Rescuing double robustness: safe estimation under complete misspecification

Lorenzo Testa[1,2], Francesca Chiaromonte[2,3], and Kathryn Roeder[1,4]

[1]Department of Statistics & Data Science, Carnegie Mellon University, Pittsburgh PA, US
[2]L'EMbeDS, Sant'Anna School of Advanced Studies, Pisa, Italy
[3]Department of Statistics, Penn State University, University Park PA, US
[4]Department of Computational Biology, Carnegie Mellon University, Pittsburgh PA, US

lorenzo@stat.cmu.edu fxc11@psu.edu roeder@andrew.cmu.edu

September 29, 2025

Abstract

Double robustness is a major selling point of semiparametric and missing data methodology. Its virtues lie in protection against partial nuisance misspecification and asymptotic semiparametric efficiency under correct nuisance specification. However, in many applications, complete nuisance misspecification should be regarded as the norm (or at the very least the expected default), and thus doubly robust estimators may behave fragilely. In fact, it has been amply verified empirically that these estimators can perform poorly when all nuisance functions are misspecified. Here, we first characterize this phenomenon of *double fragility*, and then propose a solution based on *adaptive correction clipping* (`DR+ACC`). We argue that our `DR+ACC` proposal is *safe*, in that it inherits the favorable properties of doubly robust estimators under correct nuisance specification, but its error is guaranteed to be bounded by a convex combination of the individual nuisance model errors, which prevents the instability caused by the compounding product of errors of doubly robust estimators. We also show that our proposal provides *valid* inference through the parametric bootstrap when nuisances are well-specified. We showcase the efficacy of our `DR+ACC` estimator both through extensive simulations and by applying it to the analysis of Alzheimer's disease proteomics data.

1 Introduction

Scientific progress relies on stable and reproducible evidence, which in turn demands solid statistical methodology (Yu and Barter, 2024; Yu and Kumbier, 2020). In recent decades, a paradigm shift has occurred in various scientific fields, spurred by ideas and concepts from causal inference and missing data analysis, moving from a "bottom-up" approach, where statistical models are posited first and their parameters interpreted post-hoc, to a "top-down" approach (Kennedy, 2024). This framework begins by defining a specific target of inquiry – the estimand – and then systematically lays out the assumptions and methods required to identify and estimate it from data.

Within this top-down paradigm, doubly robust (DR) estimators have emerged as a particularly powerful and celebrated tool (Bang and Robins, 2005; Robins et al., 1994; Scharfstein et al., 1999). Their appeal lies in two key properties. First, they offer protection against partial model misspecification: they yield a consistent estimate of the target parameter if at least one of the two required nuisance models – typically an outcome regression model and a propensity score model – is correctly specified. Second, if both nuisance models are correct, DR estimators achieve semiparametric efficiency, meaning they have the smallest possible asymptotic variance. These virtues have made them a cornerstone of semiparametric and missing data methodology.

However, the theoretical protection of double robustness can be misleading in practice. In applied research, it is often more realistic to assume that *all* models are misspecified to some degree, rather than expecting one to be perfectly correct (even asymptotically). In this scenario of complete nuisance misspecification, the guarantees of DR estimators vanish. Worse, they can become fragile. It has been empirically observed that DR estimators can perform poorly in this setting, sometimes exhibiting more bias than simpler estimators that rely on only one of the misspecified nuisance functions (Kang and Schafer, 2007). We term this phenomenon *double fragility*: the very mechanism designed to provide robustness can, under complete misspecification, amplify estimation errors and lead to unstable results.

This paper provides a formal characterization of double fragility, focusing on the dual nature of the correction term in doubly robust estimators. When at least one of the nuisance models is correct, this term is beneficial, guiding the estimator toward the true parameter. Conversely, when both nuisance models are misspecified, we show that their errors can compound within the correction term, actively degrading the estimator performance rather than improving it. This analysis reveals the source of fragility and directly motivates our solution: *adaptive correction clipping* (DR+ACC). Our proposal is built on the principle of *safety* – the property that an estimator performs no worse than the simpler estimators from which it is constructed (Xu et al., 2025). By adaptively clipping the correction term, our DR+ACC method preserves the consistency and efficiency of standard DR estimators in ideal settings, while crucially preventing the correction from amplifying bias when all models are misspecified.

While the non-linear nature of the adaptive correction clipping operator implies that our final estimator does not converge to a normal distribution, we show that *valid* confidence intervals can be constructed using a parametric bootstrap procedure. This approach works by simulating the joint asymptotic distribution of the estimator components, applying the adaptive clipping transformation to these simulations, and then using the empirical quantiles of the resulting distribution to form the interval.

To complement these theoretical and methodological results, we provide extensive empirical evidence of our DR+ACC estimator effectiveness. First, we conduct a simulation study that fully replicates the design of Kang and Schafer (2007), a well-known benchmark in the missing data literature. This setup is specifically designed to assess estimator performance across several scenarios of nuisance model specification, allowing for a rigorous evaluation of both fragility and safety. Second, we showcase the practical utility of our method by applying it to a substantive scientific problem: estimating the average treatment effect (ATE) of Alzheimer's disease on hundreds of peptide abundances, using data from Merrihew et al. (2023). This application demonstrates the tangible benefits of our safe estimator in a real-world setting where the risk of model misspecification is high.

1.1 A preview of our results

Before diving into the details of our approach, we provide a brief sketch of what can go wrong with DR, and some intuition on our proposal. Assume that we observe a sample of n independent and identically distributed random variables $\{\mathscr{D}_i = (X_i, R_i, R_i Y_i)\}_{i=1}^n$. Here, X_i is a vector of covariates, R_i is a binary indicator equal to 1 if Y_i is observed and 0 otherwise, and the outcome Y_i is only observed when $R_i = 1$. We let $\mathscr{D} = (X, R, RY)$ denote an independent copy of $\mathscr{D}_i = (X_i, R_i, R_i Y_i)$. We operate under the standard assumptions of missing at random ($Y \perp\!\!\!\perp R \mid X$) and positivity. Our goal is to estimate the target parameter

$$\theta^\star = \mathbb{E}[Y] = \mathbb{E}\left[\mu^\star(X) + \frac{R}{\pi^\star(X)}(Y - \mu^\star(X))\right], \tag{1}$$

where $\mu^\star(x) = \mathbb{E}[Y \mid X = x, R = 1]$ denotes the nuisance regression function and $\pi^\star(x) = \mathbb{P}[R = 1 \mid X = x]$ denotes the nuisance propensity score. Assuming we have access to pre-trained or externally fitted models for these functions, denoted $\hat{\mu}$ and $\hat{\pi}$ respectively, the well-known DR estimator can be defined as

$$\hat{\theta}_{\texttt{DR}} = \frac{1}{n}\sum_{i=1}^n \left[\hat{\mu}(X_i) + \frac{R_i}{\hat{\pi}(X_i)}(Y_i - \hat{\mu}(X_i))\right], \tag{2}$$

The DR estimator is particularly compelling because it can be algebraically decomposed into:

$$\hat{\theta}_{\texttt{DR}} = \underbrace{\frac{1}{n}\sum_{i=1}^n \hat{\mu}(X_i)}_{\hat{\theta}_{\texttt{OR}}} + \underbrace{\frac{1}{n}\sum_{i=1}^n \frac{R_i Y_i}{\hat{\pi}(X_i)}}_{\hat{\theta}_{\texttt{IPW}}} - \underbrace{\frac{1}{n}\sum_{i=1}^n \frac{R_i \hat{\mu}(X_i)}{\hat{\pi}(X_i)}}_{\text{correction}}, \tag{3}$$

which sheds light on the fact that the DR estimator is a combination of the simpler *outcome regression* (OR) and *inverse probability weighting* (IPW) estimators, respectively defined as

$$\hat{\theta}_{\texttt{OR}} = \frac{1}{n}\sum_{i=1}^n \hat{\mu}(X_i), \quad \hat{\theta}_{\texttt{IPW}} = \frac{1}{n}\sum_{i=1}^n \frac{R_i Y_i}{\hat{\pi}(X_i)}. \tag{4}$$

This simple algebraic structure has profound consequences on the property of the doubly robust estimator. Its correction term is engineered to ensure consistency and, ultimately, efficiency. This behavior can be seen in two key scenarios:

- Under partial misspecification, the term acts as a safeguard. For instance, if the outcome model $\hat{\mu}$ is correct but the propensity model $\hat{\pi}$ is not, the correction term is designed to asymptotically cancel the biased IPW component. As a result, the DR estimator converges to the same correct limit as the OR estimator. A symmetric cancellation occurs if $\hat{\pi}$ is correct instead.
- Under full correctness, when both nuisance models are well-specified, all three components of the estimator – OR, IPW, and the correction term – converge to the true parameter. In this ideal case, the correction term is effectively free to cancel either the OR or the IPW component, a flexibility that leads not only to consistency but also to semiparametric efficiency.

While this property is widely known as double robustness, we argue it can be better understood as a form of *asymptotic hard thresholding*, where the estimator effectively selects a valid component through the correction term. The first three panels of Figure 1 provide an empirical assessment of this principle. There, we show the distribution of the OR, IPW and DR estimators across 1000 replications (with $n = 1000$) under

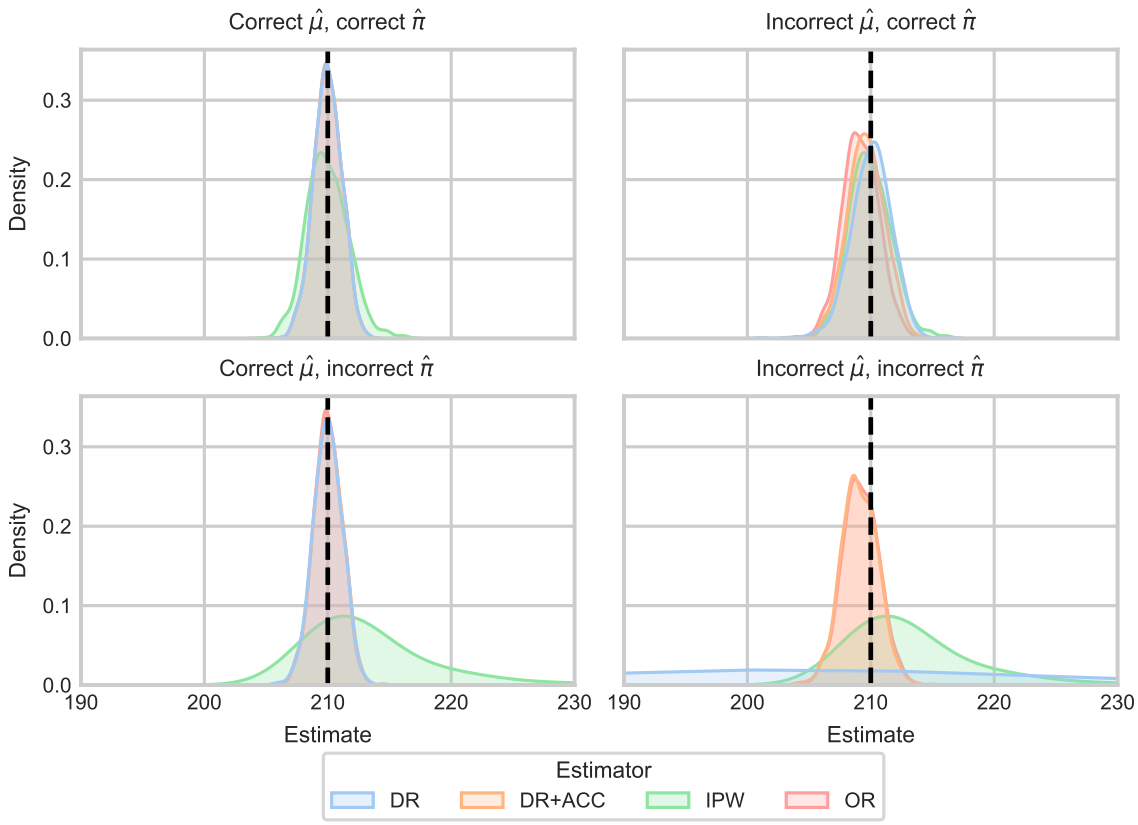

Figure 1: Sampling distributions for the Outcome Regression (`OR`), Inverse Probability Weighting (`IPW`), Doubly Robust (`DR`), and our proposed `DR+ACC` estimators from 1000 simulations with a sample size of $n = 1000$. The true parameter value is 210. The four panels show the estimators' performance under all combinations of correct and incorrect nuisance model specifications. The top row and bottom-left panel demonstrate the *asymptotic hard thresholding* property. When at least one nuisance model is correct, both `DR` and `DR+ACC` align with the correct simpler estimator. The bottom-right panel illustrates *double fragility*. When both nuisance models are wrong, the bias of the standard `DR` estimator is substantially worse than that of either the `OR` or `IPW` estimators. In this challenging scenario, our proposed `DR+ACC` estimator is shown to be *safe*, providing a much more stable and accurate estimate than the standard `DR` estimator. Full details of the simulation scenario are provided in Section 4.

scenarios where either both or at least one of the nuisance models are correct. As the panels illustrate, the `DR` estimator distribution aligns with that of the correctly specified component, thus ignoring the misspecified one. Although the full details of the simulation design are deferred to Section 4, these results offer a practical validation of the hard thresholding mechanism in action.

The problem arises when both nuisance models are misspecified – a common, if not ubiquitous, scenario in applied research. In this case, the asymptotic cancellation fails. The correction term, now a function of two incorrect models, can compound their errors in unpredictable ways, potentially introducing substantial bias that pushes the final estimate far from the true value. We call this failure mode *double fragility*. Instead of reducing bias, the correction term can exacerbate it. This phenomenon is illustrated in the final panel of Figure 1. Under complete model misspecification, the standard `DR` estimator is not only biased but also performs considerably worse than either of its simpler `OR` and `IPW` components.

To address this, we propose a simple yet effective solution: *adaptive correction clipping* (DR+ACC). Our method works by constraining, or "clipping", the correction term to ensure the final estimate is always anchored within the range defined by the simpler OR and IPW estimates. This modification enforces a safety property: it inherits the favorable properties of the standard DR estimator when its assumptions hold, but it is guaranteed to perform no worse than its constituent estimators when they fail. As Figure 1 demonstrates, the performance of our DR+ACC estimator is nearly identical to, if not better than, that of the standard DR when at least one nuisance model is correct. However, in the double fragility scenario, it substantially outperforms the standard DR estimator, validating its role as a safe and robust alternative.

1.2 Related work

Here, we discuss the relationship between our work and closely related scholarship.

Semiparametric statistics and missing data. From a technical standpoint, our work is grounded in the field of semiparametric statistics, which focuses on estimation in the presence of high-dimensional nuisance parameters. Missing data literature provides a rich toolkit for such problems, including methods such as Inverse Probability Weighting (IPW), augmented IPW, and targeted maximum likelihood estimation, many of which lead to doubly robust estimators (Hernán and Robins, 2006; Horvitz and Thompson, 1952; Van Der Laan and Rubin, 2006). The foundational concepts of influence functions, tangent spaces, and semiparametric efficiency bounds – developed in seminal works by Bickel et al. (1993); Tsiatis (2006); Van der Vaart (2000) – serve as the theoretical backbone for much of modern statistical learning. It is important to note, however, that most of these classical results are based on asymptotic arguments, which may not always hold in finite samples. Therefore, recent literature has increasingly focused on finite-sample performance of semiparametric estimators. Mou et al. (2022) analyze two-stage procedures common in causal inference and prove non-asymptotic upper bounds on the mean-squared error, revealing that to achieve optimality in finite samples, the error in estimating the nuisance function should be minimized in a specific weighted $\mathbb{L}_2$-norm. Similarly, Wang et al. (2024) provide a finite-sample analysis of doubly robust estimators using PAC-style guarantees, demonstrating that minimizing the estimation error of the treatment effect in terms of Chi-square distance is crucial for minimizing the final estimator variance. For the widely-used augmented inverse probability weighting (AIPW) estimator, Wang and Deng (2023) explore its non-asymptotic properties, showing it can achieve near-oracle performance even when the nuisance models converge at a slow, non-parametric rate. This focus on non-asymptotic guarantees has also extended to specific, challenging scenarios. Ghadiri et al. (2023) address the historical lack of non-asymptotic accuracy bounds for treatment effect estimation in the finite population setting, while Celentano and Wainwright (2023) tackle the high-dimensional $n < p$ "inconsistency regime", where they develop a novel procedure to achieve consistency when standard methods fail. Complementing these theoretical advances, empirical studies provide practical insights; for instance, Witter and Musco (2024) used a novel benchmarking framework to show that simpler, doubly robust estimators often outperform more complicated methods in practice, a finding which in turn spurred new theory on the finite-sample variance of these estimators.

Stability, safety, and robust statistics. The principles of stability and robustness are crucial for developing reliable and reproducible statistical methods. Classical robust statistics, with foundational work by Hampel (1974); Hampel et al. (2011); He and Portnoy (1992); Huber (1996), introduced concepts like the influence function to create estimators that are insensitive to outliers or deviations from assumed data distributions. More recently, this idea has been broadened to the concept of stability, which encompasses the entire data

science life cycle, emphasizing the importance of methods being resilient to perturbations in data, models, and analytical choices (Agarwal et al., 2025; Rewolinski and Yu, 2025; Yu, 2013; Yu and Barter, 2024; Yu and Kumbier, 2020). The development of tools like the *s-value* for evaluating stability against distributional shifts further highlights the field's focus on this property (Gupta and Rothenhäusler, 2023). Our work connects directly to these developments by identifying a specific instability in doubly robust estimators, which we term double fragility. The safety property we propose is a specific form of stability tailored to this problem, guaranteeing that the estimator is robust against the complete misspecification of its underlying nuisance models (Deng et al., 2024; Xu et al., 2025). We also want to note that the finite-sample instability of doubly robust estimators is a well-recognized challenge. A common *ad hoc* adjustment is propensity score *trimming*, where estimated probabilities of treatment are bounded away from 0 and 1 to prevent the inverse weights from becoming excessively large (Ma et al., 2023; Stürmer et al., 2021). Another approach is self-normalization, which is used in the Hájek estimator (Basu, 1971). Cai et al. (2024) recently proposed the C-Learner, which reframes the construction of a debiased estimator as a constrained optimization problem. Instead of first training a nuisance model to optimize prediction accuracy and then applying a *post hoc* correction, the C-Learner directly trains the outcome model to minimize prediction error subject to the constraint that the first-order error term is zero.

Semi-supervised learning and prediction-powered inference. Our setup, which we will introduce shortly, mimics the one usually studied in semi-supervised learning. For instance, Zhang et al. (2019); Zhang and Bradic (2022) have focused on semi-supervised mean estimation, including extensions to high-dimensional settings with bias-corrected inference. The problem of semi-supervised linear regression has also been explored in depth, leading to the development of asymptotically normal estimators with improved efficiency and minimax optimal estimators in high dimensions, with subsequent extensions to generalized linear models (Azriel et al., 2022; Chakrabortty and Cai, 2018). The scope of semi-supervised has broadened to encompass more general inferential tasks like M-estimation and U-statistics (Chakrabortty, 2016; Kim et al., 2024; Testa et al., 2025). A closely related paradigm is prediction-powered inference (PPI), where an analyst leverages a pre-trained, black-box machine learning model in addition to labeled and unlabeled data (Angelopoulos et al., 2023a,b; Zrnic and Candès, 2024).

1.3 Roadmap

The remainder of this paper is organized as follows. First, in Section 2 we introduce our setup and notation, and then we analyze the behavior of standard doubly robust estimators, formally characterizing the *double fragility* phenomenon that arises under complete nuisance model misspecification. Then, in Section 3 we introduce our proposed solution, `DR+ACC`, an estimator that uses *adaptive correction clipping* as a safeguard against this fragility. Here, we prove the consistency of our estimator and detail a parametric bootstrap procedure for constructing valid confidence intervals when nuisance models are well-specified. In Section 4 we provide empirical validation through an extensive simulation study, and in Section 5 we demonstrate the practical utility of our method with an application to the analysis of peptide abundance data in an Alzheimer's study. Section 6 contains some concluding remarks. Additional theoretical results and simulation details are provided in the Supplementary Material. All code for reproducing our analysis is available at `https://github.com/testalorenzo/DoubleFragility`.

2 Problem setup and review of double robustness

2.1 Setup and notation

Following traditional nomenclature from semiparametric statistics literature, we cast our problem in a missing data framework. We denote the *observed data* as $\{\mathscr{D}_i = (X_i, R_i, R_i Y_i)\}_{i=1}^n$, where R_i is a binary indicator that indicates whether observation $\mathscr{D}_i$ is labeled ($R_i = 1$), or unlabeled ($R_i = 0$). We let $\mathscr{D} = (X, R, RY)$ denote an independent copy of $\mathscr{D}_i = (X_i, R_i, R_i Y_i)$. We denote the data-generating distribution as $\mathbb{P}^\star \in \mathscr{P}$, where $\mathscr{P}$ is the set of distributions induced by a nonparametric model, and thus we write $\mathscr{D} \sim \mathbb{P}^\star$. For theoretical convenience, we also define the *full data* $\mathscr{D}^F = (X, Y)$, that is, the data that we would observe if there were no missingness mechanisms in place.

The potential discrepancy between the labeled and unlabeled datasets due to *distribution shift* naturally leads us to adopt a *missing at random* (MAR) labeling mechanism. We formally state this assumption below, along with a weak overlap assumption – as requirements for identifiability of the target parameter.

Assumption 2.1 (Identifiability)**.** Let the following assumptions hold:

a. Missing at random. $R \perp\!\!\!\perp Y \mid X$.

b. Weak overlap. $\pi(x) = \mathbb{P}[R = 1 \mid X = x] \in (0, 1)$ for all $x \in \mathbb{R}^p$ almost surely.

Throughout this paper, we focus on the estimation of a target quantity $\theta^\star \in \mathbb{R}$ that is defined as the functional solving $\theta^\star = \theta(\mathbb{P}^\star)$, with $\theta : \mathscr{P} \to \mathbb{R}$. We restrict our attention to target parameters that can be estimated using *regular* and *asymptotically linear observed-data* estimators, that is, targets that admit the expansion

$$\sqrt{n}\left(\hat{\theta} - \theta^\star\right) = n^{-1/2} \sum_{i=1}^{n} \varphi\left(\mathscr{D}_i; \theta^\star\right) + o_{\mathbb{P}}(1), \tag{5}$$

for some regular and asymptotically linear *observed-data* estimator $\hat{\theta}$, and some function $\varphi(\mathscr{D}_i; \theta^\star)$, referred to as *observed-data influence function*, evaluating the contribution of the *observed-data* sample $\mathscr{D}_i$ to the overall estimator $\hat{\theta}$ (Hampel, 1974). This framework is very general and encompasses M-estimation, Z-estimation, and many estimands in causal inference (Kennedy, 2024; Van der Vaart, 2000).

In a nonparametric setting, semiparametric theory shows that for any given target parameter, there is a unique and efficient *full-data influence function*. This theoretical building block can be projected onto the observed data to derive the *efficient observed-data influence function*, which forms the basis of many estimators encountered in practice. In particular, given a full-data influence function $\varphi^F\left(\mathscr{D}^F; \theta^\star\right)$, one can recover the efficient observed-data influence function using the following Lemma.

Lemma 2.2 (Observed-data influence function)**.** *Let $\theta^\star$ be a target parameter that admits a regular and asymptotically linear (RAL) expansion as in Eq. 5, and let $\varphi^F\left(\mathscr{D}^F; \theta^\star\right)$ be the full-data influence function associated to it. Assume the identifiability conditions described in Assumption 2.1. Then, the observed-data influence function for the target $\theta^\star$ is given by:*

$$\varphi(\mathscr{D}; \theta^\star) = \mu^\star(X) + \frac{R}{\pi^\star(X)}\left(\varphi^F(\mathscr{D}; \theta^\star) - \mu^\star(X)\right), \tag{6}$$

where $\mu^\star(x) = \mathbb{E}\left[\varphi^F\left(\mathscr{D}^F; \theta^\star\right) \mid X = x, R = 1\right]$ is the nuisance regression function, and $\pi^\star : \mathbb{R}^p \to (0, 1)$ denotes the nuisance propensity score, defined as $\pi^\star(x) = \mathbb{P}[R = 1 \mid X = x]$.

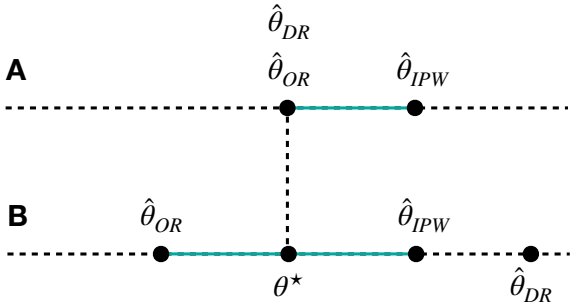

Figure 2: A conceptual illustration of the behavior of the doubly robust estimator under different model specification scenarios. Scenario A depicts the case of partial misspecification, where the `OR` estimator is consistent for $\theta^\star$ but the `IPW` estimator is biased. The `DR` estimator correctly aligns with the consistent `OR` estimator, demonstrating its asymptotic hard thresholding property. Scenario B depicts the case of complete misspecification, where both the `OR` and `IPW` estimators are biased. This illustrates double fragility: the bias of the `DR` estimator can be substantially worse than that of its constituent estimators, as its correction term can amplify, rather than reduce, the overall error.

In observational studies, the nuisance regression function $\mu^\star$ and the propensity score $\pi^\star$ are unknown and must be estimated from the data. For simplicity, we will proceed by assuming access to pre-trained models, denoted $\hat{\mu}$ and $\hat{\pi}$. However, our results can be readily extended to the practical setting where these nuisance functions are estimated from the data using techniques like *sample splitting* or *cross-fitting*. Sample splitting works as follows. We randomly split the observations $\{\mathscr{D}_1, \ldots, \mathscr{D}_n\}$ into 2 disjoint folds. We form $\hat{\mathbb{P}}$ with the first fold, and $\mathbb{P}_n$ with the second fold. Then, we learn $\hat{\mu}$ and $\hat{\pi}$ on $\hat{\mathbb{P}}$, and we compute the estimator $\hat{\theta}_{\texttt{DR}}$ by solving for $\theta^\star$ the estimating equation

$$\sum_{i \in \mathbb{P}_n} \varphi\left(\mathscr{D}_i; \theta^\star; \hat{\mu}; \hat{\pi}\right) = 0\,. \tag{7}$$

This separation of training and estimation prevents overfitting and is crucial for valid inference. The resulting estimator takes the form

$$\hat{\theta}_{\texttt{DR}} = \frac{1}{|\mathbb{P}_n|} \sum_{i \in \mathbb{P}_n} \left[\hat{\mu}(X_i) + \frac{R_i}{\hat{\pi}(X_i)} \left(\varphi^F\left(\mathscr{D}_i; \theta^\star\right) - \hat{\mu}(X_i) \right) \right], \tag{8}$$

where $|\mathbb{P}_n|$ denotes the cardinality of $\mathbb{P}_n$. By assuming access to external nuisance models, the averaging distribution $\mathbb{P}_n$ contains the entire sample at hand, so that $|\mathbb{P}_n| = n$.

2.2 What is robust in double robustness?

The celebrated *double robustness* property of estimators as in Eq. 8 is foundational to modern semiparametric statistics. This Section revisits this concept, first through a novel framing that explains its mechanism and then by analyzing the failure mode that motivates our work.

Classically, double robustness refers to the property that an estimator remains consistent if at least one of its two nuisance functions (the outcome regression $\hat{\mu}$ or the propensity score $\hat{\pi}$) is correctly specified. We argue that this behavior can be more mechanistically understood as a form of *asymptotic hard thresholding*.

In fact, the DR estimator in Eq. 8 can be written as:

$$\hat{\theta}_{\texttt{DR}} = \hat{\theta}_{\texttt{OR}} + \hat{\theta}_{\texttt{IPW}} - \frac{1}{|\mathbb{P}_n|}\sum_{i\in\mathbb{P}_n}\frac{R_i\hat{\mu}(X_i)}{\hat{\pi}(X_i)}\,. \tag{9}$$

When one nuisance model is correct, the estimator correction term is engineered to asymptotically cancel the bias introduced by the misspecified model, effectively forcing the estimator to rely solely on the correctly specified component. When both models are correct, this flexibility allows the estimator to achieve optimal semiparametric efficiency. While Figure 2 depicts this intuition, the following Proposition formalizes it.

Proposition 2.3. *Assume the identifiability conditions described in Assumption 2.1. If $\hat{\mu} = \mu^\star$, then*

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\frac{R_i\hat{\mu}(X_i)}{\hat{\pi}(X_i)}\right] = \mathbb{E}\left[\hat{\theta}_{\texttt{IPW}}\right]. \tag{10}$$

Similarly, if $\hat{\pi} = \pi^\star$, then

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\frac{R_i\hat{\mu}(X_i)}{\hat{\pi}(X_i)}\right] = \mathbb{E}\left[\hat{\theta}_{\texttt{OR}}\right]. \tag{11}$$

In either case, $\mathbb{E}\left[\hat{\theta}_{\texttt{DR}}\right] = \theta^\star$.

A more refined analysis of the error of the DR estimator can be carried out using *Von Mises expansion*. In fact, assuming that there exist $\bar{\mu}$ and $\bar{\pi}$ such that $\varphi(\mathscr{D};\theta^\star;\hat{\mu};\hat{\pi}) \xrightarrow{\mathbb{L}^2} \varphi(\mathscr{D};\theta^\star;\bar{\mu};\bar{\pi})$, we can write:

$$\hat{\theta}_{\texttt{DR}} - \theta^\star = \underbrace{(\mathbb{P}_n - \mathbb{P}^\star)[\varphi(\mathscr{D};\theta^\star;\bar{\mu};\bar{\pi})]}_{\text{CLT term}} + \underbrace{(\mathbb{P}_n - \mathbb{P}^\star)[\varphi(\mathscr{D};\theta^\star;\hat{\mu};\hat{\pi}) - \varphi(\mathscr{D};\theta^\star;\bar{\mu};\bar{\pi})]}_{\text{Empirical process term}} + \underbrace{\eta((\hat{\mu},\hat{\pi}),(\mu^\star,\pi^\star))}_{\text{Remainder term}}. \tag{12}$$

The first term is the sample average of a fixed function, and so Central Limit Theorem applies. The second term is an empirical process, and under the mild condition on the $\mathbb{L}_2$-convergence of influence functions above it can be shown to be $o_{\mathbb{P}}(n^{-1/2})$, e.g. by Lemma 2 in Kennedy et al. (2020). The third term, known as *remainder* term, can be directly evaluated, and equals the product of the estimation errors of the two nuisance models:

$$|\eta((\hat{\mu},\hat{\pi}),(\mu^\star,\pi^\star))| = \left|\mathbb{E}\left[(\hat{\mu}-\mu^\star)\left(1-\frac{\pi^\star}{\hat{\pi}}\right)\right]\right| \leq \mathbb{E}\left[|\hat{\mu}-\mu^\star|\left|1-\frac{\pi^\star}{\hat{\pi}}\right|\right], \tag{13}$$

where the last inequality holds by Cauchy-Schwarz. This property, often called *rate* double robustness, *strong* double robustness, or *product bias*, is highly advantageous when one model is easy to estimate well (e.g., at a parametric rate) while the other is not (Wager, 2024). However, this same structure becomes a critical liability under complete model misspecification. When both nuisance functions are incorrect, their errors no longer cancel. Instead, they can compound through this product term, causing the error of the DR estimator to be even larger than that of the simpler OR or IPW estimators. We term this failure mode *double fragility*.

3 Safe estimation through adaptive correction clipping

Having established that the correction term is the source of double fragility, we now introduce a solution that directly targets this mechanism. Our proposal builds on a key insight: any doubly robust estimator can be decomposed into its simpler components plus a correction term. Instead of letting the correction term vary freely, our proposed solution adaptively constrains the correction term to prevent it from destabilizing the final estimate. We define the *adaptive correction clipping* (DR+ACC) estimator as

$$\hat{\theta}_{\texttt{DR+ACC}} = \hat{\theta}_{\texttt{OR}} + \hat{\theta}_{\texttt{IPW}} - \text{clip}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{R_i\hat{\mu}(X_i)}{\hat{\pi}(X_i)}\right), \tag{14}$$

where

$$\text{clip}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{R_i\hat{\mu}(X_i)}{\hat{\pi}(X_i)}\right) = \begin{cases} \min\{\hat{\theta}_{\texttt{OR}}, \hat{\theta}_{\texttt{IPW}}\} & \text{if} \quad \frac{1}{n}\sum_{i=1}^{n}\frac{R_i\hat{\mu}(X_i)}{\hat{\pi}(X_i)} \leq \min\{\hat{\theta}_{\texttt{OR}}, \hat{\theta}_{\texttt{IPW}}\} \\ \max\{\hat{\theta}_{\texttt{OR}}, \hat{\theta}_{\texttt{IPW}}\} & \text{if} \quad \frac{1}{n}\sum_{i=1}^{n}\frac{R_i\hat{\mu}(X_i)}{\hat{\pi}(X_i)} \geq \max\{\hat{\theta}_{\texttt{OR}}, \hat{\theta}_{\texttt{IPW}}\} \\ \frac{1}{n}\sum_{i=1}^{n}\frac{R_i\hat{\mu}(X_i)}{\hat{\pi}(X_i)} & \text{otherwise}. \end{cases} \tag{15}$$

Remark 3.1. A key advantage of our DR+ACC estimator, inherited by the standard doubly robust estimator, is its flexibility; any "black-box" machine learning model can be used to fit the nuisance models for the outcome regression and the propensity score. This allows for the use of state-of-the-art predictive tools best suited for the data at hand – including contemporary deep learning models and even large language models for text (or textualized) covariates – without altering the fundamental statistical properties of the final DR+ACC estimator.

Now that we have introduced our DR+ACC estimator, we can analyze its statistical guarantees. We first focus on estimation performance, and we then move to inference.

3.1 Estimation

The simple modification induced by the clipping operator on the correction term endows the estimator with two desirable properties: *consistency* when at least one of the nuisances is well specified, and *safety* under complete nuisance misspecification. We now make these claims more formal, starting with consistency.

Theorem 3.2 (Consistency)**.** *Assume the identifiability conditions described in Assumption 2.1. Assume also that at least one of the nuisance models, $\hat{\mu}$ or $\hat{\pi}$, is well-specified. Then, the* DR+ACC *estimator is consistent for $\theta^\star$, that is*

$$\hat{\theta}_{\texttt{DR+ACC}} \xrightarrow{p} \theta^\star. \tag{16}$$

Remark 3.3 (Double robustness)**.** This consistency result shows that our DR+ACC estimator is doubly robust, inheriting this critical property from the standard DR estimator. By remaining consistent when at least one nuisance model is correctly specified, the DR+ACC estimator offers a significant advantage over methods like OR and IPW estimators, which are only valid if their respective underlying models are correct. This makes our proposal a more reliable choice for practitioners, providing the same protection against partial misspecification as the classical DR approach.

We now move to safety, which is the key statistical property of our estimator.

Theorem 3.4 (Safety)**.** *Assume the identifiability conditions described in Assumption 2.1. Define*

$$\hat{\lambda} = \frac{\hat{\theta}_{\texttt{OR}} - \text{clip}\left(\frac{1}{n}\sum_{i=1}^{n} \frac{R_i \hat{\mu}(X_i)}{\hat{\pi}(X_i)}\right)}{\hat{\theta}_{\texttt{OR}} - \hat{\theta}_{\texttt{IPW}}} \in [0,1]\,. \tag{17}$$

Then, we have

$$\begin{aligned} |\hat{\theta}_{\texttt{DR+ACC}} - \theta^\star| &\leq \hat{\lambda}|\hat{\theta}_{\texttt{OR}} - \theta^\star| + \left(1-\hat{\lambda}\right)|\hat{\theta}_{\texttt{IPW}} - \theta^\star| \\ &\leq \max\{|\hat{\theta}_{\texttt{OR}} - \theta^\star|, |\hat{\theta}_{\texttt{IPW}} - \theta^\star|\}\,. \end{aligned} \tag{18}$$

Remark 3.5. The previous result can be readily extended to a result involving the absolute bias of the estimator, by taking expectations. In this case, to simplify the analysis, it is enough to estimate $\hat{\lambda}$ on a separate independent sample with respect to $\hat{\theta}_{\texttt{OR}}$ and $\hat{\theta}_{\texttt{IPW}}$, so that one gets

$$\begin{aligned} \mathbb{E}\left[\left|\hat{\theta}_{\texttt{DR+ACC}} - \theta^\star\right|\right] &\leq \mathbb{E}\left[\hat{\lambda}\right]\mathbb{E}\left[|\hat{\mu} - \mu^\star|\right] + \left(1 - \mathbb{E}\left[\hat{\lambda}\right]\right)\mathbb{E}\left[\left|1 - \frac{\pi^\star}{\hat{\pi}}\right||\mu^\star|\right] \\ &\leq \max\left\{\mathbb{E}\left[|\hat{\mu} - \mu^\star|\right], \mathbb{E}\left[\left|1 - \frac{\pi^\star}{\hat{\pi}}\right||\mu^\star|\right]\right\}. \end{aligned} \tag{19}$$

Remark 3.6. This safety property holds regardless of the quality of the nuisance function estimates. It provides a strict guarantee that the DR+ACC estimator error is bounded by the worst of its constituent parts, which implies that the clipped estimator is better than the standard DR estimator whenever the latter fails by performing worse than its components. This, combined with the consistency result in Theorem 3.2, creates a powerful set of assurances. If at least one nuisance model is well-specified, the estimator is consistent for the true parameter. If both models are misspecified, the estimator performance is guaranteed to be no worse than the maximum of its components. This behavior is fundamentally different from the standard DR estimator, which can be catastrophically wrong in the same scenario. The reason for this difference can be understood by comparing their errors. The error of the standard DR estimator depends on the product of the errors of the two nuisance models. In contrast, the error of the DR+ACC estimator is bounded by a convex combination of the errors of the OR and IPW estimators, as shown in Theorem 3.4. In turn, this convex combination is bounded by the maximum of the two errors. When both nuisance models are substantially incorrect, the product of their errors can be far larger than their maximum. This is, again, the source of the double fragility phenomenon, where the standard DR estimator amplifies errors, while the DR+ACC contains them. Finally, it is important to note that the maximum bound provided in Theorem 3.4 is a worst-case guarantee and is often conservative. Instead, the precise evaluation of the error as a convex combination of the estimation error of the nuisance functions provides a more detailed analysis of the safety property of our DR+ACC estimator. As observed in simulations (e.g., Figure 1 and Section 4 below), the performance of the DR+ACC estimator is often much better than the maximum upper bound and can in fact be closer to the minimum of the errors of the component estimators.

3.2 Inference

The safety property of the DR+ACC estimator is achieved through the use of a non-linear clipping operator. A direct, unfortunate consequence of this is that the estimator asymptotic distribution is not normal, even when the nuisance models are correctly specified. Instead, the limiting distribution is a non-linear transformation

Algorithm 1 Parametric bootstrap

Require: α (confidence level), B (number of bootstrap samples)
Ensure: Asymptotically valid confidence intervals for $\theta^\star$, i.e. $[\hat{\theta}_{\texttt{DR+ACC}} - q_{1-\alpha/2}/\sqrt{n}, \hat{\theta}_{\texttt{DR+ACC}} - q_{\alpha/2}/\sqrt{n}]$

Compute estimators $\hat{\theta}_{\texttt{OR}}$, $\hat{\theta}_{\texttt{IPW}}$, and $\hat{C} = \frac{1}{n}\sum_{i=1}^{n} \frac{R_i \hat{\mu}(X_i)}{\hat{\pi}(X_i)}$
for $i \leftarrow 1$ to n **do**
    Estimate empirical influence functions $\varphi\left(\mathscr{D}_i; \hat{\theta}_{\texttt{OR}}; \hat{\mu}\right)$, $\varphi\left(\mathscr{D}_i; \hat{\theta}_{\texttt{IPW}}; \hat{\pi}\right)$, and $\varphi\left(\mathscr{D}_i; \hat{C}; \hat{\mu}; \hat{\pi}\right)$
end for
Estimate covariance matrix, where each element is given by $\hat{\Sigma}_{jk} = \frac{1}{n}\sum_{i=1}^{n} \varphi\left(\mathscr{D}_i; j\right) \varphi\left(\mathscr{D}_i; k\right)$
for $b \leftarrow 1$ to B **do**
    Sample $\left(Z^b_{\texttt{OR}}, Z^b_{\texttt{IPW}}, Z^b_{\text{correction}}\right) \sim \mathcal{N}\left(0, \hat{\Sigma}\right)$ and compute $W^b = Z^b_{\texttt{OR}} + Z^b_{\texttt{IPW}} - \text{clip}\left(Z^b_{\text{correction}}\right)$
end for
Find the empirical $\alpha/2$ and $1-\alpha/2$ quantiles of $\{W^b\}_{b=1}^{B}$, denoted $q_{\alpha/2}$ and $q_{1-\alpha/2}$

of jointly normal random variables, which invalidates standard inferential procedures that rely on a normal approximation. However, we propose a *parametric bootstrap* procedure to construct asymptotically valid confidence intervals. This method does not rely on a normal approximation; instead, it works by simulating the true, non-standard asymptotic distribution of our estimator and then using its empirical quantiles to form an interval.

Before presenting the next Theorem, which characterizes the asymptotic distribution of our DR+ACC estimator when both nuisance models are well-specified, we introduce some additional notation. We define the scaled errors of OR, IPW and the nonclipped correction term, along with their limits in distribution, as:

$$\begin{aligned} \sqrt{n}(\hat{\theta}_{\texttt{OR}} - \theta^\star) &\rightsquigarrow Z_{\texttt{OR}}\,, \\ \sqrt{n}(\hat{\theta}_{\texttt{IPW}} - \theta^\star) &\rightsquigarrow Z_{\texttt{IPW}}\,, \\ \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} \frac{R_i \hat{\mu}(X_i)}{\hat{\pi}(X_i)} - \theta^\star\right) &\rightsquigarrow Z_{\text{correction}}\,, \end{aligned} \tag{20}$$

where the vector $(Z_{\texttt{OR}}, Z_{\texttt{IPW}}, Z_{\text{correction}})$ follows a multivariate normal distribution with mean zero and an unknown covariance matrix Σ.

Theorem 3.7 (Asymptotic distribution of DR+ACC)**.** *Assume the identifiability conditions described in Assumption 2.1. Assume also that the nuisance models, $\hat{\mu}$ and $\hat{\pi}$, are well-specified. Then*

$$\sqrt{n}(\hat{\theta}_{\texttt{DR+ACC}} - \theta^\star) \rightsquigarrow Z_{\texttt{OR}} + Z_{\texttt{IPW}} - \text{clip}\left(Z_{\text{correction}}\right). \tag{21}$$

Remark 3.8. The Theorem recovers the previous consistency result, as the limiting distribution is centered correctly. However, the limiting distribution is not normal. In fact, the random variable $W = Z_{\texttt{OR}} + Z_{\texttt{IPW}} - \text{clip}\left(Z_{\text{correction}}\right)$ is a non-linear transformation of jointly normal variables, and thus not normal. This is the reason why standard inferential procedures based on the normal approximation are not strictly valid for our estimator. Despite this, in our simulations we find that the difference between our limit W and a Gaussian distribution with proper variance is negligible (see Supplementary Figure C.1 for an example).

Based on Theorem 3.7, we construct an asymptotically valid confidence interval by simulating the distribution of W to calculate its quantiles (see Algorithm 1). This approach is, in essence, a parametric bootstrap. The

procedure begins by computing the empirical influence functions for each of the three core components: the `OR` estimator, the `IPW` estimator, and the correction term. These influence functions, which capture the contribution of each observation to the variance, are then used to compute a consistent estimate, $\hat{\Sigma}$, of the joint asymptotic covariance matrix of the three components. Then, a large number (B) of random vectors are drawn from a multivariate normal distribution, $\mathcal{N}\left(0, \hat{\Sigma}\right)$, to simulate the joint asymptotic behavior of the unclipped components, that is $\left(Z^b_{\texttt{OR}}, Z^b_{\texttt{IPW}}, Z^b_{\text{correction}}\right) \sim \mathcal{N}\left(0, \hat{\Sigma}\right)$. For each $b = 1, \ldots, B$, we then compute $W^b = Z^b_{\texttt{OR}} + Z^b_{\texttt{IPW}} - \text{clip}\left(Z^b_{\text{correction}}\right)$. This transforms the sample of jointly normal variables into a large empirical sample from the asymptotic distribution of the final estimator. From this simulated sample, the empirical $\alpha/2$ and $1-\alpha/2$ quantiles, denoted $q_{\alpha/2}$ and $q_{1-\alpha/2}$, are calculated and used to construct the asymptotically valid confidence interval around the point estimate, that is

$$\left[\hat{\theta}_{\texttt{DR+ACC}} - \frac{q_{1-\alpha/2}}{\sqrt{n}}, \hat{\theta}_{\texttt{DR+ACC}} - \frac{q_{\alpha/2}}{\sqrt{n}}\right]. \tag{22}$$

4 Simulation study

To empirically validate the concepts of double fragility and the safety of our proposed `DR+ACC` estimator, we conduct a simulation study that fully replicates the well-known design of Kang and Schafer (2007). This setup is ideal for our purposes as it is explicitly designed to create the possible scenarios of nuisance model specification, allowing for a rigorous evaluation of estimator performance. That is, the simulation is designed to create scenarios where both the models for the regression function and the propensity score can be specified either correctly or incorrectly. For each unit $i = 1, \ldots, n$ we generate data through the following steps:

- Latent variable generation: 4 independent latent variables are drawn from a standard normal distribution:

$$(T_{i1}, T_{i2}, T_{i3}, T_{i4}) \sim \mathcal{N}\left(0, I_4\right), \tag{23}$$

  where I_4 is the 4×4 identity matrix. These latent variables form the basis for both the outcome and the missingness mechanism.

- Outcome generation: the outcome variable, Y_i, is generated as a linear function of the latent variables T_{ij} with an added standard normal error term, ε_i:

$$Y_i = 210 + 27.4T_{i1} + 13.7T_{i2} + 13.7T_{i3} + 13.7T_{i4} + \varepsilon_i, \quad \text{where} \quad \varepsilon_i \sim \mathcal{N}(0,1). \tag{24}$$

  This constitutes the true outcome model. The resulting population mean is $\theta^\star = \mathbb{E}[Y_i] = 210$.

- Missingness mechanism: the probability of the outcome Y_i being observed, denoted $\pi(T_i)$, is determined by a logistic regression model that is also linear in the latent variables:

$$\text{logit}(\pi(T_i)) = \text{logit}\left(\mathbb{P}\left[R_i = 1 \mid T_{i1}, T_{i2}, T_{i3}, T_{i4}\right]\right) = -T_{i1} + 0.5T_{i2} - 0.25T_{i3} - 0.1T_{i4}. \tag{25}$$

  This defines the true propensity score model. The response indicator for each unit, R_i, is then drawn from a Bernoulli distribution with this probability:

$$R_i \sim \text{Bernoulli}(\pi(T_i)). \tag{26}$$

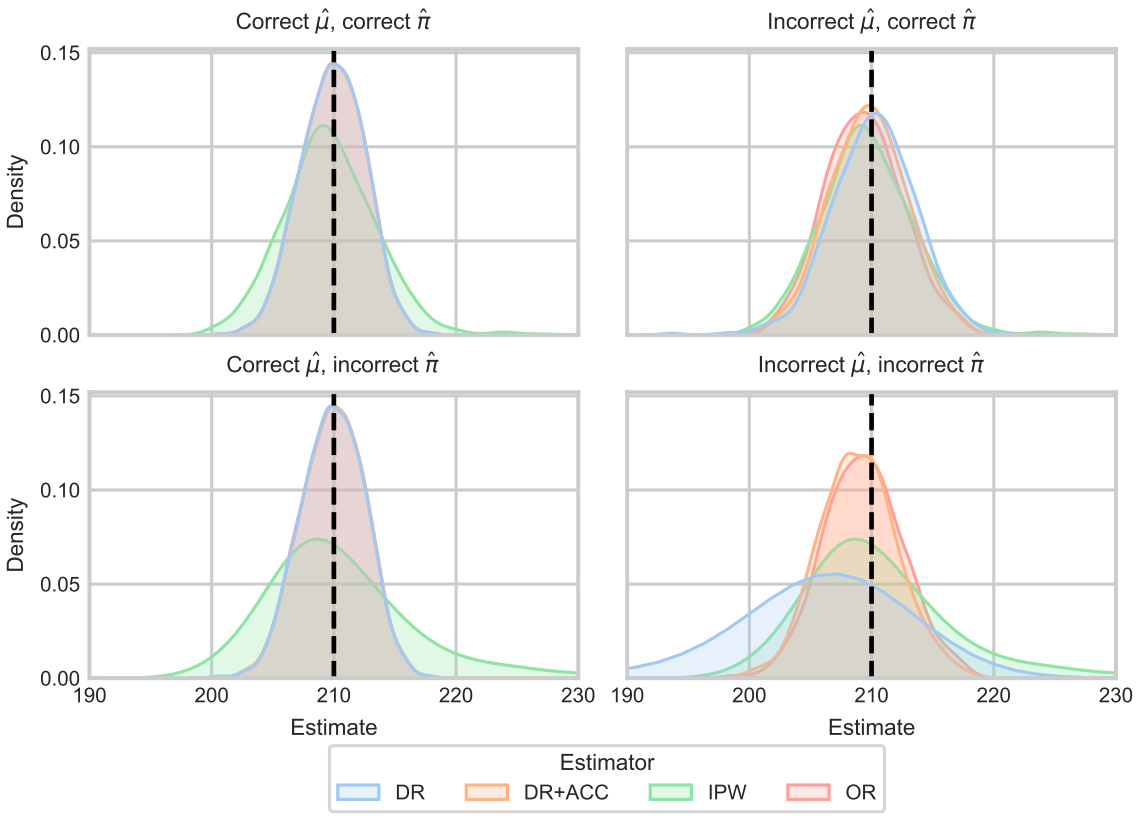

Figure 3: Sampling distributions for the Outcome Regression (`OR`), Inverse Probability Weighting (`IPW`), Doubly Robust (`DR`), and our proposed `DR+ACC` estimators from 1000 simulations with a sample size of $n = 200$. The true parameter value is 210. The four panels show the estimators' performance under all combinations of correct and incorrect nuisance model specifications. The top row and bottom-left panel demonstrate the *asymptotic hard thresholding* property. When at least one nuisance model is correct, both `DR` and `DR+ACC` align with the correct simpler estimator. The bottom-right panel illustrates *double fragility*. When both nuisance models are wrong, the bias of the standard `DR` estimator is substantially worse than that of either the `OR` or `IPW` estimators. In this challenging scenario, our proposed `DR+ACC` estimator is shown to be *safe*, providing a much more stable and accurate estimate than the standard `DR` estimator. Full details of the simulation scenario are provided in Section 4.

- Observed covariates: instead of observing the true latent variables, the analyst is presented with a set of four covariates which are non-linear transformations of the original ones:

$$X_{i1} = \exp\left(\frac{T_{i1}}{2}\right), \quad X_{i2} = \frac{T_{i2}}{1+\exp(T_{i1})} + 10, \quad X_{i3} = \left(\frac{T_{i1}T_{i3}}{25} + 0.6\right)^3, \quad X_{i4} = (T_{i2} + T_{i4} + 20)^2. \tag{27}$$

  An analyst who fits a linear model for the outcome or a logistic model for the propensity score using these observed covariates X_i will have a misspecified model. A correct specification would require the analyst to know the true latent variables T_i or the exact inverse transformations.

We first assess performance in terms of estimation accuracy, which we measure through three different metrics. *Bias* measures the difference between the estimated mean and the true mean ($\theta^\star = 210$). The *root mean squared error* (RMSE) is the square root of the average squared difference between the estimate and

Table 1: Simulation results for $n = 200$ and $n = 1000$, grouped by nuisance model specification. The results highlight the double fragility of the standard DR estimator and the safety of the DR+ACC estimator across both sample sizes.

		$n = 200$			$n = 1000$		
Scenario	**Estimator**	**Bias**	**RMSE**	**MAE**	**Bias**	**RMSE**	**MAE**
Correct $\hat{\mu}$, Correct $\hat{\pi}$	OR	-0.072	2.568	1.853	-0.002	1.128	0.761
	IPW	-0.311	3.859	2.464	-0.019	1.688	1.098
	DR	-0.072	2.570	1.851	-0.002	1.128	0.767
	DR+ACC	-0.072	2.569	1.853	-0.000	1.128	0.766
Correct $\hat{\mu}$, Incorrect $\hat{\pi}$	OR	-0.072	2.568	1.853	-0.002	1.128	0.761
	IPW	1.595	9.726	3.412	4.761	11.095	2.561
	DR	-0.076	2.574	1.840	0.018	1.314	0.786
	DR+ACC	-0.047	2.568	1.853	0.056	1.298	0.772
Incorrect $\hat{\mu}$, Correct $\hat{\pi}$	OR	-0.665	3.306	2.273	-0.814	1.678	1.179
	IPW	-0.311	3.859	2.464	-0.019	1.688	1.098
	DR	0.160	3.448	2.268	0.028	1.643	1.069
	DR+ACC	-0.257	3.202	2.126	-0.395	1.544	1.077
Incorrect $\hat{\mu}$, Incorrect $\hat{\pi}$	OR	-0.665	3.306	2.273	-0.814	1.678	1.179
	IPW	1.595	9.726	3.412	4.761	11.095	2.561
	DR	-6.395	21.932	4.026	-13.467	77.565	5.295
	DR+ACC	-1.094	3.375	2.259	-0.912	1.717	1.226

the true mean, providing a comprehensive measure of an estimator's overall accuracy by combining both its bias and variance. Finally, the *median absolute error* (MAE) is a robust measure of precision, calculated as the median of the absolute errors.

Figures 1 and 3, together with Table 1, summarize the results of our simulation study, providing strong empirical support for our central arguments across both sample sizes ($n = 200$ and $n = 1000$). In the three scenarios where at least one nuisance model is correctly specified, the results align with classical theory. Both the standard doubly robust (DR) and our proposed DR+ACC estimators perform excellently, exhibiting minimal bias and RMSE that is comparable to the best-performing correctly specified estimator (OR or IPW). This confirms their double robustness and the asymptotic hard thresholding property in practice. The most critical insights come from the scenario where both nuisance models are misspecified. Here, the DR estimator fails catastrophically, demonstrating the phenomenon of double fragility. As shown in Table 1, its RMSE explodes, increasing from approximately 2.570 to 21.932 for $n = 200$, and from approximately 1.128 to 77.565 for $n = 1000$ – an error far exceeding that of either the misspecified OR or IPW estimators. In stark contrast, our DR+ACC estimator remains stable in this challenging setting. Its RMSE is dramatically lower than that of the standard DR estimator (3.375 vs. 21.932 for $n = 200$; 1.717 vs. 77.565 for $n = 1000$) and is comparable to the better between the two simpler OR and IPW estimators. Similarly, Supplementary Figures C.2 and C.3 display the relationship between the estimates provided by DR+ACC and the other estimators. Again, DR+ACC estimates follow the standard DR when the latter is stable but remains bounded when the standard DR produces extreme, unstable estimates. These results (together with the ones for $n = 100$ in Supplementary Figures C.4 and C.5) empirically validate the safety property of our method, demonstrating that it successfully mitigates the critical failure mode of the standard DR approach without sacrificing performance in well-behaved settings.

Table 2: Empirical coverage and confidence interval (CI) widths for $n = 200$ and $n = 1000$, grouped by nuisance model specification. The nominal coverage level is 95%. We set B, the number of bootstrap samples, to 10000.

		$n = 200$		$n = 1000$	
Scenario	**Estimator**	**Coverage**	**CI Width**	**Coverage**	**CI Width**
Correct $\hat{\mu}$, Correct $\hat{\pi}$	OR	0.948	9.99	0.938	4.49
	IPW	1.000	91.12	1.000	41.32
	DR	0.956	10.30	0.939	4.52
	DR+ACC	0.951	10.16	0.939	4.51
Correct $\hat{\mu}$, Incorrect $\hat{\pi}$	OR	0.948	9.99	0.938	4.49
	IPW	1.000	138.69	1.000	169.34
	DR	0.962	12.15	0.963	7.61
	DR+ACC	0.957	11.62	0.960	6.99
Incorrect $\hat{\mu}$, Correct $\hat{\pi}$	OR	0.861	10.31	0.836	4.59
	IPW	1.000	91.12	1.000	41.32
	DR	0.965	13.80	0.953	6.60
	DR+ACC	0.934	12.02	0.921	5.46
Incorrect $\hat{\mu}$, Incorrect $\hat{\pi}$	OR	0.861	10.31	0.836	4.59
	IPW	1.000	138.69	1.000	169.34
	DR	0.915	22.02	0.713	55.43
	DR+ACC	0.904	11.68	0.862	5.37

We also evaluate performance in terms of inference, using two metrics: *empirical coverage* (the fraction of experiments in which confidence intervals contain the true parameter) and the average confidence interval (CI) *width*. The results for all four nuisance specification scenarios are presented in Table 2. The results in the ideal scenario where both the outcome regression and propensity score models are correctly specified, the only scenario in which confidence intervals are theretically guaranteed to be asymptotically valid (Kennedy, 2024), validate that our inference procedure for the DR+ACC estimator achieves the nominal 95% coverage level, performing almost identically to the standard DR and OR estimators at both sample sizes ($n = 200$ and $n = 1000$). In terms of efficiency, the IPW estimator is clearly suboptimal, producing extremely wide confidence intervals. In contrast, the DR and DR+ACC estimators are highly efficient, yielding CIs that are nearly identical in width to the efficient OR estimator. This confirms that the safety mechanism of the DR+ACC estimator does not compromise its statistical efficiency in well-behaved settings where the risk of fragility is absent. While our theoretical guarantees do not cover misspecified settings, our simulation results also provide valuable insights into the practical robustness of our approach. When one of the two nuisance functions is misspecified, our DR+ACC confidence intervals remain highly competitive with the standard DR CIs. In addition, when both nuisance functions are wrong, the advantage of our approach becomes even clearer. The standard DR confidence intervals suffer from severe undercoverage, dropping to 71.3% at $n = 1000$, and their width increases substantially. In contrast, our DR+ACC confidence intervals maintain much better coverage (86.2%) and are narrower (5.37 vs. 55.43). This, together with the results for $n = 100$ in Supplementary Table C.1, suggests that the safety property of the DR+ACC estimator not only improves point estimation but also leads to more reliable inference in the realistic setting of complete model misspecification.

5 Application to Alzheimer's disease proteomics

To demonstrate the practical efficacy of our method in a real-world scientific setting, we revisit an application from Moon et al. (2025), focusing on the impact of Alzheimer's disease (AD) on the human proteome. AD is a prominent neurodegenerative disorder, and while many contributing factors are known, its underlying biological pathways are still being discovered. Bulk peptide-level datasets offer a valuable opportunity to explore these pathways, but they also present complex modeling challenges where the risk of misspecification is high. Our objective is to estimate the average treatment effect (ATE) of an AD diagnosis on the abundances of various peptides. This estimand, which is the difference between two means, fits naturally within our missing data framework.

We use the peptide abundance data from Merrihew et al. (2023), which contains samples from individuals with and without dementia. Following previous work (Moon et al., 2025), we define our treatment variable by grouping samples into two categories: cases (individuals with autosomal dominant or sporadic AD dementia) and controls (individuals without dementia, with or without a high AD histopathologic burden). To ensure data quality, we focus on the 270 peptides that present no more than 10% missing values across the 220 observations in the data set, and we impute the remaining missing values through `MissForest` (Stekhoven and Bühlmann, 2012).

For our analysis, we estimate the two nuisance functions using standard approaches. For the outcome model, we use a difference-in-means estimator. While simple, this approach is routinely used in the analysis of peptide abundance data (Chen et al., 2020) and serves as a relevant baseline. For the propensity score model, we fit a logistic regression of the AD diagnosis on some available external covariates, namely brain region, post-mortem interval (PMI), age, and gender. To improve stability, the estimated propensity scores are rescaled in the spirit of a Hájek estimator (Basu, 1971): we first multiply each estimated propensity score $\tilde{\pi}(X_i)$ by $n^{-1}\sum_{i=1}^{n} A_i/\tilde{\pi}(X_i)$ and $n^{-1}\sum_{i=1}^{n}(1-A_i)/(1-\tilde{\pi}(X_i))$ for treatment and control observations respectively, and then we clip the result so that it ranges between 0 and 1. We then compute the `OR`, `IPW`, `DR`, and our proposed `DR+ACC` estimators.

The results, illustrated in Figure 4 and Supplementary Figure D.6, highlight the practical impact of double fragility in a real-world setting. For several peptides, the standard `DR` estimator produces ATE estimates that deviate from the `OR` and `IPW` estimates. In contrast, our `DR+ACC` estimator consistently provides more credible results that remain anchored between the two simpler `OR` and `IPW` estimators, demonstrating the importance of its safety property for drawing reliable scientific conclusions when the true data-generating process is unknown.

We now present a preliminary analysis of the above estimates. Switching from the standard `DR` to the proposed `DR+ACC` has substantial consequences. The standard `DR` estimator, with its Gaussian approximation, identifies 55 peptides as differentially abundant in cases vs. controls (at a significance level of $\alpha = 0.05$, without multiple testing correction). In contrast, the `DR+ACC` estimator, with its parametric bootstrap, identifies 95 significant peptides – the same 55, plus 40 additional findings. These additional peptides map to proteins translated from 38 distinct primary genes. Among them, 34 were *not* found using the standard `DR` estimator. Moreover, and notably, these 34 genes have been independently implicated in Alzheimer's disease by prior research, providing some external validation for our findings. A full list of the genes and the supporting literature is available in Supplementary Table D.2. While a full biological interpretation is beyond the scope of this paper, these results suggest that the `DR+ACC` estimator, by mitigating the instability

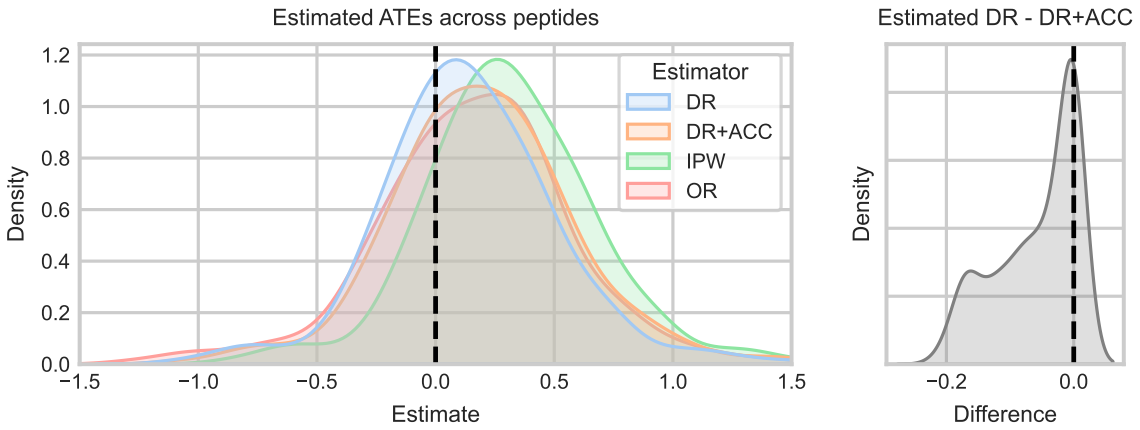

Figure 4: In the left panel, we show the distributions of the estimated average treatment effects (ATEs) of Alzheimer's disease across 270 different peptides. All estimators produce distributions centered near zero, suggesting that for most peptides, the estimated effect of AD is small. However, some of the estimates from the standard `DR` estimator are outside the range defined by the estimates of `OR` and `IPW`, providing a real-world example of double fragility. The stability of the `DR+ACC` estimator, which remains aligned with the more plausible `OR` and `IPW` results, demonstrates the practical importance of its safety property for drawing reliable scientific conclusions. In the right panel, we focus on the difference between the `DR` estimator and `DR+ACC`, plotting the distribution between the estimates provided by the two estimators.

inherent in the standard `DR` approach, can indeed uncover credible and scientifically important biological signals.

6 Conclusions

This paper identifies and formally characterizes double fragility, a critical failure mode of standard doubly robust estimators that arises under the ubiquitous scenario of complete nuisance model misspecification. While these estimators are celebrated for their theoretical properties, we show that the very mechanism that provides robustness – which we call asymptotic hard thresholding – can, in practice, amplify errors and lead to instability.

To address this, we propose the `DR+ACC` estimator. Our solution is both simple and powerful. When at least one nuisance model is correct, our `DR+ACC` estimator retains the desirable double robustness property of standard methods, ensuring consistency. At the same time, our `DR+ACC` posesses a crucial safety guarantee, which ensures that its performance cannot be worse than that of its simpler constituent parts when all models are misspecified, thus preventing catastrophic failure. Finally, we propose an inference procedure based on the parametric bootstrap, which provides asymptotically valid confidence intervals when all nuisance models are correctly specified – and appears to behave reasonably also under misspecification. Our simulation study and application to peptide abundance data confirm that the `DR+ACC` estimator provides a practical and reliable alternative, offering practitioners the benefits of double robustness with a crucial safeguard against its fragility.

This work opens several avenues for future research. First, our `DR+ACC` framework could be extended to handle multivariate or other complex target parameters. Second, while a key practical advantage of

our DR+ACC estimator is that it does not require the tuning of any additional parameters, alternative solutions for achieving safety, such as power-tuning or data-adaptive convex combinations of the OR and IPW estimators (Angelopoulos et al., 2023b), could and should be explored. Finally, our simulation study suggests that the bias of double robust estimators can be exacerbated, not reduced, as the sample size grows. Exploring the relationship between sample size and bias of DR under complete misspecification represents a further interesting research direction.

Acknowledgments

L.T. wishes to thank the members of the causal inference reading group at Carnegie Mellon University, Ana M. Kenney and the other participants to the 2025 L'EMbeDS workshop "Learning from large, complex and structured data" held in Pisa, Italy, in June 2025 for helpful discussions. This project was supported by National Institute of Mental Health (NIMH) grant R01MH123184.

References

Julia M Adams, Sanket V Rege, Angela T Liu, Ninh V Vu, Sharda Raina, Douglas Y Kirsher, Amy L Nguyen, Reema Harish, Balazs Szoke, Dino P Leone, et al. Leukotriene a4 hydrolase inhibition improves age-related cognitive decline via modulation of synaptic function. *Science Advances*, 9(46):eadf8764, 2023.

Abhineet Agarwal, Michael Xiao, Rebecca Barter, Omer Ronen, Boyu Fan, and Bin Yu. Pcs-uq: Uncertainty quantification via the predictability-computability-stability framework. *arXiv preprint arXiv:2505.08784*, 2025.

Anastasios N Angelopoulos, Stephen Bates, Clara Fannjiang, Michael I Jordan, and Tijana Zrnic. Prediction-powered inference. *Science*, 382(6671):669–674, 2023a.

Anastasios N Angelopoulos, John C Duchi, and Tijana Zrnic. Ppi++: Efficient prediction-powered inference. *arXiv preprint arXiv:2311.01453*, 2023b.

David Azriel, Lawrence D Brown, Michael Sklar, Richard Berk, Andreas Buja, and Linda Zhao. Semi-supervised linear regression. *Journal of the American Statistical Association*, 117(540):2238–2251, 2022.

Heejung Bang and James M Robins. Doubly robust estimation in missing data and causal inference models. *Biometrics*, 61(4):962–973, 2005.

D Basu. An essay on the logical foundations of survey sampling, part i. foundations of statistical inferences, vp godambe and da sprott, 1971.

Peter J Bickel, Chris AJ Klaassen, Ya'acov Ritov, and Jon A Wellner. *Efficient and adaptive estimation for semiparametric models*, volume 4. Springer, 1993.

Laura J Blair, Jeremy D Baker, Jonathan J Sabbagh, and Chad A Dickey. The emerging role of peptidyl-prolyl isomerase chaperones in tau oligomerization, amyloid processing, and alzheimer's disease. *Journal of neurochemistry*, 133(1):1–13, 2015.

Tiffany Tianhui Cai, Yuri Fonseca, Kaiwen Hou, and Hongseok Namkoong. C-learner: Constrained learning for causal inference. *arXiv preprint arXiv:2405.09493*, 2024.

Michael Celentano and Martin J Wainwright. Challenges of the inconsistency regime: Novel debiasing methods for missing data models. *arXiv preprint arXiv:2309.01362*, 2023.

Abhishek Chakrabortty. *Robust semi-parametric inference in semi-supervised settings*. PhD thesis, Harvard University, 2016.

Abhishek Chakrabortty and Tianxi Cai. Efficient and adaptive linear regression in semi-supervised settings. *The Annals of Statistics*, 46(4):1541–1572, 2018.

Chen Chen, Jie Hou, John J Tanner, and Jianlin Cheng. Bioinformatics methods for mass spectrometry-based proteomics data analysis. *International journal of molecular sciences*, 21(8):2873, 2020.

Philip S Crooke III, John T Tossberg, Rachel M Heinrich, Krislyn P Porter, and Thomas M Aune. Reduced rna adenosine-to-inosine editing in hippocampus vasculature associated with alzheimer's disease. *Brain Communications*, 4(5):fcac238, 2022.

Siyi Deng, Yang Ning, Jiwei Zhao, and Heping Zhang. Optimal and safe estimation for high-dimensional semi-supervised learning. *Journal of the American Statistical Association*, 119(548):2748–2759, 2024.

Josef Finsterer. Neuropathy due to impaired axonal transport of non-fragmented mitochondria in myh14 mutation carriers. *EBioMedicine*, 49:24, 2019.

Mehrdad Ghadiri, David Arbour, Tung Mai, Cameron Musco, and Anup B Rao. Finite population regression adjustment and non-asymptotic guarantees for treatment effect estimation. *Advances in Neural Information Processing Systems*, 36:74180–74212, 2023.

Suyash Gupta and Dominik Rothenhäusler. The s-value: evaluating stability with respect to distributional shifts. *Advances in Neural Information Processing Systems*, 36:72058–72070, 2023.

Frank R Hampel. The influence curve and its role in robust estimation. *Journal of the american statistical association*, 69(346):383–393, 1974.

Frank R Hampel, Elvezio M Ronchetti, Peter J Rousseeuw, and Werner A Stahel. *Robust Statistics: The Approach Based on Influence Functions*. John Wiley & Sons, 2011.

Jay S Hanas, James RS Hocker, Christian A Vannarath, Megan R Lerner, Scott G Blair, Stan A Lightfoot, Rushie J Hanas, James R Couch, and Linda A Hershey. Distinguishing alzheimer's disease patients and biochemical phenotype analysis using a novel serum profiling platform: potential involvement of the vwf/adamts13 axis. *Brain Sciences*, 11(5):583, 2021.

Xuming He and Stephen Portnoy. Reweighted ls estimators converge at the same rate as the initial estimator. *The Annals of Statistics*, pages 2161–2167, 1992.

Miguel A Hernán and James M Robins. Estimating causal effects from epidemiological data. *Journal of Epidemiology & Community Health*, 60(7):578–586, 2006.

James P Higham, Bilal R Malik, Edgar Buhl, Jennifer M Dawson, Anna S Ogier, Katie Lunnon, and James JL Hodge. Alzheimer's disease associated genes ankyrin and tau cause shortened lifespan and memory loss in drosophila. *Frontiers in cellular neuroscience*, 13:260, 2019.

Daniel G Horvitz and Donovan J Thompson. A generalization of sampling without replacement from a finite universe. *Journal of the American statistical Association*, 47(260):663–685, 1952.

Wanhua Hu, Xiaodong Lin, and Kelong Chen. Integrated analysis of differential gene expression profiles in hippocampi to identify candidate genes involved in alzheimer's disease. *Molecular medicine reports*, 12 (5):6679–6687, 2015.

Peter J Huber. *Robust statistical procedures*. SIAM, 1996.

Toshio Ikeda, Akihiko Nakahara, Rie Nagano, Maiko Utoyama, Megumi Obara, Hiroshi Moritake, Tamayo Uechi, Jun Mitsui, Hiroyuki Ishiura, Jun Yoshimura, et al. Tbcd may be a causal gene in progressive neurodegenerative encephalopathy with atypical infantile spinal muscular atrophy. *Journal of human genetics*, 62(4):473–480, 2017.

Joseph DY Kang and Joseph L Schafer. Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. *Statistical Science*, 22(4):523–539, 2007.

Edward H Kennedy. Semiparametric doubly robust targeted double machine learning: a review. *Handbook of statistical methods for precision medicine*, pages 207–236, 2024.

Edward H Kennedy, Sivaraman Balakrishnan, and Max G'Sell. Sharp instruments for classifying compliers and generalizing causal effects. *The Annals of Statistics*, 48(4):2008–2030, 2020.

Ilmun Kim, Larry Wasserman, Sivaraman Balakrishnan, and Matey Neykov. Semi-supervised u-statistics. *arXiv preprint arXiv:2402.18921*, 2024.

Rachel E Lackie, Jose Marques-Lopes, Valeriy G Ostapchenko, Sarah Good, Wing-Yiu Choy, Patricija van Oosten-Hawle, Stephen H Pasternak, Vania F Prado, and Marco AM Prado. Increased levels of stress-inducible phosphoprotein-1 accelerates amyloid-β deposition in a mouse model of alzheimer's disease. *Acta neuropathologica communications*, 8(1):143, 2020.

Guangpu Li. Rab gtpases, membrane trafficking and diseases. *Current drug targets*, 12(8):1188–1193, 2011.

Jie Li, Lingfang Li, Shanshan Cai, Kun Song, and Shenghui Hu. Identification of novel risk genes for alzheimer's disease by integrating genetics from hippocampus. *Scientific Reports*, 14(1):27484, 2024.

Qingqin S Li and Louis De Muynck. Differentially expressed genes in alzheimer's disease highlighting the roles of microglia genes including olr1 and astrocyte gene cdk2ap1. *Brain, behavior, & immunity-health*, 13:100227, 2021.

Ekaterina A Litus, Marina P Shevelyova, Alisa A Vologzhannikova, Evgenia I Deryusheva, Andrey V Machulin, Ekaterina L Nemashkalova, Maria E Permyakova, Andrey S Sokolov, Valeria D Alikova, Vladimir N Uversky, et al. Binding of pro-inflammatory proteins s100a8 or s100a9 to amyloid-β peptide suppresses its fibrillation. *Biomolecules*, 15(3):431, 2025.

Lei Liu, Naoki Watanabe, Hiroyasu Akatsu, and Masaki Nishimura. Neuronal expression of ilei/fam3c and its reduction in alzheimer's disease. *Neuroscience*, 330:236–246, 2016.

T Liu, K Hou, J Li, T Han, S Liu, and Jianshe Wei. Alzheimer's disease and aging association: identification and validation of related genes. *The Journal of Prevention of Alzheimer's Disease*, 11(1):196–213, 2024.

Jian Qi Luo, Luca A Hategan, Samantha Creighton, Shan Hua, Timothy AB McLean, Tarkan Ahmad Dahi, Zhenhong Jin, Fardad Pirri, Stephen M Winston, Isaiah L Reeves, et al. Opposing effects of histone h2a.

z on memory, transcription and pathology in male and female alzheimer's disease mice and patients. *bioRxiv*, pages 2025–05, 2025.

Yukun Ma, Pedro HC Sant'Anna, Yuya Sasaki, and Takuya Ura. Doubly robust estimators with weak overlap. *arXiv preprint arXiv:2304.08974*, 2023.

Bryan Maloney, Yokesh Balaraman, Yunlong Liu, Nipun Chopra, Howard J Edenberg, John Kelsoe, John I Nurnberger, and Debomoy K Lahiri. Lithium alters expression of rnas in a type-specific manner in differentiated human neuroblastoma neuronal cultures, including specific genes involved in alzheimer's disease. *Scientific reports*, 9(1):18261, 2019.

Gennifer E Merrihew, Jea Park, Deanna Plubell, Brian C Searle, C Dirk Keene, Eric B Larson, Randall Bateman, Richard J Perrin, Jasmeer P Chhatwal, Martin R Farlow, et al. A peptide-centric quantitative proteomics dataset for the phenotypic assessment of alzheimer's disease. *Scientific Data*, 10(1):206, 2023.

Haeun Moon, Jin-Hong Du, Jing Lei, and Kathryn Roeder. Augmented doubly robust post-imputation inference for proteomic data. *bioRxiv*, pages 2024–03, 2025.

Wenlong Mou, Martin J Wainwright, and Peter L Bartlett. Off-policy estimation of linear functionals: Non-asymptotic theory for semi-parametric efficiency. *arXiv preprint arXiv:2209.13075*, 2022.

Claudius Mueller, Weidong Zhou, Amy VanMeter, Michael Heiby, Shino Magaki, Mark M Ross, Virginia Espina, Matthew Schrag, Cindy Dickson, Lance A Liotta, et al. The heme degradation pathway is a promising serum biomarker source for the early detection of alzheimer's disease. *Journal of Alzheimer's Disease*, 19(3):1081–1091, 2010.

Cynthia Picard, Henrik Zetterberg, Kaj Blennow, Sylvia Villeneuve, Judes Poirier, and Prevent-Ad Research Group. Brain and csf alzheimer's biomarkers are associated with serpine1 gene expression. *Genes*, 16(7): 818, 2025.

Raghavan Pillai Raju, Lun Cai, Alpna Tyagi, and Subbiah Pugazhenthi. Interactions of cellular energetic gene clusters in the alzheimer's mouse brain. *Molecular neurobiology*, 61(1):476–486, 2024.

Zachary T Rewolinski and Bin Yu. Pcs workflow for veridical data science in the age of ai. *arXiv preprint arXiv:2508.00835*, 2025.

James M Robins, Andrea Rotnitzky, and Lue Ping Zhao. Estimation of regression coefficients when some regressors are not always observed. *Journal of the American statistical Association*, 89(427):846–866, 1994.

Matthew J Rosene, Niko-Petteri Nykanen, Logan Brase, Oscar Harari, and Bruno A Benitez. A cell-autonomous role of dnajc5 in microglial-mediated proteostasis in alzheimer's disease. *Alzheimer's & Dementia*, 19:e071075, 2023.

Daniel O Scharfstein, Andrea Rotnitzky, and James M Robins. Adjusting for nonignorable drop-out using semiparametric nonresponse models. *Journal of the American Statistical Association*, 94(448):1096–1120, 1999.

Alberto Serrano-Pozo, Huan Li, Zhaozhi Li, Clara Muñoz-Castro, Methasit Jaisa-Aad, Molly A Healey, Lindsay A Welikovitch, Rojashree Jayakumar, Annie G Bryant, Ayush Noori, et al. Astrocyte transcriptomic

changes along the spatiotemporal progression of alzheimer's disease. *Nature neuroscience*, 27(12): 2384–2400, 2024.

Mohammed R Shaker, Amna Kahtan, Renuka Prasad, Ju-Hyun Lee, Giovanni Pietrogrande, Hannah C Leeson, Woong Sun, Ernst J Wolvetang, and Andrii Slonchak. Neural epidermal growth factor-like like protein 2 is expressed in human oligodendroglial cell types. *Frontiers in Cell and Developmental Biology*, 10:803061, 2022.

Hongtao Shen, Yuying Xie, Yan Wang, Yusheng Xie, Yongxiang Wang, Zhenyan Su, Laixi Zhao, Shi Yao, Xiaoling Cao, Jinglan Liang, et al. The er protein canx (calnexin)-mediated autophagy protects against alzheimer disease. *Autophagy*, 21(5):1096–1115, 2025.

Serena Silvestro, Luigi Chiricosta, Agnese Gugliandolo, Renato Iori, Patrick Rollin, Daniele Perenzoni, Fulvio Mattivi, Placido Bramanti, and Emanuela Mazzon. The moringin/α-cd pretreatment induces neuroprotection in an in vitro model of alzheimer's disease: A transcriptomic study. *Current Issues in Molecular Biology*, 43(1):197–214, 2021.

Daniel J Stekhoven and Peter Bühlmann. Missforest—non-parametric missing value imputation for mixed-type data. *Bioinformatics*, 28(1):112–118, 2012.

Til Stürmer, Michael Webster-Clark, Jennifer L Lund, Richard Wyss, Alan R Ellis, Mark Lunt, Kenneth J Rothman, and Robert J Glynn. Propensity score weighting and trimming strategies for reducing variance and bias of treatment effect estimates: a simulation study. *American journal of epidemiology*, 190(8): 1659–1670, 2021.

Qiushan Tao, Ting Fang Alvin Ang, Samia C Akhter-Khan, Indira Swetha Itchapurapu, Ronald Killiany, Xiaoling Zhang, Andrew E Budson, Katherine W Turk, Lee Goldstein, Jesse Mez, et al. Impact of c-reactive protein on cognition and alzheimer disease biomarkers in homozygous apoe e4 carriers. *Neurology*, 97 (12):e1243–e1252, 2021.

Lorenzo Testa, Qi Xu, Jing Lei, and Kathryn Roeder. Semiparametric semi-supervised learning for general targets under distribution shift and decaying overlap. *arXiv preprint arXiv:2505.06452*, 2025.

Anastasios A Tsiatis. *Semiparametric theory and missing data*, volume 4. Springer, 2006.

Mark J Van Der Laan and Daniel Rubin. Targeted maximum likelihood learning. *The international journal of biostatistics*, 2(1), 2006.

Aad W Van der Vaart. *Asymptotic statistics*, volume 3. Cambridge university press, 2000.

Stefan Wager. Causal inference: A statistical learning approach, 2024.

Fei Wang and Yuhao Deng. Non-asymptotic bounds of aipw estimators for means with missingness at random. *Mathematics*, 11(4):818, 2023.

Hansen Wang, Declan Williams, Jennifer Griffin, Takashi Saito, Takaomi C Saido, Paul E Fraser, Ekaterina Rogaeva, and Gerold Schmitt-Ulms. Time-course global proteome analyses reveal an inverse correlation between aβ burden and immunoglobulin m levels in the appnl-f mouse model of alzheimer disease. *PLoS One*, 12(8):e0182844, 2017.

Yuhao Wang, Arnab Bhattacharyya, Jin Tian, and NV Vinodchandran. Pac style guarantees for doubly robust generalized front-door estimator. In *9th Causal Inference Workshop at UAI 2024*, 2024.

Jamal B Williams, Qing Cao, and Zhen Yan. Transcriptomic analysis of human brains with alzheimer's disease reveals the altered expression of synaptic genes linked to cognitive deficits. *Brain communications*, 3(3):fcab123, 2021.

R Teal Witter and Christopher Musco. Benchmarking estimators for natural experiments: A novel dataset and a doubly robust algorithm. *Advances in Neural Information Processing Systems*, 37:82594–82626, 2024.

Zichun Xu, Daniela Witten, and Ali Shojaie. A unified framework for semiparametrically efficient semi-supervised learning. *arXiv preprint arXiv:2502.17741*, 2025.

Ming-Xuan Yang, Zhuo-Ran Wang, Yan-Li Zhang, Zhi-Na Zhang, Yan-Li Li, Rui Wang, Qiang Su, and Jun-Hong Guo. Albumin antagonizes alzheimer's disease-related tau pathology and enhances cognitive performance by inhibiting aberrant tau aggregation. *Experimental Neurology*, 386:115155, 2025.

Yu Yang, Xu Wang, Weina Ju, Li Sun, and Haining Zhang. Genetic and expression analysis of copi genes and alzheimer's disease susceptibility. *Frontiers in Genetics*, 10:866, 2019.

Takashi Yasukawa, Aya Tsutsui, Chieri Tomomori-Sato, Shigeo Sato, Anita Saraf, Michael P Washburn, Laurence Florens, Tohru Terada, Kentaro Shimizu, Ronald C Conaway, et al. Nrbp1-containing crl2/crl4a regulates amyloid β production by targeting bri2 and bri3 for degradation. *Cell reports*, 30(10):3478–3491, 2020.

Shan-Ju Yeh, Ming-Hsun Chung, and Bor-Sen Chen. Investigating pathogenetic mechanisms of alzheimer's disease by systems biology approaches for drug discovery. *International Journal of Molecular Sciences*, 22 (20):11280, 2021.

Bin Yu. Stability. *Bernoulli*, 19(4):1484–1500, 2013.

Bin Yu and Rebecca L Barter. *Veridical data science: The practice of responsible data analysis and decision making*. MIT Press, 2024.

Bin Yu and Karl Kumbier. Veridical data science. *Proceedings of the National Academy of Sciences of the United States of America*, 117(8):3920–3929, 2020.

Anru Zhang, Lawrence D Brown, and T Tony Cai. Semi-supervised inference: General theory and estimation of means. *The Annals of Statistics*, 47(5):2538–2566, 2019.

Yuqian Zhang and Jelena Bradic. High-dimensional semi-supervised learning: in search of optimal inference of the mean. *Biometrika*, 109(2):387–403, 2022.

Tijana Zrnic and Emmanuel J Candès. Cross-prediction-powered inference. *Proceedings of the National Academy of Sciences*, 121(15):e2322083121, 2024.

Supplementary Material

A Technical lemmas

Lemma A.1 (Continuity of the clip operator)**.** *The function* $\text{clip}(x, y, z) = \max(y, \min(x, z))$ *is a continuous function from* $\mathbb{R}^3$ *to* $\mathbb{R}$*.*

Proof. The functions $g_1(x, y, z) = x$, $g_2(x, y, z) = y$, and $g_3(x, y, z) = z$ are continuous. The binary operators $\min(a, b)$ and $\max(a, b)$ are continuous functions from $\mathbb{R}^2 \to \mathbb{R}$. The composition of continuous functions is continuous. Since $\text{clip}(x, y, z)$ is a composition of these continuous functions, it is itself continuous. □

Lemma A.2 (Exchange of limit and clip)**.** *Let* $\{f_n\}_{n=1}^{\infty}$, $\{L_n\}_{n=1}^{\infty}$, *and* $\{U_n\}_{n=1}^{\infty}$ *be sequences of real numbers that converge to the limits* f*,* L*, and* U *respectively, that is,*

$$\lim_{n\to\infty} f_n = f\,, \quad \lim_{n\to\infty} L_n = L\,, \quad \lim_{n\to\infty} U_n = U\,.$$

Then, the limit of the clipped sequence is the clip of the limits:

$$\lim_{n\to\infty} \text{clip}(f_n, L_n, U_n) = \text{clip}(f, L, U) \tag{28}$$

Proof. Let the vector sequence be defined as $V_n = (f_n, L_n, U_n) \in \mathbb{R}^3$. By the definition of convergence of a vector sequence, since each component converges, the vector sequence converges to the vector limit $V = (f, L, U)$, that is,

$$\lim_{n\to\infty} V_n = V\,.$$

By definition of function continuity for sequences, if a function $g : \mathbb{R}^3 \to \mathbb{R}$ is continuous at a point V, and a sequence $V_n \to V$, then the sequence $g(V_n)$ must converge to $g(V)$, that is,

$$\lim_{n\to\infty} g(V_n) = g(V)\,.$$

From Lemma A.1, we know that the function $g(V) = \text{clip}(V)$ is continuous everywhere. We can therefore substitute our sequence V_n and its limit V into the continuity definition, getting

$$\lim_{n\to\infty} \text{clip}(f_n, L_n, U_n) = \text{clip}(f, L, U)\,,$$

which is equivalent to:

$$\lim_{n\to\infty} \text{clip}(f_n, L_n, U_n) = \text{clip}\Big(\lim_{n\to\infty} f_n, \lim_{n\to\infty} L_n, \lim_{n\to\infty} U_n\Big)\,.$$

□

Lemma A.3 (Asymptotic expansion of clip)**.** *Let* $\hat{C} = n^{-1}\sum_{i=1}^{n} R_i\hat{\mu}(X_i)/\hat{\pi}(X_i)$*. Assume the asymptotic expansions in Equation 20 hold. Then we have*

$$\text{clip}\big(\hat{C}, L_n, U_n\big) = \theta^\star + \frac{1}{\sqrt{n}}\text{clip}\left(Z_{\text{correction}}, \min\{Z_{\texttt{OR}}, Z_{\texttt{IPW}}\}, \max\{Z_{\texttt{OR}}, Z_{\texttt{IPW}}\}\right) + o_p(n^{-1/2})\,, \tag{29}$$

where $L_n = \min\{\hat{\theta}_{\texttt{OR}}, \hat{\theta}_{\texttt{IPW}}\}$ *and* $U_n = \max\{\hat{\theta}_{\texttt{OR}}, \hat{\theta}_{\texttt{IPW}}\}$.

Proof. The proof proceeds by substituting the given asymptotic expansions into the clip function and simplifying, leveraging the properties of the min, max, and clip operators. We first expand the clipping bounds. For L_n:

$$L_n = \min\{\hat{\theta}_{\texttt{OR}}, \hat{\theta}_{\texttt{IPW}}\} = \min\left\{\theta^\star + \frac{Z_{\texttt{OR}}}{\sqrt{n}} + o_p(n^{-1/2}), \theta^\star + \frac{Z_{\texttt{IPW}}}{\sqrt{n}} + o_p(n^{-1/2})\right\}. \tag{30}$$

Since $\min(a+x, a+y) = a + \min(x,y)$, we can factor out the common $\theta^\star$ term, getting:

$$L_n = \theta^\star + \min\left\{\frac{Z_{\texttt{OR}}}{\sqrt{n}} + o_p(n^{-1/2}), \frac{Z_{\texttt{IPW}}}{\sqrt{n}} + o_p(n^{-1/2})\right\}. \tag{31}$$

The same logic applies to the upper bound U_n. Now we can substitute the expansions for $\hat{C}$, L_n, and U_n into the main expression:

$$\text{clip}\left(\hat{C}, L_n, U_n\right) = \text{clip}\left(\theta^\star + \frac{Z_{\text{correction}}}{\sqrt{n}} + o_p(n^{-1/2}), \theta^\star + \min\{\dots\}, \theta^\star + \max\{\dots\}\right). \tag{32}$$

The clip function is translation equivariant ($\text{clip}(x+a, y+a, z+a) = a + \text{clip}(x,y,z)$), so that we can factor out the common $\theta^\star$ term:

$$\text{clip}\left(\hat{C}, L_n, U_n\right) = \theta^\star + \text{clip}\left(\frac{Z_{\text{correction}}}{\sqrt{n}} + o_p(n^{-1/2}), \min\left\{\frac{Z_{\texttt{OR}}}{\sqrt{n}} + \dots\right\}, \max\left\{\frac{Z_{\texttt{IPW}}}{\sqrt{n}} + \dots\right\}\right) \tag{33}$$

The clip, min, and max functions are also scale equivariant for positive constants. We can factor out the $1/\sqrt{n}$ term:

$$\text{clip}\left(\hat{C}, L_n, U_n\right) = \theta^\star + \frac{1}{\sqrt{n}}\text{clip}\left(Z_{\text{correction}} + o_p(1), \min\{Z_{\texttt{OR}} + o_p(1), Z_{\texttt{IPW}} + o_p(1)\}, \max\{\dots\}\right) \tag{34}$$

Because the clip, min, and max functions are all continuous, the small $o_p(1)$ terms (which converge to zero in probability) can be pulled out of the functions, resulting in a single remainder term:

$$\text{clip}\left(\hat{C}, L_n, U_n\right) = \theta^\star + \frac{1}{\sqrt{n}}\text{clip}\left(Z_{\text{correction}}, \min\{Z_{\texttt{OR}}, Z_{\texttt{IPW}}\}, \max\{Z_{\texttt{OR}}, Z_{\texttt{IPW}}\}\right) + o_p(n^{-1/2}). \tag{35}$$

This completes the proof. □

B Proof of main results

B.1 Proof of Lemma 2.2

Proof. We write φ^F to indicate the full-data influence function for notational simplicity. Given a full-data influence function φ^F, by Theorem 7.2 in Tsiatis (2006), we know that the space of associated observed-data influence functions $\Lambda^\perp$ is given by

$$\Lambda^\perp = \left\{\frac{R}{\pi^\star(X)}\varphi^F \oplus \Lambda_2\right\}, \tag{36}$$

where $\Lambda_2 = \left\{L_2(\mathscr{D}) : \mathbb{E}\left[L_2(\mathscr{D}) \mid \mathscr{D}^F\right] = 0\right\}$.

By Theorem 10.1 in Tsiatis (2006), for a fixed φ^F, the optimal observed-data influence function among the class in Equation 36 is obtained by choosing

$$L_2(\mathscr{D}) = -\Pi\left(\frac{R}{\pi^\star(X)}\varphi^F \mid \Lambda_2\right), \tag{37}$$

where the operator Π projects the element $R\varphi^F/\pi^\star(X)$ onto the space Λ_2 – see Theorem 2.1 in Tsiatis (2006).

Finally, by Theorem 10.2 in Tsiatis (2006), we know that

$$\Pi\left(\frac{R}{\pi^\star(X)}\varphi^F \mid \Lambda_2\right) = \left(\frac{R-\pi^\star(X)}{\pi^\star(X)}\right)h_2(X) \in \Lambda_2\,, \tag{38}$$

where $h_2(X) = \mathbb{E}\left[\varphi^F \mid X\right]$. This concludes the proof. □

B.2 Proof of Proposition 2.3

Proof. Under Assumption 2.1, we have

$$\mathbb{E}\left[\hat{\theta}_{\texttt{IPW}}\right] = \mathbb{E}\left[\frac{RY}{\hat{\pi}(X)}\right] = \mathbb{E}\left[\frac{\pi^\star(X)\mu^\star(X)}{\hat{\pi}(X)}\right]. \tag{39}$$

Then, given the fact that $\hat{\mu} = \mu^\star$, we have

$$\mathbb{E}\left[\hat{\theta}_{\texttt{IPW}}\right] = \mathbb{E}\left[\frac{\pi^\star(X)\mu^\star(X)}{\hat{\pi}(X)}\right] = \mathbb{E}\left[\frac{\pi^\star(X)\hat{\mu}(X)}{\hat{\pi}(X)}\right], \tag{40}$$

which is the expected value of the quantity $n^{-1}\sum_{i=1}^n R_i\hat{\mu}(X_i)/\hat{\pi}(X_i)$.

Similarly, under Assumption 2.1 and $\hat{\pi} = \pi^\star$, we have

$$\mathbb{E}\left[\hat{\theta}_{\texttt{OR}}\right] = \mathbb{E}[\hat{\mu}(X)] = \mathbb{E}\left[\frac{\pi^\star(X)\hat{\mu}(X)}{\pi^\star(X)}\right] = \mathbb{E}\left[\frac{R\hat{\mu}(X)}{\hat{\pi}(X)}\right], \tag{41}$$

which is the expected value of $n^{-1}\sum_{i=1}^n R_i\hat{\mu}(X_i)/\hat{\pi}(X_i)$. These facts imply $\mathbb{E}\left[\hat{\theta}_{\texttt{DR}}\right] = \theta^\star$. □

B.3 Proof of Theorem 3.2

Proof. The proof relies on Lemma A.2, Proposition 2.3, and the property of DR estimators. We know that, if a nuisance is well-specified, then $\hat{\theta}_{\texttt{DR}} \xrightarrow{p} \theta^\star$. This can be written as

$$\operatorname{plim}\left(\hat{\theta}_{\texttt{DR}}\right) = \operatorname{plim}\left(\hat{\theta}_{\texttt{OR}}\right) + \operatorname{plim}\left(\hat{\theta}_{\texttt{IPW}}\right) - \operatorname{plim}\left(\hat{C}\right) = \theta^\star\,, \tag{42}$$

where $\hat{C} = n^{-1}\sum_{i=1}^n R_i\hat{\mu}(X_i)/\hat{\pi}(X_i)$ denotes the correction term. Assume that $\hat{\mu} = \mu^\star$ (the proof for $\hat{\pi} = \pi^\star$ is identical). Then, by Proposition 2.3, we have

$$\operatorname{plim}\left(\hat{C}\right) = \operatorname{plim}\left(\hat{\theta}_{\texttt{IPW}}\right). \tag{43}$$

By Lemma A.2, we also know that

$$\operatorname{clip}\left(\operatorname{plim}\left(\hat{C}\right)\right) = \operatorname{plim}\left(\hat{C}\right), \tag{44}$$

as the limit $\operatorname{plim}\left(\hat{\theta}_{\texttt{IPW}}\right)$ is the boundary of the clipping operator. This implies

$$\operatorname{plim}\left(\hat{\theta}_{\texttt{DR+ACC}}\right) = \operatorname{plim}\left(\hat{\theta}_{\texttt{OR}}\right) + \operatorname{plim}\left(\hat{\theta}_{\texttt{IPW}}\right) - \operatorname{plim}\left(\operatorname{clip}\left(\hat{C}\right)\right) = \operatorname{plim}\left(\hat{\theta}_{\texttt{DR}}\right) = \theta^{\star}\,, \tag{45}$$

which delivers the result. □

B.4 Proof of Theorem 3.4

Proof. Let $\hat{C} = n^{-1}\sum_{i=1}^{n} R_i\hat{\mu}(X_i)/\hat{\pi}(X_i)$. The DR+ACC estimator is defined as

$$\hat{\theta}_{\texttt{DR+ACC}} = \hat{\theta}_{\texttt{OR}} + \hat{\theta}_{\texttt{IPW}} - \operatorname{clip}\left(\hat{C}\right). \tag{46}$$

The proof proceeds by first establishing that the value of $\hat{\theta}_{\texttt{DR+ACC}}$ is always bounded by the values of $\hat{\theta}_{\texttt{OR}}$ and $\hat{\theta}_{\texttt{IPW}}$. We consider three exhaustive cases for the value of $\hat{C}$:

1. If $\hat{C} \leq \min\{\hat{\theta}_{\texttt{OR}}, \hat{\theta}_{\texttt{IPW}}\}$, the clipped value is $\min\{\hat{\theta}_{\texttt{OR}}, \hat{\theta}_{\texttt{IPW}}\}$. The estimator becomes:

$$\hat{\theta}_{\texttt{DR+ACC}} = \hat{\theta}_{\texttt{OR}} + \hat{\theta}_{\texttt{IPW}} - \min\{\hat{\theta}_{\texttt{OR}}, \hat{\theta}_{\texttt{IPW}}\} = \max\{\hat{\theta}_{\texttt{OR}}, \hat{\theta}_{\texttt{IPW}}\}. \tag{47}$$

2. If $\hat{C} \geq \max\{\hat{\theta}_{\texttt{OR}}, \hat{\theta}_{\texttt{IPW}}\}$, the clipped value is $\max\{\hat{\theta}_{\texttt{OR}}, \hat{\theta}_{\texttt{IPW}}\}$. The estimator becomes:

$$\hat{\theta}_{\texttt{DR+ACC}} = \hat{\theta}_{\texttt{OR}} + \hat{\theta}_{\texttt{IPW}} - \max\{\hat{\theta}_{\texttt{OR}}, \hat{\theta}_{\texttt{IPW}}\} = \min\{\hat{\theta}_{\texttt{OR}}, \hat{\theta}_{\texttt{IPW}}\}. \tag{48}$$

3. If $\min\{\hat{\theta}_{\texttt{OR}}, \hat{\theta}_{\texttt{IPW}}\} < \hat{C} < \max\{\hat{\theta}_{\texttt{OR}}, \hat{\theta}_{\texttt{IPW}}\}$, the clip has no effect. The estimator is $\hat{\theta}_{\texttt{DR+ACC}} = \hat{\theta}_{\texttt{OR}} + \hat{\theta}_{\texttt{IPW}} - \hat{C}$. As $\hat{C}$ is between $\hat{\theta}_{\texttt{OR}}$ and $\hat{\theta}_{\texttt{IPW}}$, the value of $\hat{\theta}_{\texttt{OR}} + \hat{\theta}_{\texttt{IPW}} - \hat{C}$ must also lie in the interval $[\hat{\theta}_{\texttt{OR}}, \hat{\theta}_{\texttt{IPW}}]$.

In all possible cases, the estimator value is guaranteed to be in the closed interval defined by its components:

$$\min\{\hat{\theta}_{\texttt{OR}}, \hat{\theta}_{\texttt{IPW}}\} \leq \hat{\theta}_{\texttt{DR+ACC}} \leq \max\{\hat{\theta}_{\texttt{OR}}, \hat{\theta}_{\texttt{IPW}}\}. \tag{49}$$

This establishes the key "interval property" of the estimator. This implies that the DR+ACC estimator is a convex combination of the OR and IPW estimators, that is

$$\hat{\theta}_{\texttt{DR+ACC}} = \hat{\lambda}\hat{\theta}_{\texttt{OR}} + \left(1-\hat{\lambda}\right)\hat{\theta}_{\texttt{IPW}}\,, \tag{50}$$

where $\hat{\lambda} = (\hat{\theta}_{\texttt{OR}} - \hat{C})/(\hat{\theta}_{\texttt{OR}} - \hat{\theta}_{\texttt{IPW}}) \in [0,1]$.

Next, we relate this property to the estimation error. By subtracting $\theta^{\star}$ on both sides of the previous Equation, taking absolute values, and by triangle inequality, we get

$$|\hat{\theta}_{\texttt{DR+ACC}} - \theta^{\star}| \leq \hat{\lambda}|\hat{\theta}_{\texttt{OR}} - \theta^{\star}| + \left(1-\hat{\lambda}\right)|\hat{\theta}_{\texttt{IPW}} - \theta^{\star}|\,. \tag{51}$$

If $\hat{\lambda}$ is an independent estimate with respect to $\hat{\theta}_{\texttt{OR}}$ and $\hat{\theta}_{\texttt{IPW}}$, then we can compute the bias of DR+ACC:

$$\mathbb{E}\left[|\hat{\theta}_{\texttt{DR+ACC}} - \theta^{\star}|\right] \leq \mathbb{E}\left[\hat{\lambda}\right]\mathbb{E}\left[|\mu^{\star} - \hat{\mu}|\right] + \left(1 - \mathbb{E}\left[\hat{\lambda}\right]\right)\mathbb{E}\left[\left|\frac{\pi^{\star}}{\hat{\pi}} - 1\right||\mu^{\star}|\right]. \tag{52}$$

We can also easily derive the maximum bound. For any point x in an interval $[a, b]$ and any external point c, the distance $|x - c|$ is maximized at one of the endpoints, a or b. Therefore, the absolute error of our estimator is bounded by the maximum of the absolute errors of the OR and IPW estimators for any given sample:

$$|\hat{\theta}_{\texttt{DR+ACC}} - \theta^\star| \leq \max\{|\hat{\theta}_{\texttt{OR}} - \theta^\star|, |\hat{\theta}_{\texttt{IPW}} - \theta^\star|\}. \tag{53}$$

This inequality shows that the in-sample error of the DR+ACC estimator is always less than or equal to the maximum of the in-sample errors of its components. Computing the absolute bias of the OR and IPW estimators, we arrive at the final result:

$$\mathbb{E}\left[\left|\hat{\theta}_{\texttt{DR+ACC}} - \theta^\star\right|\right] \leq \max\left\{\mathbb{E}\left[|\mu^\star - \hat{\mu}|\right], \mathbb{E}\left[\left|\frac{\pi^\star}{\hat{\pi}} - 1\right| |\mu^\star|\right]\right\}. \tag{54}$$

□

B.5 Proof of Theorem 3.7

Proof. Let $\hat{C} = n^{-1}\sum_{i=1}^n R_i\hat{\mu}(X_i)/\hat{\pi}(X_i)$. Our goal is to find the limiting distribution of $\sqrt{n}(\hat{\theta}_{\texttt{DR+ACC}} - \theta^\star)$. We start by expressing the component estimators using an asymptotic expansion based on their limiting random variables:

$$\begin{aligned}
\hat{\theta}_{\texttt{OR}} &= \theta^\star + \frac{Z_{\texttt{OR}}}{\sqrt{n}} + o_p(n^{-1/2}), \\
\hat{\theta}_{\texttt{IPW}} &= \theta^\star + \frac{Z_{\texttt{IPW}}}{\sqrt{n}} + o_p(n^{-1/2}), \\
\hat{C} &= \theta^\star + \frac{Z_{\text{correction}}}{\sqrt{n}} + o_p(n^{-1/2}).
\end{aligned} \tag{55}$$

The clipping bounds can be expanded similarly:

$$\begin{aligned}
L_n &= \min\{\hat{\theta}_{\texttt{OR}}, \hat{\theta}_{\texttt{IPW}}\} = \theta^\star + \frac{1}{\sqrt{n}}\min\{Z_{\texttt{OR}}, Z_{\texttt{IPW}}\} + o_p(n^{-1/2}), \\
U_n &= \max\{\hat{\theta}_{\texttt{OR}}, \hat{\theta}_{\texttt{IPW}}\} = \theta^\star + \frac{1}{\sqrt{n}}\max\{Z_{\texttt{OR}}, Z_{\texttt{IPW}}\} + o_p(n^{-1/2}).
\end{aligned} \tag{56}$$

Now we analyze the clipped term. By Lemma A.3 we have:

$$\text{clip}\left(\hat{C}, L_n, U_n\right) = \theta^\star + \frac{1}{\sqrt{n}}\text{clip}\left(Z_{\text{correction}}, \min\{Z_{\texttt{OR}}, Z_{\texttt{IPW}}\}, \max\{Z_{\texttt{OR}}, Z_{\texttt{IPW}}\}\right) + o_p(n^{-1/2}) \tag{57}$$

Finally, we substitute these expansions into the expression for the scaled error of the final estimator:

$$\begin{aligned}
\sqrt{n}(\hat{\theta}_{\texttt{DR+ACC}} - \theta^\star) &= \sqrt{n}\left(\hat{\theta}_{\texttt{OR}} + \hat{\theta}_{\texttt{IPW}} - \text{clip}\left(\hat{C}, L_n, U_n\right) - \theta^\star\right) \\
&= \sqrt{n}\left(\left(\theta^\star + \frac{Z_{\texttt{OR}}}{\sqrt{n}}\right) + \left(\theta^\star + \frac{Z_{\texttt{IPW}}}{\sqrt{n}}\right) - \left(\theta^\star + \frac{\text{clip}\left(Z_{\text{correction}}\right)}{\sqrt{n}}\right) - \theta^\star\right) + o_p(1) \\
&= \sqrt{n}\left(\frac{Z_{\texttt{OR}} + Z_{\texttt{IPW}} - \text{clip}\left(Z_{\text{correction}}\right)}{\sqrt{n}}\right) + o_p(1) \\
&= Z_{\texttt{OR}} + Z_{\texttt{IPW}} - \text{clip}\left(Z_{\text{correction}}, \min\{Z_{\texttt{OR}}, Z_{\texttt{IPW}}\}, \max\{Z_{\texttt{OR}}, Z_{\texttt{IPW}}\}\right) + o_p(1).
\end{aligned} \tag{58}$$

As $n \to \infty$, the $o_p(1)$ term vanishes, which completes the proof. □

C Additional simulation results

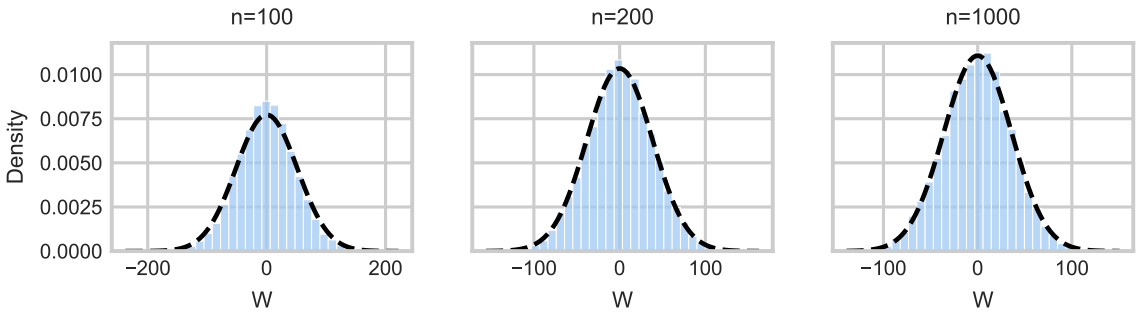

Figure C.1: Comparison between empirical distribution of W and Gaussian approximation of DR (black dashed line) for a random simulation draw, when both nuisance models are well-specified. Full details of the simulation scenario are provided in Section 4.

Table C.1: Empirical coverage and confidence interval (CI) widths for a sample size of $n = 100$, grouped by nuisance model specification. The nominal coverage level is 95%.

Scenario	Estimator	Coverage	CI Width
Correct $\hat{\mu}$, Correct $\hat{\pi}$	OR	0.957	14.17
	IPW	1.000	133.95
	DR	0.968	15.38
	DR+ACC	0.964	14.90
Correct $\hat{\mu}$, Incorrect $\hat{\pi}$	OR	0.959	14.06
	IPW	1.000	195.62
	DR	0.968	19.54
	DR+ACC	0.967	18.35
Incorrect $\hat{\mu}$, Correct $\hat{\pi}$	OR	0.904	14.62
	IPW	1.000	135.91
	DR	0.964	19.70
	DR+ACC	0.947	17.17
Incorrect $\hat{\mu}$, Incorrect $\hat{\pi}$	OR	0.904	14.62
	IPW	1.000	195.62
	DR	0.947	29.46
	DR+ACC	0.932	17.69

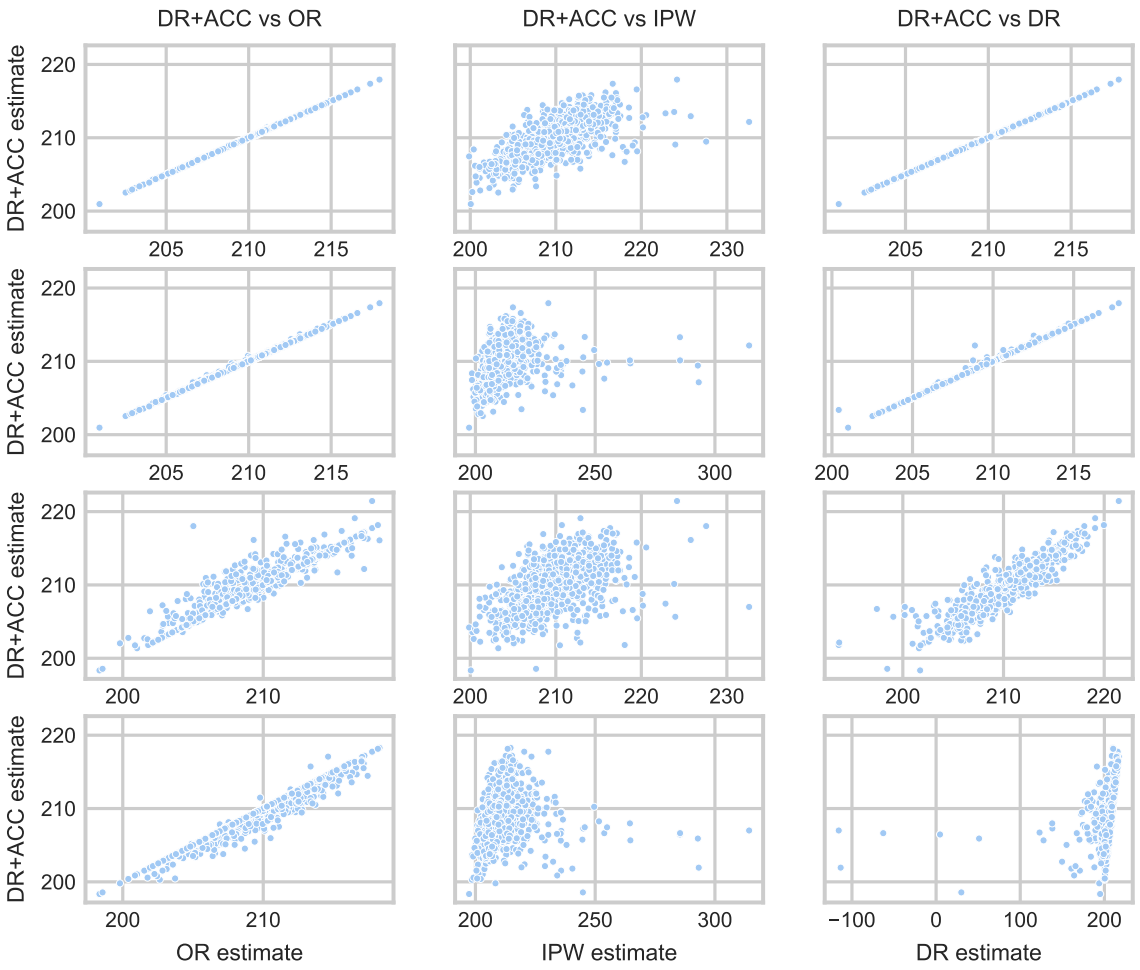

Figure C.2: Scatterplots comparing the performance of the DR+ACC estimator (y-axis) against the OR, IPW, and standard DR estimators (x-axes) for a sample size of $n = 200$. Each row corresponds to one of the four nuisance model specification scenarios (from top to bottom: correct $\hat{\mu}$/correct $\hat{\pi}$; correct $\hat{\mu}$/incorrect $\hat{\pi}$; incorrect $\hat{\mu}$/correct $\hat{\pi}$; incorrect $\hat{\mu}$/incorrect $\hat{\pi}$. The top three rows demonstrate that when at least one nuisance model is correct, the DR+ACC estimator closely tracks the standard DR estimator, which in turn tracks the well-specified estimator between OR and IPW, confirming its consistency. The bottom row illustrates the critical case of complete misspecification. The rightmost plot provides a direct visualization of the adaptive correction clipping mechanism: the DR+ACC estimate follows the standard DR when the latter is stable but remains bounded when the standard DR produces extreme, unstable estimates. This empirically demonstrates how the safety property of the DR+ACC estimator protects against the double fragility of the standard DR estimator. Full details of the simulation scenario are provided in Section 4.

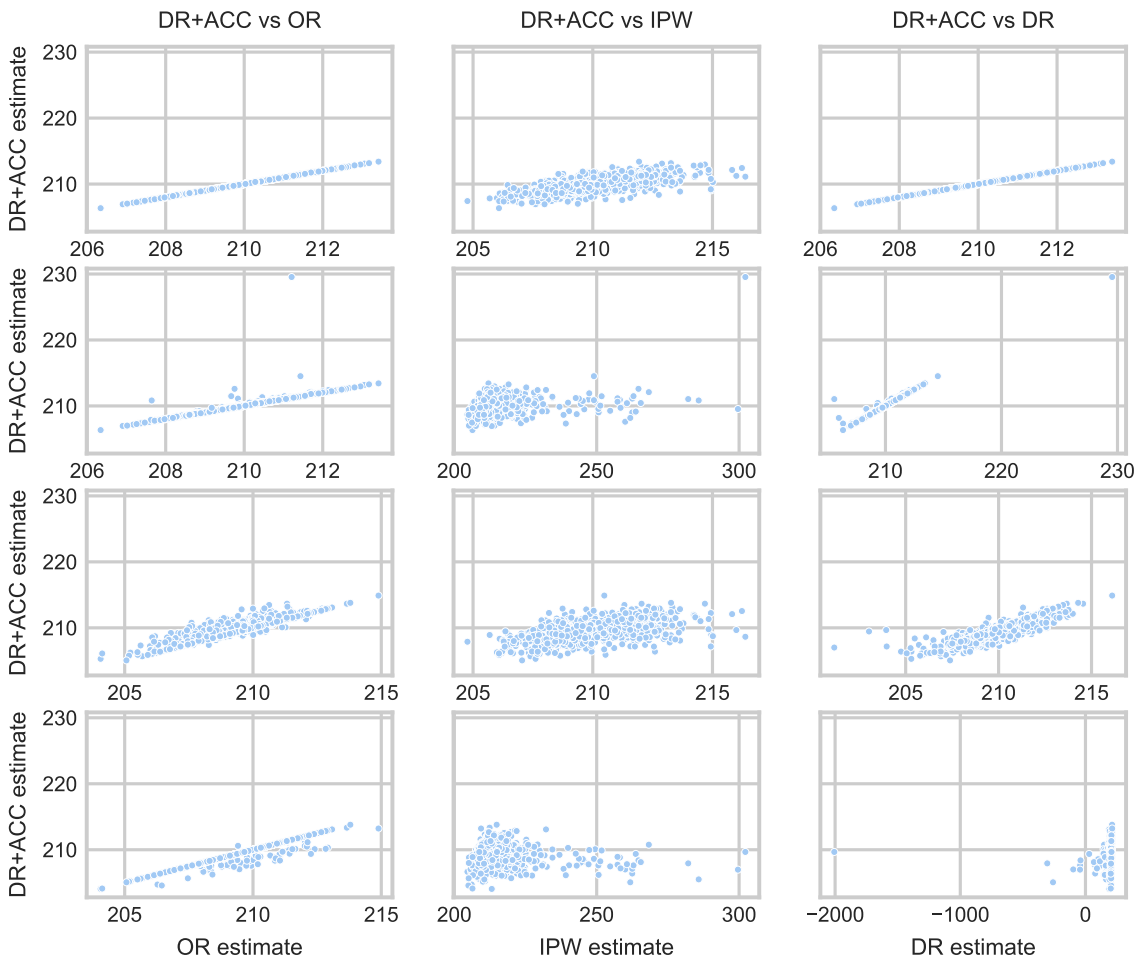

Figure C.3: Scatterplots comparing the performance of the DR+ACC estimator (y-axis) against the OR, IPW, and standard DR estimators (x-axes) for a sample size of $n = 1000$. Each row corresponds to one of the four nuisance model specification scenarios (from top to bottom: correct $\hat{\mu}$/correct $\hat{\pi}$; correct $\hat{\mu}$/incorrect $\hat{\pi}$; incorrect $\hat{\mu}$/correct $\hat{\pi}$; incorrect $\hat{\mu}$/incorrect $\hat{\pi}$. The top three rows demonstrate that when at least one nuisance model is correct, the DR+ACC estimator closely tracks the standard DR estimator, which in turn tracks the well-specified estimator between OR and IPW, confirming its consistency. The bottom row illustrates the critical case of complete misspecification. The rightmost plot provides a direct visualization of the adaptive correction clipping mechanism: the DR+ACC estimate follows the standard DR when the latter is stable but remains bounded when the standard DR produces extreme, unstable estimates. This empirically demonstrates how the safety property of the DR+ACC estimator protects against the double fragility of the standard DR estimator. Full details of the simulation scenario are provided in Section 4.

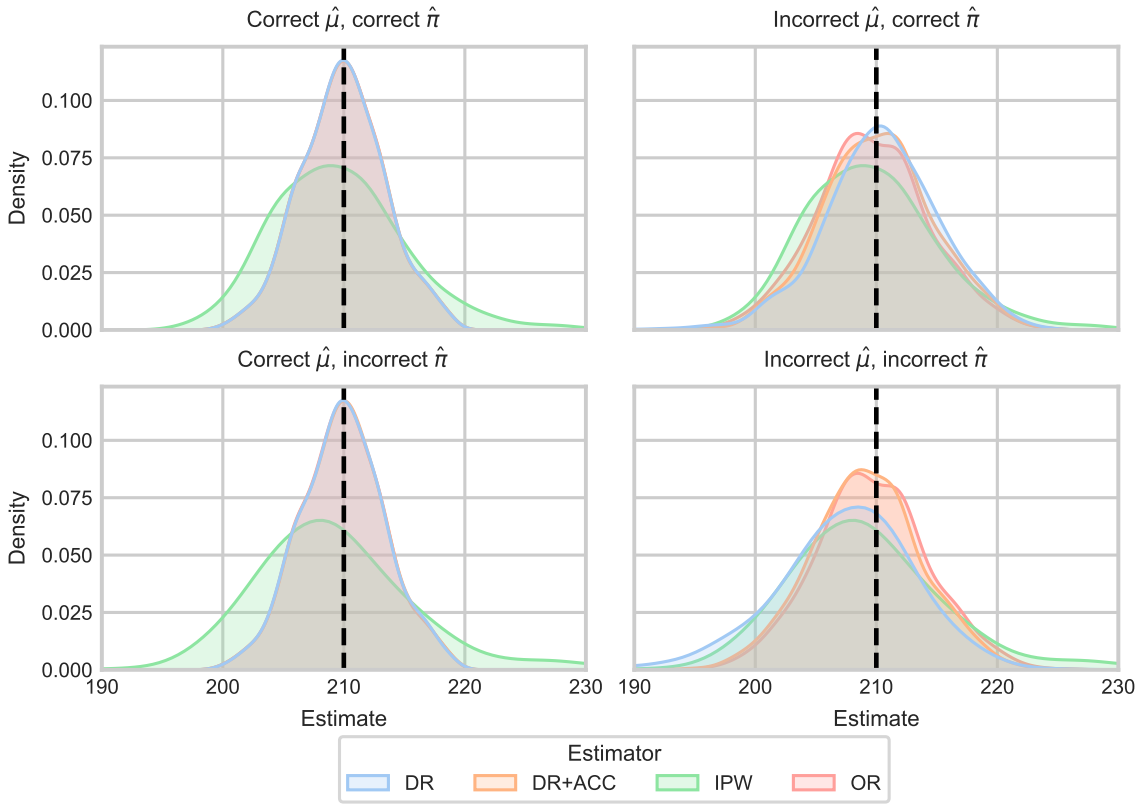

Figure C.4: Sampling distributions for the Outcome Regression (`OR`), Inverse Probability Weighting (`IPW`), Doubly Robust (`DR`), and our proposed `DR+ACC` estimators from 1000 simulations with a sample size of $n = 100$. The true parameter value is 210. The four panels show the estimators' performance under all combinations of correct and incorrect nuisance model specifications. The top row and bottom-left panel demonstrate the *asymptotic hard thresholding* property. When at least one nuisance model is correct, both `DR` and `DR+ACC` align with the correct simpler estimator. The bottom-right panel illustrates *double fragility*. When both nuisance models are wrong, the bias of the standard `DR` estimator can be worse than that of either the `OR` or `IPW` estimators. In this challenging scenario, our proposed `DR+ACC` estimator is shown to be *safe*, providing a much more stable and accurate estimate than the standard `DR` estimator. Full details of the simulation scenario are provided in Section 4.

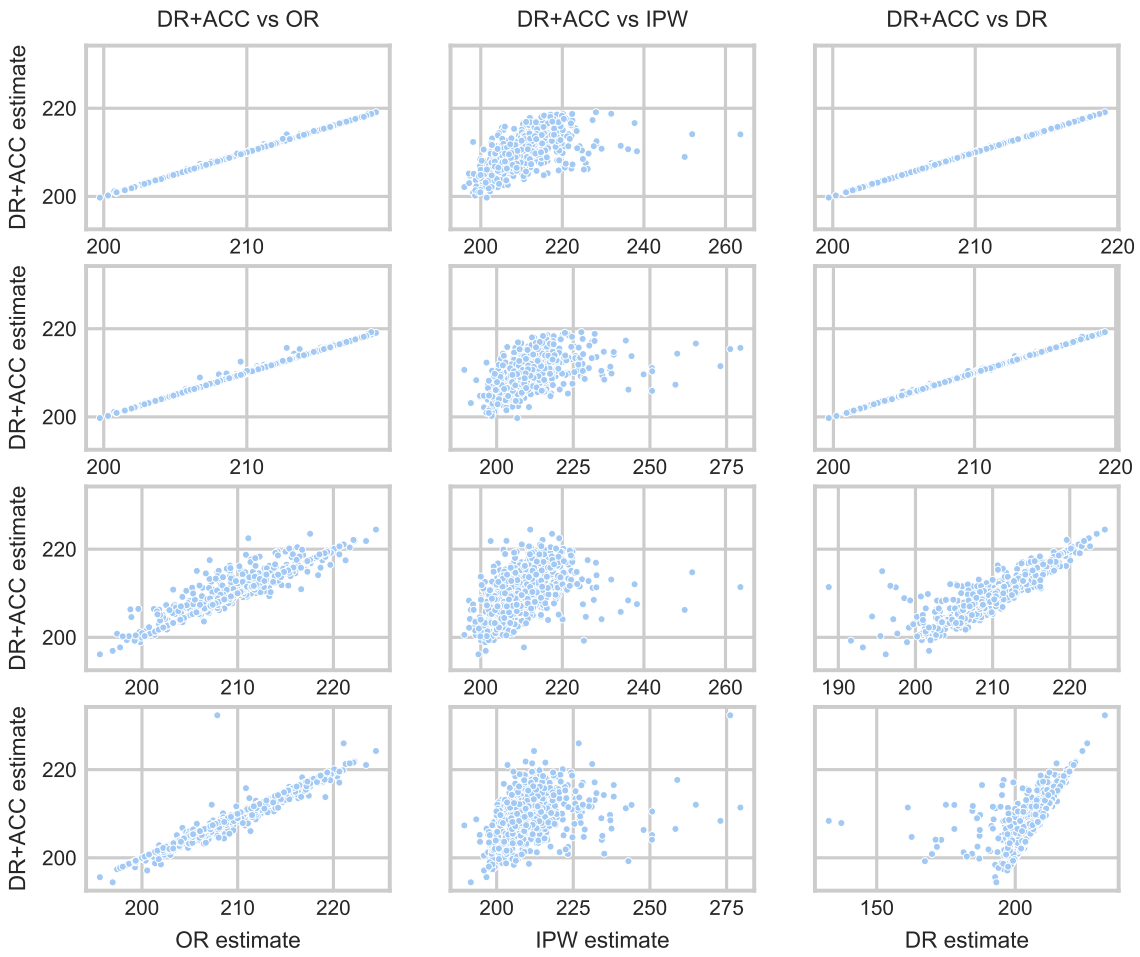

Figure C.5: Scatterplots comparing the performance of the DR+ACC estimator (y-axis) against the OR, IPW, and standard DR estimators (x-axes) for a sample size of $n = 100$. Each row corresponds to one of the four nuisance model specification scenarios (from top to bottom: correct $\hat{\mu}$/correct $\hat{\pi}$; correct $\hat{\mu}$/incorrect $\hat{\pi}$; incorrect $\hat{\mu}$/correct $\hat{\pi}$; incorrect $\hat{\mu}$/incorrect $\hat{\pi}$. The top three rows demonstrate that when at least one nuisance model is correct, the DR+ACC estimator closely tracks the standard DR estimator, which in turn tracks the well-specified estimator between OR and IPW, confirming its consistency. The bottom row illustrates the critical case of complete misspecification. The rightmost plot provides a direct visualization of the adaptive correction clipping mechanism: the DR+ACC estimate follows the standard DR when the latter is stable but remains bounded when the standard DR produces extreme, unstable estimates. This empirically demonstrates how the safety property of the DR+ACC estimator protects against the double fragility of the standard DR estimator. Full details of the simulation scenario are provided in Section 4.

D Additional application details

Table D.2: List of significant peptides and their related genes. Significance level 0.05. No multiple testing correction applied. We retrieve the UniProt Accession Number from the peptide string and use it to query the UniProt database API to get primary gene information.

Peptide	Gene name	Reference
IAAQGFTVAAILLGLAVTAMK	*HIGD2A*	Raju et al. (2024)
EGVEKAEETEQMIEK	*VAT1L*	Picard et al. (2025)
MGVAAHKK	*S100A8*	Litus et al. (2025)
RQDNEILIFWSK	*CRP*	Tao et al. (2021)
LLSMTLSPDLHMR	*TBCD*	Ikeda et al. (2017)
DAEVERDEER	*MYH14*	Finsterer (2019)
HVLVEYPMTLSLAAAQELWELAEQK	*BLVRA*	Mueller et al. (2010)
YC[+57]ADC[+57]EAK	*SMAP1*	Li and De Muynck (2021)
AREEQTPLHIASR	*ANK2*	Higham et al. (2019)
QNC[+57]ELFEQLGEYKFQNALLVR	*ALB*	Yang et al. (2025)
GDEEEEGEEKLEEK	*CANX*	Shen et al. (2025)
DGYHDNGMFSPSGESC[+57]...	*NELL2*	Shaker et al. (2022)
VPIPC[+57]YLIALVVGALESR	*LTA4H*	Adams et al. (2023)
TANKDHLVTAYNHLFETK	*FKBP3*	Blair et al. (2015)
TSVKEDLNVNEVFK	*RAB23*	Li (2011)
DAVTYTEHAKR	*H4C1*	Silvestro et al. (2021)
SLSTSGESLYHVLGLDK	*DNAJC5*	Rosene et al. (2023)
KGFSEGLWEIENNPTVK	*HDGF*	Hu et al. (2015)
LGAQLADLHLDNKK	*FN3KRP*	
SMDHATC[+57]ESR	*DAAM2*	Williams et al. (2021)
KPIDYTVLDDVGHGVK	*ABI1*	Yeh et al. (2021)
ALDLDSSC[+57]KEAADGYQR	*STIP1*	Lackie et al. (2020)
AQHEQYVAEAEEK	*CLIP2*	Serrano-Pozo et al. (2024)
SGGTDKDISAK, TMLPGEHQQVLSNLQSR	*DST*	
FTC[+57]TVTHTDLPSPLK	*IGHM*	Wang et al. (2017)
LTDIHGNVLQYHK	*BPNT1*	Maloney et al. (2019)
IPTHLFTFIQFK	*RO60*	Crooke III et al. (2022)
C[+57]IC[+57]PPGYSLQNEK	*FBN1*	Hanas et al. (2021)
FGQGAHHAAGQAGNEAGR	*SBSN*	Liu et al. (2016)
TNNVSEHEDTDKYR	*COPB1*	Yang et al. (2019)
VGATAAVYSAAILEYLTAEVLELAGNASK	*H2AZ1*	Luo et al. (2025)
LKHEC[+57]GAAFTSK	*CUL4A*	Yasukawa et al. (2020)
TAEHEAAQQDLQSK	*KTN1*	Li et al. (2024)
AVEEEDKMTPEQLAIK	*PSMD14*	Liu et al. (2024)

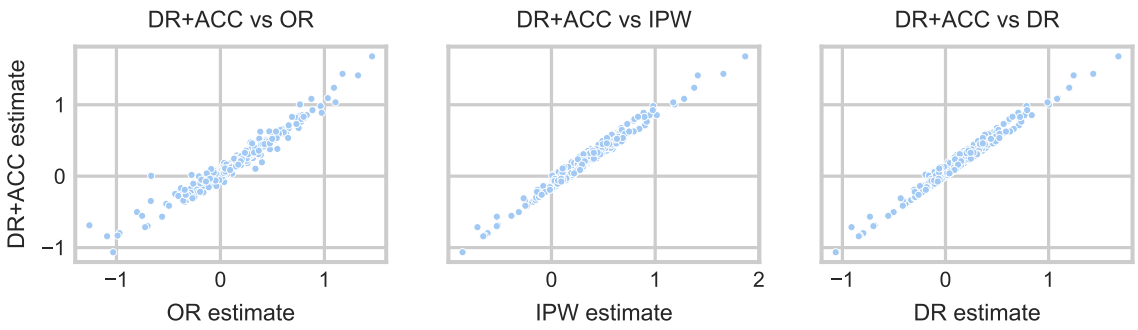

Figure D.6: Scatterplots comparing the estimates of the `DR+ACC` estimator (y-axis) against the `OR`, `IPW`, and standard `DR` estimators (x-axes) for the ATE associated to each of the 270 peptides. Full details of the application study are provided in Section 5.