Statistica Sinica Preprint No: SS-2025-0393	
Title	Response to Discussions of "Causal and Counterfactual Views of Missing Data Models"
Manuscript ID	SS-2025-0393
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202025.0393
Complete List of Authors	Razieh Nabi, Rohit Bhattacharya, Ilya Shpitser and James M. Robins
Corresponding Authors	Razieh Nabi
E-mails	razieh.nabi@emory.edu

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0393

RESPONSE TO DISCUSSIONS OF "CAUSAL AND COUNTERFACTUAL VIEWS OF MISSING DATA MODELS"

Razieh Nabi, Rohit Bhattacharya, Ilya Shpitser, James M. Robins

1. Introduction

We are grateful to the discussants – Levis and Kennedy [2025], Luo and Geng [2025], Wang and van der Laan [2025], and Yang and Kim [2025] – for their thoughtful comments on our paper [Nabi et al., 2025]. Below we summarize our main contributions before responding to each discussion in turn.

Graphical models have emerged as an important tool for clarifying identifying assumptions made in both causal inference [Pearl and Robins, 1995, Pearl, 2000] and missing data [Robins and Gill, 1997, Bhattacharya et al., 2019, Nabi et al., 2020, Malinsky et al., 2021, Mohan and Pearl, 2021]. Our paper [Nabi et al., 2025] shows how recent techniques motivated by causal graphical modeling may be fruitfully applied to obtain identification in missing data models. These recent techniques allow us to obtain novel identification results that would not be possible to obtain in standard causal inference problems, except by imposing implausible additional assumptions, such as rank preservation.

Specifically, we formalize how identifiability of the target (i.e., complete) data law can be viewed as identification of a joint distribution over counterfactuals $L^{(1)}$, the variables that

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0393

would have been observed if all missingness indicators R were set to one (that is, if no data were missing).

This reframing yields a counterfactual analogue of the g-formula. When assumptions encoded in the Markov properties of a missing data DAG (m-DAG) allow us to express the counterfactual g-formula in terms of the factuals, we obtain (nonparametric) identification. That is, noting that the target law $p(l^{(1)})$ satisfies $p(l^{(1)}) = p(l, R = 1)/p(R = 1 \mid l^{(1)})$, it follows that when the missingness selection model $p(R = 1 \mid L^{(1)})$ is identified from the observed law, so is $p(l^{(1)})$.

In our paper we used the word "nonparametric" in two different ways. The joint distribution of $(L, L^{(1)}, R)$ is Markov to a given m-DAG if it factorizes as the product of the conditional densities of each variable given its parents. In the graphical causal modeling literature, the model is said to be "nonparametric" just when these conditional densities are left unrestricted (with the exception that L is a deterministic function of its parents). The term "nonparametric identification" refers to identification in such a model. In contrast, in the missing data literature in statistics, a model is said to be nonparametric if it places no restrictions on either the observed data law or the target data law $p(l^{(1)})$. It is said to be nonparametric just identified (NPI) if the target law is identified from the observed data law. The permutation model [Robins, 1997] we discuss in our paper is known to be NPI. In contrast, all the other identified missing data models in our paper place testable restrictions on the joint distribution of the observed data. In fact, we conjecture that the permutation model is the only m-DAG model that is NPI. In the following, we use nonparametric in the graphical causal modeling sense.

A key message of our paper is that features specific to missing data models (in particular the partial observability of the counterfactuals through the proxies L and the structural restriction that R and L do not cause $L^{(1)}$) can deliver parameter identification in settings

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0393

where analogous parameters in hidden variable causal DAGs are not identified. We catalog several such nonparametric identification techniques in m-DAGs, and we clarify similarities and differences with standard causal identification, including when additional assumptions like rank preservation might be needed for causal analogues.

Each discussion extends or challenges our framework in important ways. Levis and Kennedy [2025] highlight identification strategies complementary to ours, based on the existence of instrumental and shadow variables, discuss semiparametric estimation theory for functionals arising from such strategies, as well as discuss connections between m-DAGs and Single World Intervention Graphs (SWIGs). Luo and Geng [2025] analyze self-censoring MNAR mechanisms with binary variables, deriving identifiability results that leverage auxiliary variables. Wang and van der Laan [2025] emphasize, as we do, viewing missingness as interventions, and connect our perspective to censoring in survival models. Yang and Kim [2025] and Wang and van der Laan [2025] both examine challenges of applying m-DAGs in applications, particularly assumption validation, scalability, and sensitivity analysis.

2. Response to the discussions

2.1 On the discussion by Levis and Kennedy

Levis and Kennedy [2025] draw attention to additional causal identification tools and highlight implications for estimation. In particular, they emphasize the relevance of instrumental variables, shadow variables, and SWIGs as complementary devices for reasoning about identification in missing data problems. They also consider how identification results derived from m-DAGs can be translated into practical estimation procedures with desirable asymptotic properties. Their discussion situates our contribution within a broader pipeline of causal inference methods and points toward promising directions for future methodological development.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0393

2.1 On the discussion by Levis and Kennedy

We appreciate Levis and Kennedy's thoughtful remarks. On their first point, we agree that instrumental variables [Sun et al., 2018], shadow variables [Miao et al., 2024], and related tools [Li et al., 2023] are useful complements. Our emphasis in the paper, however, was on a nonparametric identification framework: aside from the Markov restrictions encoded via independence assumptions in m-DAGs, we make no further distributional assumptions. In contrast, the use of IVs or shadow variables begins in settings where nonparametric identification fails, and proceeds by imposing additional restrictions (such as functional form constraints e.g., homogeneity assumptions when using of IVs), relevance assumptions which result in generic identification, or restrictions on the support of variables to restore identification. These extra ingredients move the analysis outside the purely nonparametric domain that was our focus.

On estimation, we appreciate the worked example they provide. While our goal in this paper was primarily conceptual, highlighting the philosophical parallels and differences between missing data and causal identification, their discussion underscores the importance of connecting identification results to practical estimation and efficiency theory. We agree this is important future work.

We also agree with Levis and Kennedy that causal inference problems in the presence of confounding often occur together with censoring, and that obtaining identification when both complications are present is challenging. We also appreciate their worked example illustrating these complications.

We would like to offer two notes of caution regarding generalizing the construction presented by Levis and Kennedy, first on the appropriate generalization of the SWIG splitting construction from causal diagrams to m-DAGs, and second on the general utility of using SWIGs for obtaining identification in missing data settings.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0393

On the translation of m-DAGs to SWIGs

For simple MAR models where the m-DAG's structure resembles that of the standard conditionally ignorable model in causal inference, the translation of an m-DAG to an equivalent SWIG is fairly natural. There are, however, important distinctions even in these simple settings. In causal inference, SWIGs provide a template indexed by treatment interventions (two templates if the treatment is binary) [Richardson and Robins, 2013]; in the missing data analogue, only the template for $R = 1$ is meaningful, since the counterfactual $L^{(0)}$ is not defined. This difference is highlighted in Figures 1(a, b) for a causal model satisfying ignorability and an analogous missing data model that is MCAR. Similarly, Figures 1(c, d) highlight the same for a conditionally ignorable causal model and a MAR missing data model.

Caution must be exercised when SWIGs are constructed from m-DAGs representing MNAR mechanisms. For instance, in a missing data model with self-censoring, a natural idea for what the SWIG should look like would be the graph shown in Figure 2(a). However, if we try to collapse the random and fixed nodes in this SWIG to reconstruct the observed data graph, we obtain the graph shown in Figure 2(b), which erroneously implies a cycle in the underlying data generating process.

The problem arises from conflating the full data variable $L^{(1)}$ in m-DAGs, and the observed data variable L relabeled to be counterfactual in the SWIG construction. An appropriate view of m-DAGs that avoids these types of difficulties is to view full data variables such as $L^{(1)}$ as unobserved confounders U, with additional structure imposed by missing data consistency. In this view, the SWIG corresponding to the self-censoring model is better represented by Figure 2(c), where the variable $L^{(1)}$ is viewed as an unobserved confounder U with special structure, which influences both the observability indicator R, and the observed proxy variable L. The m-DAG corresponding to this view of the self-censoring model is shown in Figure 2(d).

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0393

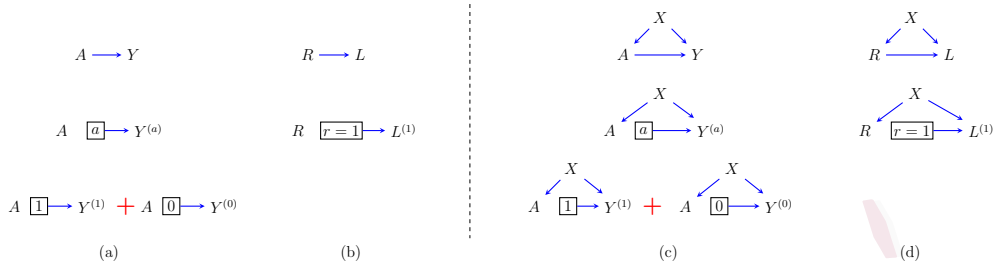

Figure 1: (a) Ignorable treatment model with SWIGs; (b) Its MCAR analogue, with a single relevant SWIG; (c) Conditionally ignorable treatment model with SWIGs; (d) Its MAR analogue, again with only one relevant SWIG.

SWIGs do not make identification arguments for MNAR problems clearer

The primary motivation for SWIGs in the causal inference context is to provide a graphical representation for independences that arise in identification arguments. The function of m-DAGs is to provide precisely the same graphical representation in missing data problems, rendering SWIGs unnecessary.

Take, for example, the independences that can be read from the m-DAG in Figure 3(a) and the corresponding SWIG in Figure 3(b). From either graph, we are able to extract independences of the form $R_k \perp\!\!\!\perp L_k^{(1)}, R_{-k} \mid L_{-k}^{(1)}$, for $k \in \{1, 2\}$ that define the binary block-parallel MNAR model. That is, the m-DAG itself is just as expressive as the SWIG for reading off independences important for identification in missing data models. In fact, we now argue that attempting identification using standard causal inference arguments from the SWIG may be problematic

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0393

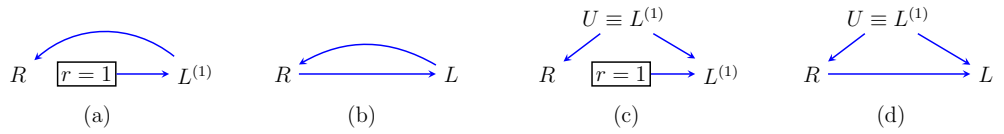

Figure 2: (a) A possible SWIG representation of a self-censoring missing data model; (b) Stitching SWIGs into an observed-data model produces a cycle; (c) A SWIG that draws a distinction between the full data variable $L^{(1)}$, viewed as an unobserved confounder U with extra restrictions, and the observed variable L under an intervention where R is set to 1; (d) The full data graph equating the full data variable $L^{(1)}$ with an unobserved confounder U with special structure.

for missing data problems due to the single counterfactual nature of missing data, which as we discussed in Section 6 of our paper is more akin to the rank preservation assumption in causal inference.

To obtain the identification of the target law $p(l_1^{(1)}, l_2^{(1)})$ of the block-parallel MNAR model, we showed in Section 5.2 of our paper that a parallel use of the g-formula was required, which never arose in standard causal inference settings. If just the marginal $p(l_2^{(1)})$ is desired, this is obtained by marginalization. Now suppose instead, we attempted to perform identification arguments for $p(l_2^{(1)})$ using the SWIG, in a manner similar to the derivation of the adjustment formula for causal inference. This would proceed as follows. First notice from the SWIG in Figure 3(b) that $L_2^{(1)} \perp\!\!\!\perp R_2 \mid U_1, R_1$ – this is analogous to conditional ignorability except U_1 is

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0393

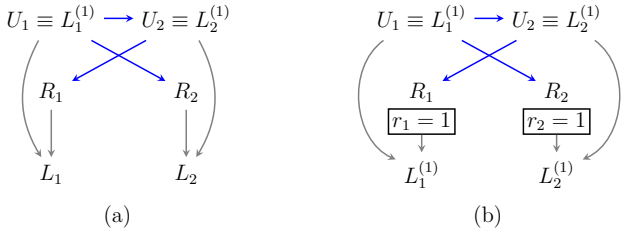

Figure 3: (a) The block-parallel MNAR model, labeling missing variables as unmeasured for the purposes of constructing a SWIG; (b) The corresponding SWIG obtained from (a).

observed only when $R_1 = 1$. Thus we have,

$$\begin{aligned} p(l_2^{(1)}) &= \sum_{r_1,u_1} p(r_1, u_1, l_2^{(1)}) \\ &= \sum_{r_1,u_1} p(r_1, u_1)\, p(l_2^{(1)} \mid r_1, u_1) \\ &= \sum_{r_1,u_1} p(r_1, u_1)\, p(l_2^{(1)} \mid r_1, u_1, r_2 = 1) \\ &= \sum_{r_1,u_1} p(r_1, u_1)\, p(l_2 \mid r_1, u_1, r_2 = 1). \end{aligned}$$

Despite the final expression above appearing to be a function that is devoid of counterfactuals, the appearance of the variable U_1 in the expression prevents identification, as U_1 is observed only when $R_1 = 1$. Since R_1 cannot be set to the value 1 in the formula above, we are unable to establish identification via the arguments presented. In fact, no sequential strategy that we are aware of based on the SWIG in Figure 3(b) would help establish identification. That is, the parallel fixing arguments presented in Section 5.2 of our paper are required here.

In short, while applying the SWIG construction to m-DAGs offers useful insights in causal problems with simple types of missingness, such as MCAR or MAR, we caution that such

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0393

constructions must be done with care, as Levis and Kennedy have done.

To illustrate, Figure 4(a) extends the example of Levis and Kennedy to include missingness in both the outcome Y and treatment A. The goal here is to identify the average causal effect of A on Y in the absence of censoring.

Just as in the previous example shown in Figures 3(a, b), an argument based solely on SWIGs does not yield identification. Instead, identification strategies that leverage restrictions encoded in both m-DAGs and SWIGs are necessary. In particular, we first obtain the target law $p(X, A^{(1)}, Y^{(1)})$ by applying inverse probability weighting with the product of the propensity scores for R_1 and R_2, resulting in the graph in Figure 4(b). We then apply the g-formula on $A^{(1)}$ to adjust for confounding by X, taking advantage of the restrictions in the SWIG shown in Figure 4(c). This yields $p(Y^{(1,a)})$, the distribution of the outcome Y, had it not been censored and had treatment been set to value a, from which the average causal effect of interest may be obtained.

Finally, we note that there is currently no complete graphical identification theory for causal parameters associated with arbitrary m-DAGs that encode both MNAR and (possible) confounding by unmeasured common causes U. We expect this theory to employ both SWIG and m-DAG constructions, as in the worked example of Levis and Kennedy, and our examples above.

2.2 On the discussion by Luo and Geng

Luo and Geng [2025] extend our framework by focusing on self-censoring MNAR mechanisms, where the missingness of a variable depends directly on its unobserved value. They study such models with binary outcomes and establish identifiability results under explicit, testable conditions. Their discussion highlights how auxiliary variables, either baseline or follow-up

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0393

measurements, can be leveraged to restore identification. In doing so, they illustrate that self-censoring structures, which were not emphasized in our paper, can still yield identification in important cases. They also point to future directions for extending these ideas beyond the binary case to more general settings.

We thank Luo and Geng for their examples of self-censoring mechanisms. In our paper, we noted self-censoring as a form of MNAR but did not explore it further, since m-DAG models with self centering are never nonparametrically identified [Mohan et al., 2013]. Since our emphasis was on nonparametric identification, we did not employ any restrictions not implied by the m-DAG factorization, including (i) relevant restrictions needed by instrumental variable or proxy methods, or (ii) constraints implied by state space restrictions. As Luo and Geng [2025] illustrated, such assumptions can sometimes yield identifiability, and their results highlight concrete testable conditions under which this occurs.

We fully agree that self-censoring is of practical importance, with income surveys, sensitive health questions, and other social science settings providing common examples. Their discussion usefully demonstrates how auxiliary information, through baseline or follow-up variables, can be leveraged to address such scenarios.

More broadly, we believe that in MNAR submodels that identify the target law, restrictions implied by the m-DAG always yield testable implications for the observed data law, with the exception of the permutation model [Robins, 1997]. Examples of such tests have been developed in prior work on graphical models [Mohan and Pearl, 2014, Nabi and Bhattacharya, 2023, Guo et al., 2023, Chen et al., 2023].

Overall, we view their discussion as complementary to our focus: while our framework aimed to catalog nonparametric identification results, their work illustrates how additional structure can both sharpen identifiability and connect to practical applications.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0393

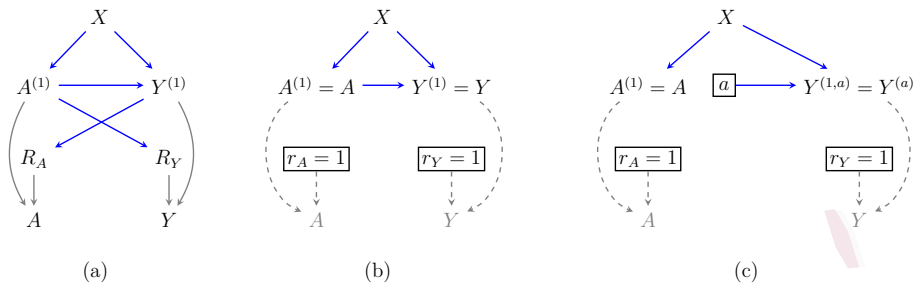

Figure 4: (a) Extension of Levis and Kennedy's example to have two missing variables. (b) Graph obtained after fixing R_A and R_Y in parallel. (c) SWIG obtained by splitting the treatment variable A after already having fixed R_A and R_Y.

2.3 On the discussion by Wang and van der Laan

Wang and van der Laan [2025], like us, approach missingness through the lens of interventions. They emphasize the close relationship between missing data models and multivariate censoring models in survival analysis. They carefully prove that the permutation missingness model is MNAR rather than MAR and flesh out our example of why the permutation model is substantively plausible in some real world applications. They also discuss substantive settings where the block sequential model is plausible.

They are less certain that the other identifiable MNAR models discussed in our paper are substantively plausible. In our view, it is difficult to establish plausibility of *any* missing data model in practice, whether it is formulated graphically or not. One advantage of the sort of theory we present is a single formulation for a large class of identifiable models (including existing models in the literature such as those described in [Zhou et al., 2010, Robins, 1997]).

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0393

Since no single model may be plausible in applications, an alternative strategy is to perform the analysis of interest under a wide class of identifiable missingness models, as a form of sensitivity analysis.

We make one final observation regarding the plausibility of (or the lack there of) the identifiable MNAR models considered in the paper. Identification in every case we discuss relies on an assumption that one or more counterfactuals $L_j^{(1)}$ suffice to control confounding of the "effect" of R_k on L_k where $j \neq k$. But why should we privilege $L_j^{(1)}$ (or a set of such variables) as a set sufficient for adjustment? It seems much more plausible that there exist other unmeasured common causes of R_k and L_k (equivalently $L_k^{(1)}$) that would also need to be adjusted for to eliminate confounding. Are we assuming the $L_j^{(1)}$ suffice to control confounding because we believe it to be so or rather because we want to achieve identification for identification's sake? In fact, similar plausibility concerns arise in non-graphical identifiable models, such as non-monotone MAR. In our view, plausible models of missingness feature a description of data generation that follows a temporal order (often representable as a DAG). Unfortunately, to be realistic, even models of this type would feature unobserved confounding of the sort that would prevent identifiability.

2.4 On the discussion by Yang and Kim

Yang and Kim [2025] emphasize the practical challenges of applying the m-DAG framework. They point out that while m-DAGs provide a principled way to encode assumptions and extend identification theory, their utility in practice depends critically on correct graph specification, which may be difficult to achieve in applied settings. They raise concerns about scalability in high-dimensional problems, where the number of variables and missingness patterns grows quickly, and about the feasibility of validating the conditional independence assumptions en-

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0393

coded in an m-DAG. They further stress the importance of developing systematic approaches to sensitivity analysis, noting that small misspecifications of the graph can have consequences for identification and downstream tasks.

We thank Yang and Kim for raising important questions about the practical utility of m-DAGs, and graphical modeling more broadly. Our goal in this paper was primarily conceptual: to provide a general framework that connects causal and missing data perspectives and to catalog identification results that arise from this connection. Their emphasis on practical feasibility is appreciated, and complementary to our focus. That said, continuing work in the graphical modeling community over the last two decades has rendered much of the critiques of the graphical modeling approach to causal inference and missing data problems out of date.

Identification of causal and missing data parameters must, by necessity, rely on equality restrictions in the full data distribution. Graphical models aim to encode these restrictions in a way that allows the construction of a plausible data generating mechanism consistent with a temporal order of events. Models that obtain identifiability without a corresponding graphical representation, for instance the non-monotone missing at random (MAR) model, often lack such a plausible mechanism. Indeed, existing efforts that argue for the plausibility of models such as non-monotone MAR embed them into a graphical model or a mixture of such models [Robins and Gill, 1997].

Over the last decade, an explosion of easy to use open-source packages that take advantage of graphical models have been developed for all tasks in causal inference and missing data, including establishing identification, constructing estimators and applying them to data, conducting sensitivity analyses, establishing bounds on non-identified parameters, and model selection. A non-exhaustive list of these packages includes: Ananke [Lee et al., 2023], autobounds [Duarte et al., 2024], dosearch [Tikka et al., 2021], DAGitty [Textor et al., 2016], software developed

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0393

as part of the Tetrad project (equipped with Python and R interfaces) [Ramsey and Andrews, 2023], as well the pcalg package [Kalisch et al., 2012], and extensions to deal with missing values [Andrews et al., 2024].

Further, sensitivity analyses results can be fruitfully applied to graphical models. For example, nongraphical results in Robins et al. [1999] can easily be represented in graphical structures that encode MAR, conditional or sequential ignorability, and other types of Markov restrictions. For instance, the permutation model in Robins [1997] is shown to be a graphical model in our paper.

In addition, graphical models allow a particularly powerful form of nonparametric sensitivity analysis, where consistent inferences are made in union models defined over a potentially large class of graphs that the analyst is uncertain about; see for example, [VanderWeele and Shpitser, 2011, Yang et al., 2024, Shpitser and VanderWeele, 2010, Shpitser et al., 2010, Chang et al., 2024, Wang et al., 2025]. For instance, the result in [VanderWeele and Shpitser, 2011] states that identification of the causal effect by covariate adjustment may be formulated without precise knowledge of the graph, but only via the set of common causes of the treatment and outcome, provided *an* adjustment set exists. The method developed by Yang et al. [2024] provide similar robustness guarantees with more complex types of identifying functionals, including the front-door functional, and the ratio functional arising in instrumental variable analysis.

We share the authors' view that algorithmic and computational advances will be needed to bring graphical identification methods to bear on high-dimensional problems with complex missingness patterns. We see this as an exciting frontier for future work, and one where continued integration of causal inference tools with missing data methodology will be especially fruitful. We believe smoothing and sparsity methods, as well as sum-product algorithm type methods developed in the graphical modeling literature may all be relevant for these advances

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0393

[Lee et al., 2021].

3. Conclusion

We thank the discussants for their thoughtful contributions. Together, their contributions point toward a roadmap: combine graphical identification, auxiliary information, intervention-based censoring perspectives, efficiency-oriented estimation, and structured sensitivity analysis into a practical toolkit for MNAR problems.

References

Ryan M. Andrews, Christine W. Bang, Vanessa Didelez, Janine Witte, and Ronja Foraita. Software application profile: tpc and micd—R packages for causal discovery with incomplete cohort data. *International Journal of Epidemiology*, 53(5):dyae113, 2024.

Rohit Bhattacharya, Razieh Nabi, Ilya Shpitser, and James Robins. Identification in missing data models represented by directed acyclic graphs. In *Proceedings of the Thirty Fifth Conference on Uncertainty in Artificial Intelligence (UAI-35th)*. AUAI Press, 2019.

Ting-Hsuan Chang, Zijian Guo, and Daniel Malinsky. Post-selection inference for causal effects after causal discovery. *arXiv preprint arXiv:2405.06763*, 2024.

Jacob M. Chen, Daniel Malinsky, and Rohit Bhattacharya. Causal inference with outcome-dependent missingness and self-censoring. In *Uncertainty in Artificial Intelligence*, pages 358–368. PMLR, 2023.

Guilherme Duarte, Noam Finkelstein, Dean Knox, Jonathan Mummolo, and Ilya Shpitser. An automated approach to causal inference in discrete settings. *Journal of the American Statistical Association*, 119(547):1778–1793, 2024.

REFERENCES

Anna Guo, Jiwei Zhao, and Razieh Nabi. Sufficient identification conditions and semiparametric estimation under missing not at random mechanisms. In *Uncertainty in Artificial Intelligence*, pages 777–787. PMLR, 2023.

Markus Kalisch, Martin Mächler, Diego Colombo, Marloes H. Maathuis, and Peter Bühlmann. Causal inference using graphical models with the R package pcalg. *Journal of Statistical Software*, 47:1–26, 2012.

Jaron J.R. Lee, Agatha S. Mallett, Ilya Shpitser, Aimee Campbell, Edward Nunes, and Daniel O. Scharfstein. Markov-restricted analysis of randomized trials with non-monotone missing binary outcomes. *arXiv preprint arXiv:2105.08868*, 2021.

Jaron JR Lee, Rohit Bhattacharya, Razieh Nabi, and Ilya Shpitser. Ananke: A python package for causal inference using graphical models. *arXiv preprint arXiv:2301.11477*, 2023.

Alexander W. Levis and Edward H. Kennedy. Discussion of "Causal and counterfactual views of missing data models" by Razieh Nabi, Rohit Bhattacharya, Ilya Shpitser, and James M. Robins. *Statistica Sinica*, 2025. doi: 10.5705/ss.202025.0210. URL `https://www3.stat.sinica.edu.tw/preprint/SS-2025-0210_Preprint.pdf`.

Yilin Li, Wang Miao, Ilya Shpitser, and Eric J Tchetgen Tchetgen. A self-censoring model for multivariate nonignorable nonmonotone missing data. *Biometrics*, 79(4):3203–3214, 2023.

Shanshan Luo and Zhi Geng. Discussion on "Causal and counterfactual views of missing data models". *Statistica Sinica*, 2025. doi: 10.5705/ss.202025.0152. URL `https://www3.stat.sinica.edu.tw/preprint/SS-2025-0152_Preprint.pdf`.

Daniel Malinsky, Ilya Shpitser, and Eric J Tchetgen Tchetgen. Semiparametric inference for

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0393

nonmonotone missing-not-at-random data: the no self-censoring model. *Journal of the American Statistical Association*, pages 1–9, 2021.

Wang Miao, Lan Liu, Yilin Li, Eric J. Tchetgen Tchetgen, and Zhi Geng. Identification and semiparametric efficiency theory of nonignorable missing data with a shadow variable. *ACM/JMS Journal of Data Science*, 1(2):1–23, 2024.

Karthika Mohan and Judea Pearl. On the testability of models with missing data. In *Artificial Intelligence and Statistics*, pages 643–650. PMLR, 2014.

Karthika Mohan and Judea Pearl. Graphical models for processing missing data. *Journal of the American Statistical Association*, pages 1–16, 2021.

Karthika Mohan, Judea Pearl, and Jin Tian. Graphical models for inference with missing data. In *Advances in Neural Information Processing Systems 26*, pages 1277–1285. Curran Associates, Inc., 2013.

Razieh Nabi and Rohit Bhattacharya. On testability and goodness of fit tests in missing data models. In *Uncertainty in Artificial Intelligence*, pages 1467–1477. PMLR, 2023.

Razieh Nabi, Rohit Bhattacharya, and Ilya Shpitser. Full law identification in graphical models of missing data: Completeness results. In *Proceedings of the Twenty Seventh International Conference on Machine Learning (ICML-20)*, 2020.

Razieh Nabi, Rohit Bhattacharya, Ilya Shpitser, and James M. Robins. Causal and counterfactual views of missing data models. *Statistica Sinica*, 2025. doi: 10.5705/ss.202023.0382. URL `https://www3.stat.sinica.edu.tw/preprint/SS-2023-0382_Preprint.pdf`.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0393

REFERENCES

Judea Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, 2000. ISBN 0-521-77362-8.

Judea Pearl and J. M. Robins. Probabilistic evaluation of sequential plans from causal models with hidden variables. In *Uncertainty in Artificial Intelligence*, volume 11, pages 444–453, 1995.

Joseph Ramsey and Bryan Andrews. Py-Tetrad and RPy-Tetrad: A new Python interface with R support for Tetrad causal search. In *Causal Analysis Workshop Series*, pages 40–51. PMLR, 2023.

Thomas S. Richardson and James M. Robins. Single world intervention graphs (SWIGs): A unification of the counterfactual and graphical approaches to causality. *Center for the Statistics and the Social Sciences, University of Washington Series. Working Paper*, 2013.

James M. Robins. Non-response models for the analysis of non-monotone non-ignorable missing data. *Statistics in Medicine*, 16:21–37, 1997.

James M. Robins and R. D. Gill. Non-response models for the analysis of non-monotone ignorable missing data. *Statistics in Medicine*, 16(1-3):39–56, 1997.

James M. Robins, Andrea Rotnitzky, and Daniel O. Scharfstein. Sensitivity analysis for selection bias and unmeasured confounding in missing data and causal inference models. In M. E. Halloran and D. Berry, editors, *Statistical Models in Epidemiology: the Environment and Clinical Trials*, pages 1–92. NY: Springer-Verlag, 1999.

Ilya Shpitser and Tyler J. VanderWeele. A complete graphical criterion for the adjustment formula in mediation analysis. *International Journal of Biostatistics*, 2010.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0393

REFERENCES

Ilya Shpitser, Tyler VanderWeele, and James M. Robins. On the validity of covariate adjustment for estimating causal effects. In *Proceedings of the Twenty Sixth Conference on Uncertainty in Artificial Intelligence (UAI-10)*, pages 527–536. AUAI Press, 2010.

BaoLuo Sun, Lan Liu, Wang Miao, Kathleen Wirth, James Robins, and Eric J Tchetgen Tchetgen. Semiparametric estimation with data missing not at random using an instrumental variable. *Statistica Sinica*, 28(4):1965, 2018.

Johannes Textor, Benito van der Zander, Mark K. Gilthorpe, Maciej Liskiewicz, and George T.H. Ellison. Robust causal inference using directed acyclic graphs: The R package 'dagitty'. *International Journal of Epidemiology*, 45(6):1887–1894, 2016.

Santtu Tikka, Antti Hyttinen, and Juha Karvanen. Causal effect identification from multiple incomplete data sources: A general search-based approach. *Journal of Statistical Software*, 99:1–40, 2021.

Tyler J. VanderWeele and Ilya Shpitser. A new criterion for confounder selection. *Biometrics*, 67(4):1406–1413, 2011.

Y Samuel Wang, Mladen Kolar, and Mathias Drton. Confidence sets for causal orderings. *Journal of the American Statistical Association*, (just-accepted):1–25, 2025.

Zeyi Wang and Mark J. van der Laan. Discussion of "Causal and counterfactual views of missing data models" by Razieh Nabi, Rohit Bhattacharya, Ilya Shpitser, James M. Robins. *Statistica Sinica*, 2025. doi: 10.5705/ss.202025.0164. URL `https://www3.stat.sinica.edu.tw/preprint/SS-2025-0164_Preprint.pdf`.

Junhui Yang, Rohit Bhattacharya, Youjin Lee, and Ted Westling. Statistical and causal ro-

bustness for causal null hypothesis tests. In *Uncertainty in Artificial Intelligence*, pages 3956–3978. PMLR, 2024.

Shu Yang and Jae Kwang Kim. Discussion on "Causal and counterfactual views of missing data models". *Statistica Sinica*, 2025. doi: 10.5705/ss.202025.0165. URL `https://www3.stat.sinica.edu.tw/preprint/SS-2025-0165_Preprint.pdf`.

Yan Zhou, Roderick J. A. Little, and Kalbfleisch John D. Block-conditional missing at random models for missing data. *Statistical Science*, 25(4):517–532, 2010.