On Testability and Goodness of Fit Tests in Missing Data Models

Razieh Nabi, Ph.D.

Department of Biostatistics and Bioinformatics
Rollins School of Public Health
Emory University
razieh.nabi@emory.edu

Joint work with:
Rohit Bhattacharya (Department of Computer Science, Williams College)

Missing data and causal inference

- We encounter missing data for a variety of reasons.
  - Data collection, dropouts, censoring, death, etc.
- Causal inference and missing data are analogous in terminology, theory of identification, and statistical inference.
  - Causal inference has been viewed as a missing data problem, where responses to some (hypothetical) treatment interventions are not observed.
  - Missing data has also been viewed as a causal inference problem, where missingness indicators are treated as intervenable treatments.
- Similar to the use of graphical models in causal inference, we may represent missing data models graphically.
- The question is whether the encoded assumptions on a missing data graph are testable or not? If so, how?

Missing data notation

- $X = (X_1, \ldots, X_K)^T$: a vector of K random variables
- Given a finite sample from $p(X)$:
  - $R = (R_1, \ldots, R_K)^T$: binary missingness indicators

    $R_k = 1$ if X_k is observed, and $R_k = 0$ otherwise
  - $X^* = (X_1^*, \ldots, X_K^*)^T$: coarsened version of X

    $X_k^* = X_k$ if $R_k = 1$, and $X_k^* = ?$ otherwise
- Causal interpretation of the tuple (X_k, R_k, X_k^*):
  - R_k: a treatment variable that can be intervened on
  - X_k: a counterfactual – had we intervened and set $R_k = 1$
  - X_k^*: a factual variable
- For simplicity of notations, we assume all variables have missing values.

Missing data models

- A missing data model is a set of distributions defined over variables in $\{X, R, X^*\}$.

- By chain rule of probability, we can factorize $p(X, R, X^*)$ as follows:

$$\underbrace{\underbrace{p(X)}_{\text{target law}} \quad \times \quad \underbrace{p(R \mid X)}_{\text{missingness mechanism}}}_{\text{full law } p(X,R)} \quad \times \quad \underbrace{p(X^* \mid X, R)}_{\text{deterministic terms}}.$$

- Consistency assumption: $X_k^* = \begin{cases} X_k & \text{if } R_k = 1 \\ ? & \text{if } R_k = 0 \end{cases}$

- Observed data law is $p(R, X^*)$, where counterfactuals are marginalized out.

A missing data workflow

1. Define the **estimand** (often done in the absence of missing data).
   - A function of target law $p(X)$ or full law $p(X, R)$.
2. Assume a **model** that links the counterfactual, factual, and missingness indicator variables.
   - Often using **Directed Acyclic Graphs (DAGs)** to encode the modeling assumptions (Mohan et al., 2013; Bhattacharya et al., 2019).
3. Determine whether the estimand is **identifiable** in the assumed model.
   - Focus on **identification of the target and full laws** (Bhattacharya et al., 2019; Nabi et al., 2020, 2022).
4. If estimand is identifiable, find the best **estimation** strategy, and if it is not, perhaps stronger assumptions are needed (or alternatively obtaining bounds).
5. Conduct **sensitivity analysis** to reflect on the assumptions.

However . . .

- The validity of identification and estimation results using such techniques rely on the independence assumptions encoded by the graph/model holding true.

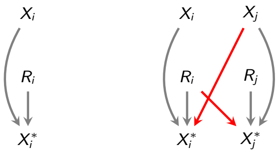
- In order to test an independence assumption in a missing data model, we have to examine its implications on the observed data distribution.
  - This enables the design of empirical testing procedures from finite (but partially unobserved) samples.
  - Unfortunately, we may not always be able to test all the encoded restrictions. But sometimes we can!

- The contributions of this work:
  - We expand on **testable implications** of missing data models that resemble ordinary conditional independencies in the underlying full law, but manifest as **generalized a.k.a. Verma independencies** in the observed law.
  - We design **empirical tests for restrictions** in three broad classes of missing data models that use ideas from weighted likelihood-ratio tests and odds-ratio parameterizations of joint distributions.

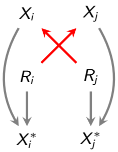
Introducing missing data DAGs

- Define missing data models via restrictions on the full data distribution that can be represented by a DAG (similar to causal inference).

- In missing data DAGs: (Mohan et al., 2013; Bhattacharya et al., 2019)
  1. Observed and counterfactual variables appear on the same graph
  2. There are certain edge restrictions: (marked in red)



deterministic edges



"no interference"

- The "no interference" assumption can be relaxed (Srinivasan et al., 2023).

Missing data DAG models

- Denote the missing data DAG (m-DAG) defined over $V = (X, R, X^*)$ via $\mathcal{G}(V)$.
- The statistical model of m-DAG $\mathcal{G}(V)$ is a set of distributions that factorize as:

$$
\begin{aligned}
p(X, R, X^*) &= \prod_{V_i \in V} p(V_i \mid \mathrm{pa}_{\mathcal{G}}(V_i)) \\
&= \prod_{V_i \ \in \ \{X,R\}} p(V_i \mid \mathrm{pa}_{\mathcal{G}}(V_i)) \times \prod_{X_i^* \ \in \ X^*} p(X_i^* \mid X_i, R_i).
\end{aligned}
$$

Familiar graphical concepts like d-separation and Markov properties carry over.

- **Factorization**: probability distribution as a set of small factors.
- **Local Markov property:** a small but complete set of indep constraints.

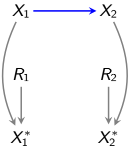
$$V_i \perp\!\!\!\perp \mathrm{nd}_{\mathcal{G}}(V_i) \setminus \mathrm{pa}_{\mathcal{G}}(V_i) \mid \mathrm{pa}_{\mathcal{G}}(V_i), \ \forall V_i \in V.$$

- **Global Markov property:** all independence constraints in the model.

$$\text{Given } X, Y, Z \in V : \quad (X \perp\!\!\!\perp_{\text{d-sep}} Y \mid Z)_{\mathcal{G}(V)} \implies (X \perp\!\!\!\perp Y \mid Z)_{p(V)}.$$

- All three properties are equivalent.

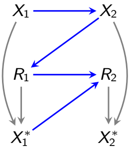
Examples: m-DAG models



$R_1 \perp\!\!\!\perp R_2, X_1, X_2$

$R_2 \perp\!\!\!\perp R_1, X_1, X_2$

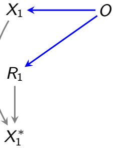
MCAR



$R_1 \perp\!\!\!\perp X_1 \mid X_2$

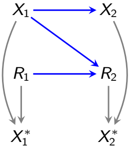
$R_2 \perp\!\!\!\perp X_1, X_2 \mid R_1, X_1^*$

Permutation (Robins, 1997)



$R_1 \perp\!\!\!\perp X_1 \mid O$

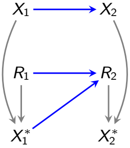
MAR



$R_1 \perp\!\!\!\perp X_1, X_2$

$R_2 \perp\!\!\!\perp X_1 \mid R_1, X_1$

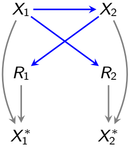
Block-conditional (Zhou et al., 2010)



$R_1 \perp\!\!\!\perp X_1, X_2$

$R_2 \perp\!\!\!\perp X_1, X_2 \mid R_1, X_1^*$

MAR



$R_1 \perp\!\!\!\perp R_2, X_1 \mid X_2$

$R_2 \perp\!\!\!\perp R_1, X_2 \mid X_1$

Block-parallel (Mohan et al., 2013)

Nonparametric saturation

- Similar to DAGs, absence of an edge in a m-DAG implies a restriction of the form $A \perp\!\!\!\perp B \mid C$. Is this restriction testable from observed finite samples?
- If all the restrictions encoded in a m-DAG are provably untestable (i.e., no restriction on the observed data law), the full law Markov relative to the m-DAG is said to be **non-parametric saturated** (NPS) (Robins, 1997).
  - Permutation model is NPS (has a DAG representation) (Robins, 1997).
  - No self-censoring model is NPS (has a chain graph representation) (Shpitser, 2016; Sadinle and Reiter, 2017; Malinsky et al., 2021).
- Submodels of a non-parametric saturated model can still be tested using partially observed data.

Example: Permutation model is NPS

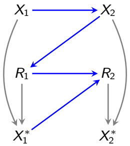
- Given an ordering on variables in X, indexed by $k \in \{1, \ldots, K\}$, each missingness indicator R_k is independent of the current and past variables in X given the past observed variables in R, X^* and future variables in X, i.e.,

$$R_k \perp\!\!\!\perp X_{\prec k+1} \mid R_{\prec k}, X^*_{\prec k}, X_{\succ k}, \ \forall k \quad \textit{(permutation model)}$$

  where $V_{\prec k} = \{V_1, \ldots, V_{k-1}\}, V_{\succ k} = \{V_{k+1}, \ldots, V_K\}$ (Robins, 1997).

- The *permutation model* is an example of a non-parametric saturated model. (This claim can be proved by deriving *tangent space* of the observed data law, and realizing it includes all observed data distributions without any restrictions.)

- Example of m-DAG representation for the permutation model with two variables:



$$R_1 \perp\!\!\!\perp X_1 \mid X_2, \qquad R_2 \perp\!\!\!\perp X_1, X_2 \mid R_1, X_1^*$$

- Submodels of the permutation model can still be tested: Introducing *sequential MAR* and *sequential MNAR* as submodels.

Sequential MAR model

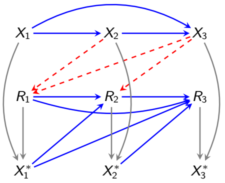
- We call a missing data model a *sequential MAR* model if under an ordering $\prec$ that indexes variables by $k = 1, \ldots, K$, the following restrictions hold:

$$R_k \perp\!\!\!\perp X \mid R_{\prec k}, X^*_{\prec k}, \ \forall k \qquad \textit{(sequential-MAR)}$$

- In addition to the assumptions in the permutation model, the sequential MAR model assumes the following independence restrictions:

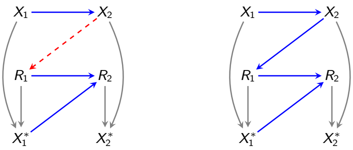
$$R_k \perp\!\!\!\perp X_{\succ k} \mid R_{\prec k}, X^*_{\prec k}$$

- Example of a sequential MAR model (without the dashed edges) along with its permutation supermodel (with the dashed edges) with three substantive variables:



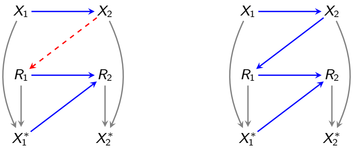
Prior developments on testable implications do not suffice

- Example of a MAR model (left) as a submodel of permutation model (right)



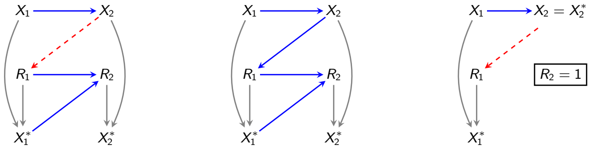
- We want to test the absence of $X_2 \to R_1$ edge, i.e., whether $R_1 \perp\!\!\!\perp X_2$?
- Let's apply the criterion proposed by Mohan and Pearl (2014):
  - A d-separation condition displayed in a m-DAG is testable if the missingness indicators associated with all partially observed variables involved in the relation are either already present in the separating set, or can be added to the set without spoiling the separation.
- $R_1 \perp\!\!\!\perp_{\text{d-sep}} X_2 \mid R_2$ does *not* hold due to the open collider R_2 on the path $R_1 \to R_2 \leftarrow X_1^* \leftarrow X_1 \to X_2$.
- Therefore, one might conclude that $R_1 \perp\!\!\!\perp X_2$ is not testable.

Testable implications using pattern-mixture parameterization



- Assume momentarily that X consists of binary variables.
- Let us compare number of parameters in the full law using m-DAG factorization against the saturated observed data law using pattern-mixture factorization (Rubin, 1976); $p(R, X^*) = p(R) \times p(X^* \mid R)$.
    - The full law $p(R, X)$ requires 7 parameters: 3 for $p(X)$, 1 for $p(R_1)$, and 3 for $p(R_2|R_1, X_1^*)$.
    - Saturated observed law requires 8 parameters: 3 for $p(R)$ and 5 for $p(X^*|R)$.
- Thus, $R_1 \perp\!\!\!\perp X_2$ must impose constraints on the observed data law, at least in the discrete case. This contradicts the earlier conclusions.
    - Contradictions are expected as the criterion of Mohan and Pearl is sufficient but not necessary for testability.

Testing the sequential MAR model via Verma constraints

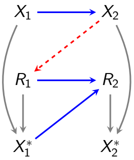


(a) (b) (c)

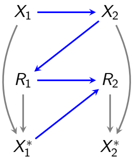
- The issue with testing $R_1 \perp\!\!\!\perp X_2$ was the collider at R_2, $R_1 \rightarrow R_2 \leftarrow X_1^*$
- From a causal perspective, removal of these edges corresponds to an intervention on R_2, resulting in the m-conditional DAG in (c).
- Following Pearl's do-calculus notation, the corresponding intervention distribution is denoted by $p(X, R_1, X^* \mid \text{do}(R_2 = 1))$, or $p(. \mid \text{do}(R_2 = 1))$.
- This intervention distribution is obtainable via truncation of the full law factorization by dropping the propensity score of R_2, $p(R_2 \mid \text{pa}_{\mathcal{G}}(R_2))$, i.e.,

$$p(. \mid \text{do}(R_2 = 1)) = \left. \frac{p(X, R, X^*)}{p(R_2 \mid R_1, X_1^*)} \right|_{R_2=1}.$$

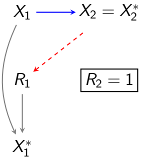
Empirical tests via weighted likelihood ratio tests



MAR model



Permutation supermodel

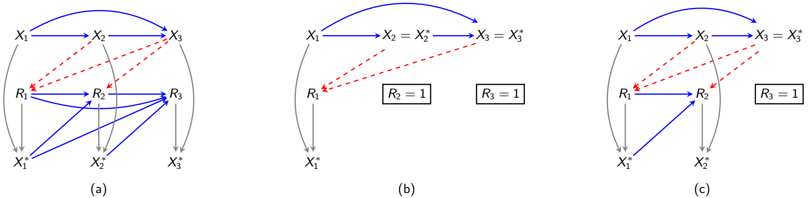


Intervention on R_2

- Is $R_1 \perp\!\!\!\perp X_2$?
- Fit models for $p(R_1)$ and $p(R_1 \mid X_2)$ and compare the goodness of fits.
- How to estimate parameters in $p(R_1 \mid X_2; \alpha_{r_1})$?
  - Assume $\mathbb{P}_n[U(.; \alpha_{r_1})] = 0$ with respect to the full law.
  - Use a weighted estimating equation to estimate α_{r_1} in $p(R_1 \mid X_2; \alpha_{r_1})$

$$\mathbb{P}_n\left[\frac{R_2}{p(R_2 \mid R_1, X_1^*; \widehat{\alpha}_{r_2})} \times U(.; \alpha_{r_1})\right] = 0,$$

Sequential MAR model with $K = 3$



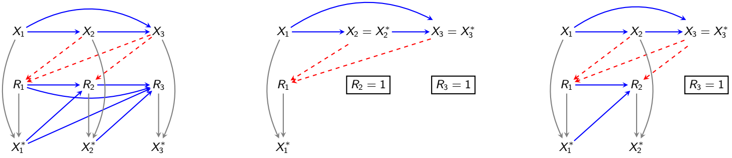
- **Null:** $\mathcal{M}_o$ is the statistical model of the m-DAG in (a) without dashed edges
- **Alternative:** $\mathcal{M}_a$ is the permutation supermodel in (a) with the dashed edges
- **Objective:** testing absence of the dashed edges which imply

$$R_1 \perp\!\!\!\perp X_2, X_3 \qquad \text{and} \qquad R_2 \perp\!\!\!\perp X_3 \mid R_1, X_1^*.$$

- **Solution:** translate these into independence restrictions in

$$p(. \mid \mathrm{do}(R_2 = 1, R_3 = 1)) \qquad \text{and} \qquad p(. \mid \mathrm{do}(R_3 = 1)).$$

Sequential MAR model with $K = 3$ ctd.



- E.g., testing $R_1 \perp\!\!\!\perp X_2, X_3$ entails fitting the parameters in $p(R_1 \mid X_2, X_3; \beta_{r_1})$ using:

$$\mathbb{P}_n \left[\frac{R_2 \times R_3}{p(R_2 \mid \text{pa}_{\mathcal{G}}(R_2); \widehat{\beta_{r_2}}) \times p(R_3 \mid \text{pa}_{\mathcal{G}}(R_3); \widehat{\beta_{r_3}})} \times U(\beta_{r_1}) \right] = 0,$$

- Estimating β_{r_3} is straightforward since $p(R_3 \mid \text{pa}_{\mathcal{G}}(R_3)) = p(R_3 \mid R_1, R_2, X_1^*, X_2^*)$ is a direct function of observed data.
- Estimating β_{r_2} is more involved since $p(R_2 \mid \text{pa}_{\mathcal{G}}(R_2)) = p(R_2 \mid R_1, X_1^*, X_3)$ is not a direct function of observed data.
  - We can estimate β_{r_2} using an intervention distribution where R_3 is intervened on, i.e., $p(X, R, X^*)/p(R_3 \mid \text{pa}_{\mathcal{G}}(R_3))$ evaluated at $R_3 = 1$. This means using $p(R_3 \mid \text{pa}_{\mathcal{G}}(R_3); \widehat{\beta_{r_3}})$ as inverse weights to estimate β_{r_2}.

Generalization to $K > 3$

- In the general case with K variables, it is better to carry out tests backwards by first testing restrictions involving R_{K-1}, moving to R_{K-2}, and so on.
- If the current test succeeds, the corresponding model for the null can be re-used to produce weights for future estimating equations; if the test fails, then the assumptions of sequential MAR does not hold.
- As we proceed with the tests, we are restricted to fewer and fewer samples which impacts the power of our tests.
- A future direction is developing semiparametric methods to use data more efficiently.
- The general theory and the corresponding algorithm are described in the paper.

Algorithm 1 TESTING SEQUENTIAL MAR $(\prec, \mathcal{M}, \mathcal{D}_n)$

Let $\prec$ index variables by $k = 1, \ldots, K$.
Let $W_K(\beta_K^o) := p(R_K | R_{\prec K}, X^*_{\prec K}; \beta_K^o)$.
Estimate β_K^o (denote it by $\widehat{\beta}_K^o$).
for $k \in \{K-1, \ldots, 1\}$ **do**
    Let $W_k(\beta_k^o) := p(R_k | R_{\prec k}, X^*_{\prec k}; \beta_k^o)$ and $W_k(\beta_k^a) := p(R_k | R_{\prec k}, X^*_{\prec k}, X_{\succ k}; \beta_k^a)$.
    Estimate β_k^o (denote it by $\widehat{\beta}_k^o$).
    Estimate β_k^a via the weighted estimating equation:

$$\mathbb{P}_n\left[\frac{\mathbb{I}(R_{\succ k} = 1)}{\prod_{j \succ k}^{K} W_j(\widehat{\beta}_j^o)} \times U(\beta_k^a)\right] = 0,$$

    where $\mathbb{P}_n\left[U(\beta_k^a)\right] = 0$ is an unbiased estimating equation for β_k^a wrt the full law (denote it by $\widehat{\beta}_k^a$).
    Compute a weighted likelihood-ratio as follows:

$$\rho = n\mathbb{P}_n\left[\frac{\mathbb{I}(R_{\succ k} = 1)}{\prod_{j \succ k}^{K} W_j(\widehat{\beta}_j^o)} \times \log\left(\frac{W_k(\widehat{\beta}_k^a)}{W_k(\widehat{\beta}_k^o)}\right)\right].$$

    Test ρ with α significance level.
    if $\mathcal{M}_o$ is rejected (i.e., $R_k \not\perp\!\!\!\perp X_{\succ k} | R_{\prec k}, X^*_{\prec k}$) **then**
        return not sequential MAR
return sequential MAR

Sequential MNAR model

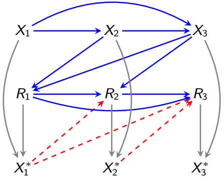
- We call a missing data model a *sequential MNAR* model if under an ordering $\prec$ that indexes variables by $k = 1, \ldots, K$, the following restrictions hold:

$$R_k \perp\!\!\!\perp X_{\prec k+1}, X^*_{\prec k} \mid R_{\prec k}, X_{\succ k}, \ \forall k \qquad \textit{(sequential-MNAR)}$$

- In addition to the assumptions in the permutation model, the sequential MNAR model assumes the following independence restrictions:

$$R_k \perp\!\!\!\perp X^*_{\prec k} \mid R_{\prec k}, X_{\succ k}, \ \forall k$$

- Example of a sequential MNAR model (without the dashed edges) along with its permutation supermodel (with the dashed edges) with three substantive variables:



- The additional restrictions can be tested using similar ideas as in the sequential MAR model (discussed in detail in the paper.)

Example: No self-censoring model is NPS

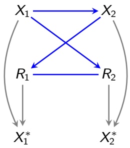
- No variable directly causes its own missingness status, i.e.,

$$R_k \perp\!\!\!\perp X_k \mid R_{-k}, X_{-k}, \ \forall k \qquad \textit{(no self-censoring)}$$

where $V_{-k} = V \setminus V_k$ (Shpitser, 2016; Sadinle and Reiter, 2017).

- The *no self-censoring model* is an example of a non-parametric saturated model. (see Malinsky et al. (2021) for more details.)

- Example of m-DAG representation for the no self-censoring model with two substantive variables:



$$R_1 \perp\!\!\!\perp X_1 \mid R_2, X_2, \qquad R_2 \perp\!\!\!\perp X_2 \mid R_1, X_1$$

- Submodels of the no self-censoring model can still be tested: Introducing *block-parallel MNAR* as submodels.

Block-parallel MNAR model

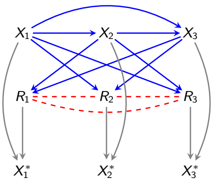
- We call a missing data model a *block-parallel MNAR* model if the following restrictions hold:

$$R_k \perp\!\!\!\perp R_{-k}, X_k \mid X_{-k}, \forall k \quad \textit{(block-parallel MNAR)}$$

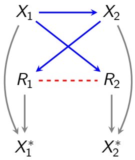
- In addition to the assumptions in the no self-censoring model, the block-parallel MNAR model assumes the following independence restrictions:

$$R_k \perp\!\!\!\perp R_j \mid X, \forall j \neq k.$$

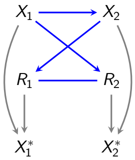
- Example of a block-parallel MNAR model (without the dashed edges) along with its no self-censoring supermodel (with the dashed edges) with three substantive variables:



Testing the block-parallel model via odds ratio parameterization



MNAR model



No self-censoring supermodel

- Is $R_1 \perp\!\!\!\perp R_2 \mid X$?
- This translates into whether or not $\mathrm{OR}(R_1 = 0, R_2 = 0 \mid X) = 1$?
- The odds ratio parameter can be estimated using the estimating equation $\mathbb{P}_n\big[U(.;\mathrm{OR})\big] = 0$, where

$$U(.;\mathrm{OR}) = \frac{R_1 \times R_2}{p(R = 1 \mid X)} \times p(R = 0 \mid X) - (1 - R_1)(1 - R_2), \text{ and}$$

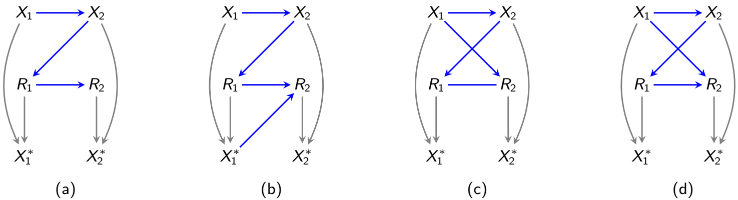
$$p(R = 1 \mid X) = p(R_1 = 1 \mid R_2 = 1, X_2) \times p(R_2 = 1 \mid R_1 = 1, X_1),$$
$$p(R = 0 \mid X) = p(R_1 = 0 \mid R_2 = 1, X_2) \times p(R_2 = 0 \mid R_1 = 1, X_1) \times \mathrm{OR}(R_1 = R_2 = 0 \mid X).$$

- The generalization to $K > 2$ is discussed in the paper.

The role of supermodels in designing empirical tests

The (sequential) MNAR model (a) was treated as a submodel of the permutation model (b), but it can also be treated as a submodel of:

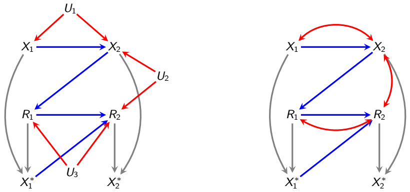
- The saturated no self-censoring model (c), or
- The so-called criss-cross model (d).



- However, the test statistic is not identifiable under the criss-cross supermodel.
- Neither the target law nor the full law is identified in (d); counterexamples are provided in the paper.

Extensions to settings with unmeasured confounders

- What if there exist variables that are not just missing but completely unobserved?
- Summarize the observed data distribution with a missing data acyclic directed mixed graph (ADMG).



(a) $\mathcal{G}(V, U)$ (b) $\mathcal{G}(V)$

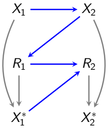
X_1: smoking, X_2: lung cancer

U_1: genotypic traits, U_2: occupation, U_3: ethnicity

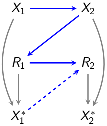
- Generalization of the results to m-ADMGs is simple as long as the full law remains identified.
- There exist *sound* and *complete* identification results for full laws representable via m-DAGs or m-ADMGs (Nabi et al., 2020).

Summary

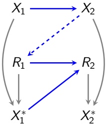
- Designing empirical tests for restrictions in three broad classes of missing data models via weighted likelihood-ratio tests and odds-ratio parameterizations.



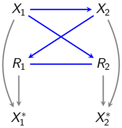
Permuation-model
(Robins; 1997)



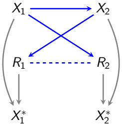
MNAR submodel



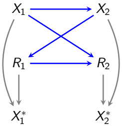
MAR submodel



No self-censoring
(Malinsky et al., 2021)



MNAR submodel



Criss-cross model
Neither full nor target law is ID.
(Nabi and Bhattacharya, 2023)
(Guo et al., 2023)

- Developing estimation methods that would complement our proposals by allowing a more efficient use of data in performing goodness-of-fit tests.
- Designing data-driven structure learning approaches.

References I

R. Bhattacharya, R. Nabi, I. Shpitser, and J. 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.

A. Guo, J. Zhao, and R. Nabi. Sufficient identification conditions and semiparametric estimation under missing not at random mechanisms. 2023.

D. Malinsky, I. Shpitser, and E. J. Tchetgen Tchetgen. Semiparametric inference for nonmonotone missing-not-at-random data: the no self-censoring model. *Journal of the American Statistical Association*, pages 1–9, 2021.

K. Mohan and J. Pearl. Graphical models for recovering probabilistic and causal queries from missing data. In *Advances in Neural Information Processing Systems 27*, pages 1520–1528. Curran Associates, Inc., 2014.

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

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

References II

R. Nabi, R. Bhattacharya, and I. 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.

R. Nabi, R. Bhattacharya, I. Shpitser, and J. Robins. Causal and counterfactual views of missing data models. *arXiv preprint arXiv:2210.05558*, 2022.

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

D. B. Rubin. Causal inference and missing data (with discussion). *Biometrika*, 63: 581–592, 1976.

M. Sadinle and J. P. Reiter. Itemwise conditionally independent nonresponse modeling for incomplete multivariate data. *Biometrika*, 104(1):207–220, 2017.

I. Shpitser. Consistent estimation of functions of data missing non-monotonically and not at random. *Advances in Neural Information Processing Systems*, 29:3144–3152, 2016.

R. Srinivasan, R. Bhattacharya, R. Nabi, E. L. Ogburn, and I. Shpitser. Graphical models of entangled missingness. *arXiv preprint arXiv:2304.01953*, 2023.

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

Sincerely,

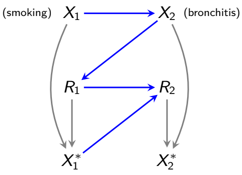
Razieh Nabi, Ph.D.
Rollins Assistant Professor
Department of Biostatistics and Bioinformatics
Emory University
@raziehnabi
razieh.nabi@emory.edu
https://raziehnabi.com

Appendix: Permutation model

How do models differ in telling a story about the missingness mechanisms?

- X_1: true smoking status of an individual.
- X_2: diagnosis of bronchitis.
- R_1, R_2: encode whether these variables have been measured or not.

- $X_2 \rightarrow R_1$
  A doctor inquires about the patient's smoking status on a *suspected* diagnosis of bronchitis before administering the test.
- $R_1 \rightarrow R_2 \leftarrow X_1^*$
  Whether the true bronchitis status is measured via a diagnostic test depends on the doctor's awareness of the individual's smoking status (R_1) and their observed value of smoking (X_1^*).



MNAR: permutation

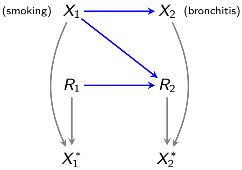
(Robins, 1997)

Appendix: Block-conditional model

How do models differ in telling a story about the missingness mechanisms?

- X_1: true smoking status of an individual.
- X_2: diagnosis of bronchitis.
- R_1, R_2: encode whether these variables have been measured or not.

- R_1 has no parent
  Inquiry into smoking status is random (e.g., as in random screening programs or surveys).
- $R_1 \ \to R_2 \ \leftarrow \ X_1$
  Administration of a diagnostic test depends on the inquiry into smoking, as well as the potentially unobserved past history of smoking.



MNAR: block-conditional

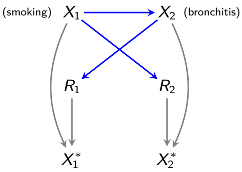
(Zhou et al., 2010)

Appendix: Block-parallel model

How do models differ in telling a story about the missingness mechanisms?

- X_1: true smoking status of an individual.
- X_2: diagnosis of bronchitis.
- R_1, R_2: encode whether these variables have been measured or not.

- $R_1 \leftarrow X_2$
  Inquiry into smoking status depends on a suspected diagnosis of bronchitis.
- $R_2 \leftarrow X_1$
  Administration of the diagnostic test depends on the suspected smoking status of an individual.



MNAR: block-parallel

(Mohan et al., 2013)

Appendix: Simulations

- Results on testing sequential MAR models.
- (top row) The sequential MAR model captures the true underlying missingness mechanism.
- (bottom row) The assumptions of sequential MAR model are violated.