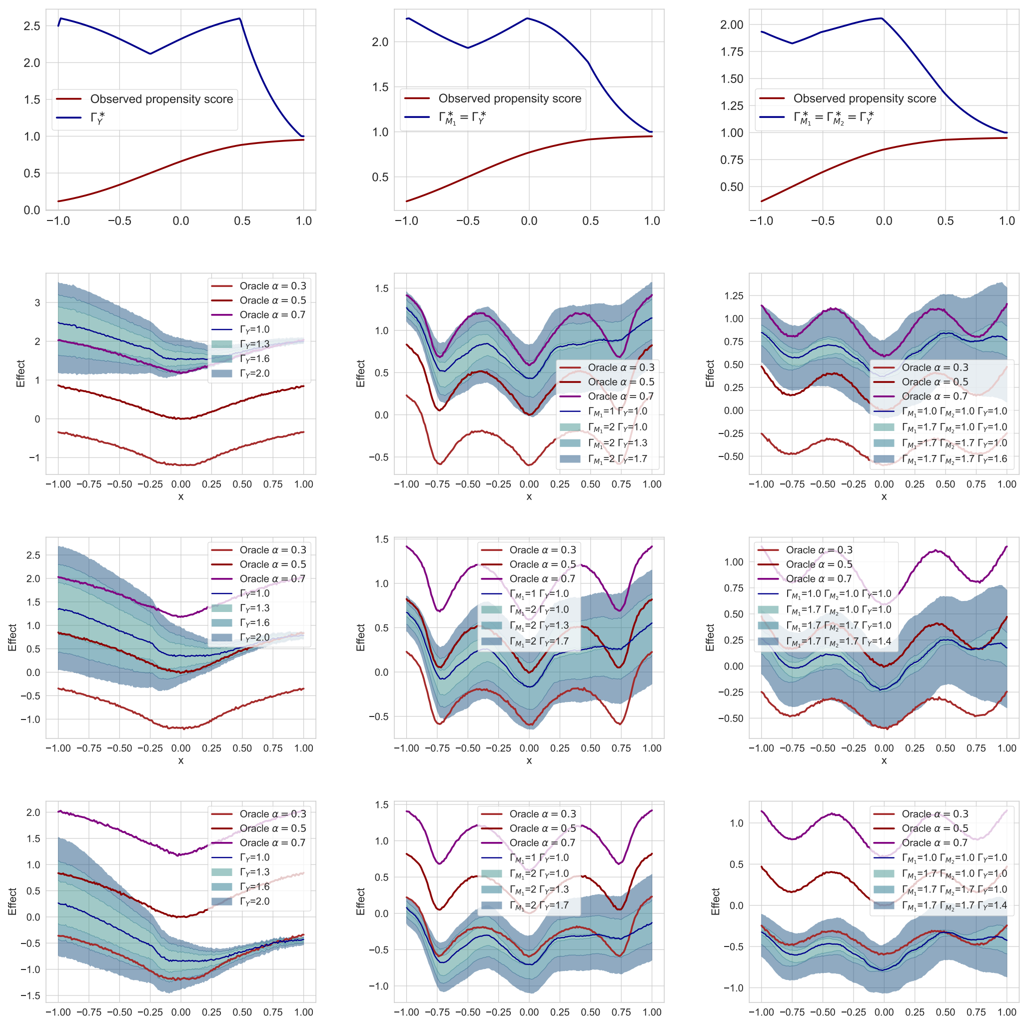

Figure 7: Results for distributional effects in the binary treatment setting using the same treatments as in Fig. 3 (main paper). Settings (i)–(iii) are ordered from left to right. The top row shows the oracle sensitivity parameter Γ_W^* (depending on x). Rows 2, 3, and 4 show the bounds for the α-quantiles of the interventional distribution with $\alpha = 0.7$, $\alpha = 0.5$, and $\alpha = 0.3$.

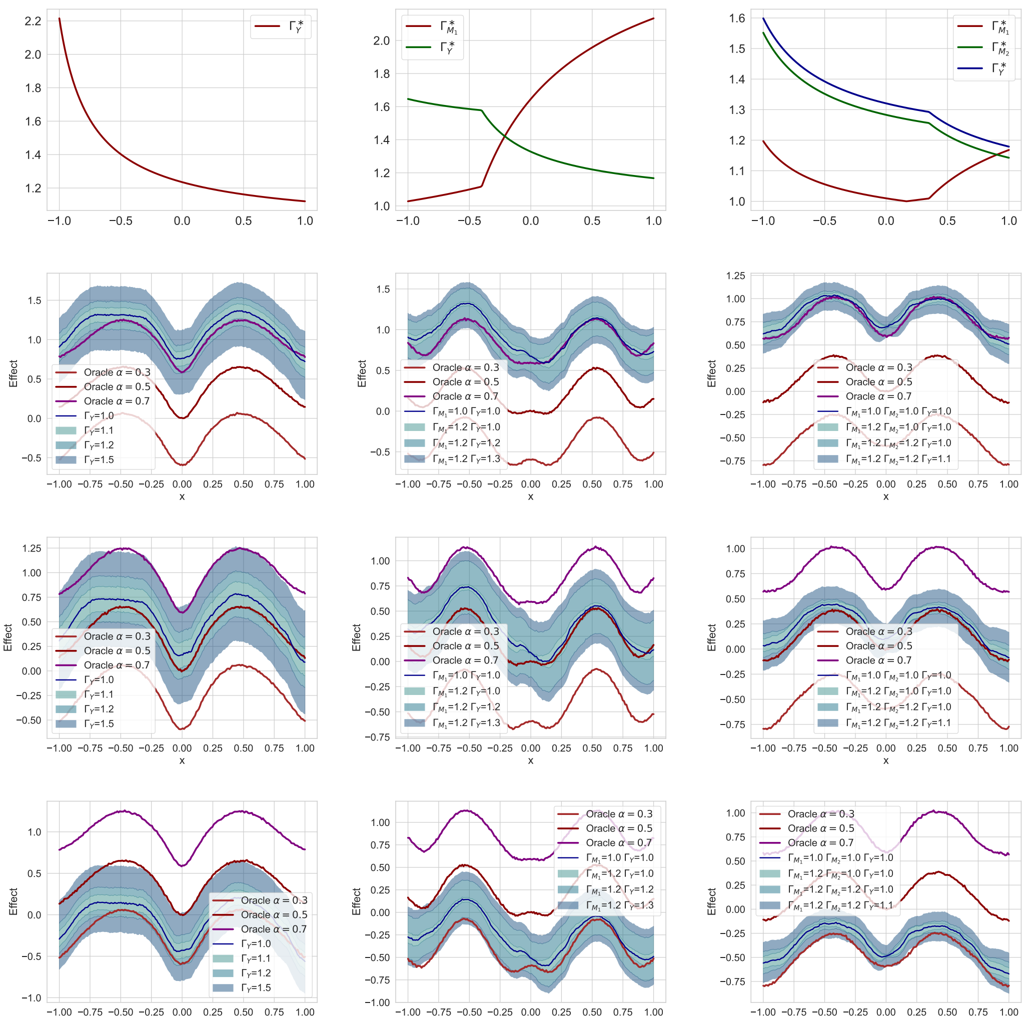

Figure 8: Results for distributional effects in the continuous treatment setting using the same treatments as in Fig. 4 (main paper). Settings (i)–(iii) are ordered from left to right. The top row shows the oracle sensitivity parameter Γ_W^* (depending on x). Rows 2, 3, and 4 show the bounds for the α-quantiles of the interventional distribution with $\alpha = 0.7$, $\alpha = 0.5$, and $\alpha = 0.3$.

References

[1] Vahid Balazadeh, Vasilis Syrgkanis, and Rahul G. Krishnan. "Partial identification of treatment effects with implicit generative models". In: *NeurIPS*. 2022.

[2] Andrew Bennett, Nathan Kallus, and Tobias Schnabel. "Deep generalized method of moments for instrumental variable analysis". In: *NeurIPS*. 2019.

[3] Ioana Bica et al. "Estimating counterfactual treatment outcomes over time through adversarially balanced representations". In: *ICLR*. 2020.

[4] Matteo Bonvini et al. "Sensitivity analysis for marginal structural models". In: *arXiv preprint* arXiv:2210.04681 (2022).

[5] Kan Chen, Bingkai Wang, and Dylan S. Small. "A differential effect approach to partial identification of treatment effects". In: *arXiv preprint* arXiv:2303.06332 (2023).

[6] Victor Chernozhukov, Ivan Fernández-Val, and Blaise Melly. "Inference on counterfactual distributions". In: *Econometrica* 81.6 (2013), pp. 2205–2268.

[7] Victor Chernozhukov et al. "Double/debiased machine learning for treatment and structural parameters". In: *The Econometrics Journal* 21.1 (2018), pp. C1–C68.

[8] Alicia Curth and Mihaela van der Schaar. "Nonparametric estimation of heterogeneous treatment effects: From theory to learning algorithms". In: *AISTATS*. 2021.

[9] Conor Durkan et al. "Neural spline flows". In: *NeurIPS*. 2019.

[10] Helmut Farbmacher et al. "Causal mediation analysis with double machine learning". In: *The Econometrics Journal* 25.2 (2022), pp. 277–300.

[11] Dennis Frauen and Stefan Feuerriegel. "Estimating individual treatment effects under unobserved confounding using binary instruments". In: *ICLR*. 2023.

[12] Dennis Frauen et al. "Estimating average causal effects from patient trajectories". In: *AAAI*. 2023.

[13] Florian Gunsilius. "A path-sampling method to partially identify causal effects in instrumental variable models". In: *arXiv preprint* arXiv:1910.09502 (2020).

[14] Wenshuo Guo et al. "Partial identification with noisy covariates: A robust optimization approach". In: *CLeaR*. 2022.

[15] Jason Hartford et al. "Deep IV: A flexible approach for counterfactual prediction". In: *ICML*. 2017.

[16] Andrew Jesson et al. "Quantifying ignorance in individual-level causal-effect estimates under hidden confounding". In: *ICML*. 2021.

[17] Andrew Jesson et al. "Scalable sensitivity and uncertainty analysis for causal-effect estimates of continuous-valued interventions". In: *NeurIPS*. 2022.

[18] Nathan Kallus, Xiaojie Mao, and Angela Zhou. "Interval estimation of individual-level causal effects under unobserved confounding". In: *AISTATS*. 2019.

[19] Edward H. Kennedy. "Towards optimal doubly robust estimation of heterogeneous causal effects". In: *arXiv preprint* (2022).

[20] Edward H. Kennedy, Sivaraman Balakrishnan, and Larry Wasserman. "Semiparametric counterfactual density estimation". In: *Biometrika* (2023).

[21] Niki Kilbertus, Matt J. Kusner, and Ricardo Silva. "A class of algorithms for general instrumental variable models". In: *NeurIPS*. 2020.

[22] Diederik P. Kingma and Jimmy Ba. "Adam: A method for stochastic optimization". In: *ICLR*. 2015.

[23] Sören R. Künzel et al. "Metalearners for estimating heterogeneous treatment effects using machine learning". In: *Proceedings of the National Academy of Sciences (PNAS)* 116.10 (2019), pp. 4156–4165.

[24] Bryan Lim, Ahmed M. Alaa, and Mihaela van der Schaar. "Forecasting treatment responses over time using recurrent marginal structural networks". In: *NeurIPS*. 2018.

[25] Valentyn Melnychuk, Dennis Frauen, and Stefan Feuerriegel. "Causal transformer for estimating counterfactual outcomes". In: *ICML*. 2022.

[26] Valentyn Melnychuk, Dennis Frauen, and Stefan Feuerriegel. "Normalizing flows for interventional density estimation". In: *ICML*. 2023.

[27] Krikamol Muandet et al. “Counterfactual mean embeddings”. In: *Journal of Machine Learning Research* 22 (2021), pp. 1–71.

[28] Kirtan Padh et al. “Stochastic causal programming for bounding treatment effects”. In: *CLeaR*. 2023.

[29] Joel Persson, Jurriaan F. Parie, and Stefan Feuerriegel. “Monitoring the COVID-19 epidemic with nationwide telecommunication data”. In: *Proceedings of the National Academy of Sciences of the United States of America* 118.26 (2021).

[30] Jonas Peters, Dominik Janzig, and Bernhard Schölkopf. *Elements of causal inference: Foundations and learning algorithms*. Cambridge, Massachusetts: MIT Press, 2017.

[31] Danilo Jimenez Rezende and Shakir Mohamed. “Variational inference with normalizing flows”. In: *ICML*. 2015.

[32] Uri Shalit, Fredrik D. Johansson, and David Sontag. “Estimating individual treatment effect: Generalization bounds and algorithms”. In: *ICML*. 2017.

[33] Claudia Shi, David M. Blei, and Victor Veitch. “Adapting neural networks for the estimation of treatment effects”. In: *NeurIPS*. 2019.

[34] Rahul Singh, Maneesh Sahani, and Arthur Gretton. “Kernel instrumental variable regression”. In: *NeurIPS*. 2019.

[35] Vasilis Syrgkanis et al. “Machine learning estimation of heterogeneous treatment effects with instruments”. In: *NeurIPS*. 2019.

[36] Zhiqiang Tan. “A distributional approach for causal inference using propensity scores”. In: *Journal of the American Statistical Association* 101.476 (2006), pp. 1619–1637.

[37] Eric J. Tchetgen Tchetgen and Ilya Shpitser. “Semiparametric theory for causal mediation analysis: Efficiency bounds, multiple robustness, and sensitivity analysis”. In: *Annals of Statistics* 40.3 (2012), pp. 1816–1845.

[38] Mark J. van der Laan and Donald B. Rubin. “Targeted maximum likelihood learning”. In: *The International Journal of Biostatistics* 2.1 (2006).

[39] Stefan Wager and Susan Athey. “Estimation and inference of heterogeneous treatment effects using random forests”. In: *Journal of the American Statistical Association* 113.523 (2018), pp. 1228–1242.

[40] Christina Winkler et al. “Learning likelihoods with conditional normalizing flows”. In: *arXiv preprint* arXiv:1912.00042 (2019).

[41] Kevin Xia et al. “The causal-neural connection: Expressiveness, learnability, and inference”. In: *NeurIPS*. 2021.

[42] Liyuan Xu et al. “Learning deep features in instrumental variable regression”. In: *ICLR*. 2021.

[43] Jinsung Yoon, James Jordon, and Mihaela van der Schaar. “GANITE: Estimation of individualized treatment effects using generative adversarial nets”. In: *ICLR*. 2018.