Causal effect estimation in the presence of misclassified binary mediators

Kimberly A. H. Webb and Martin T. Wells

Women in Statistics and Data Science

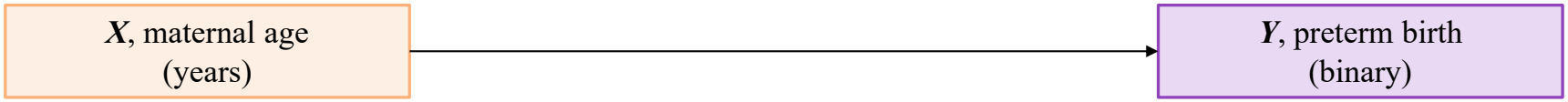
October 18, 2024

Mediation analysis

- **Mediation analysis** quantifies the effect of an **exposure (X)** on an **outcome (Y)**, mediated by some **intermediate (M)**.

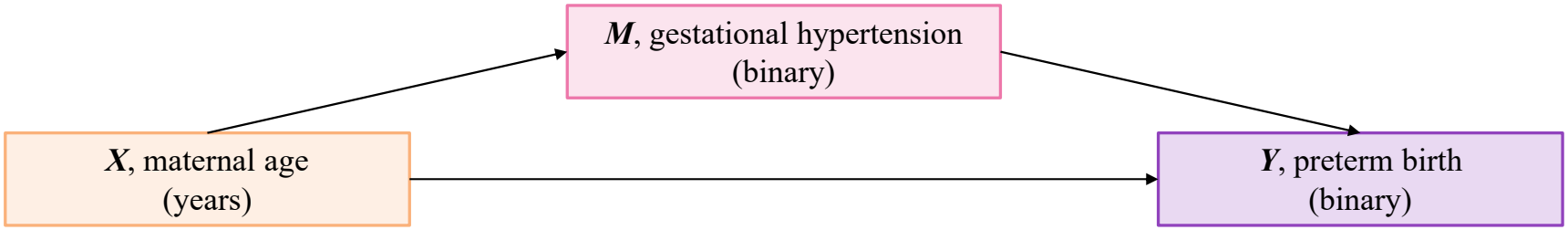
Mediation analysis

- **Mediation analysis** quantifies the effect of an **exposure (X)** on an **outcome (Y)**, mediated by some **intermediate (M)**.



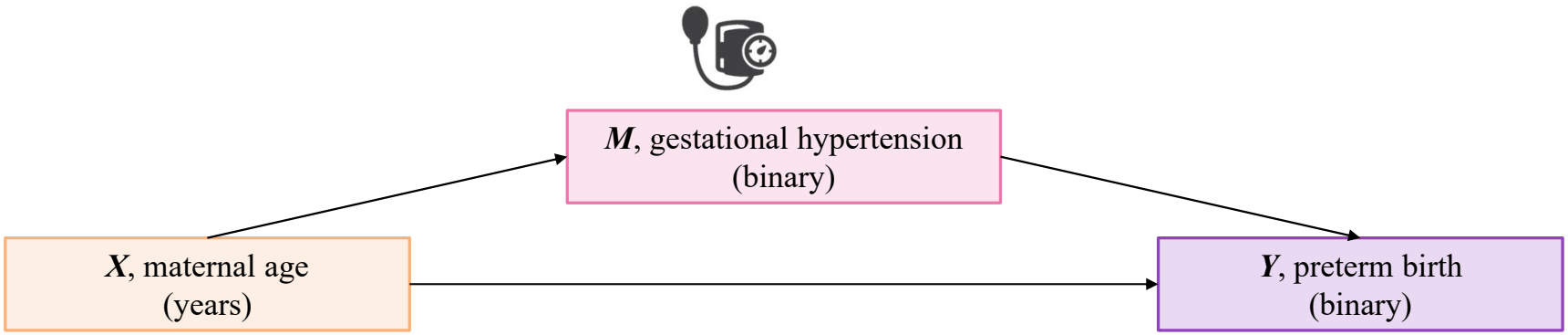
Mediation analysis

- **Mediation analysis** quantifies the effect of an **exposure (X)** on an **outcome (Y)**, mediated by some **intermediate (M)**.



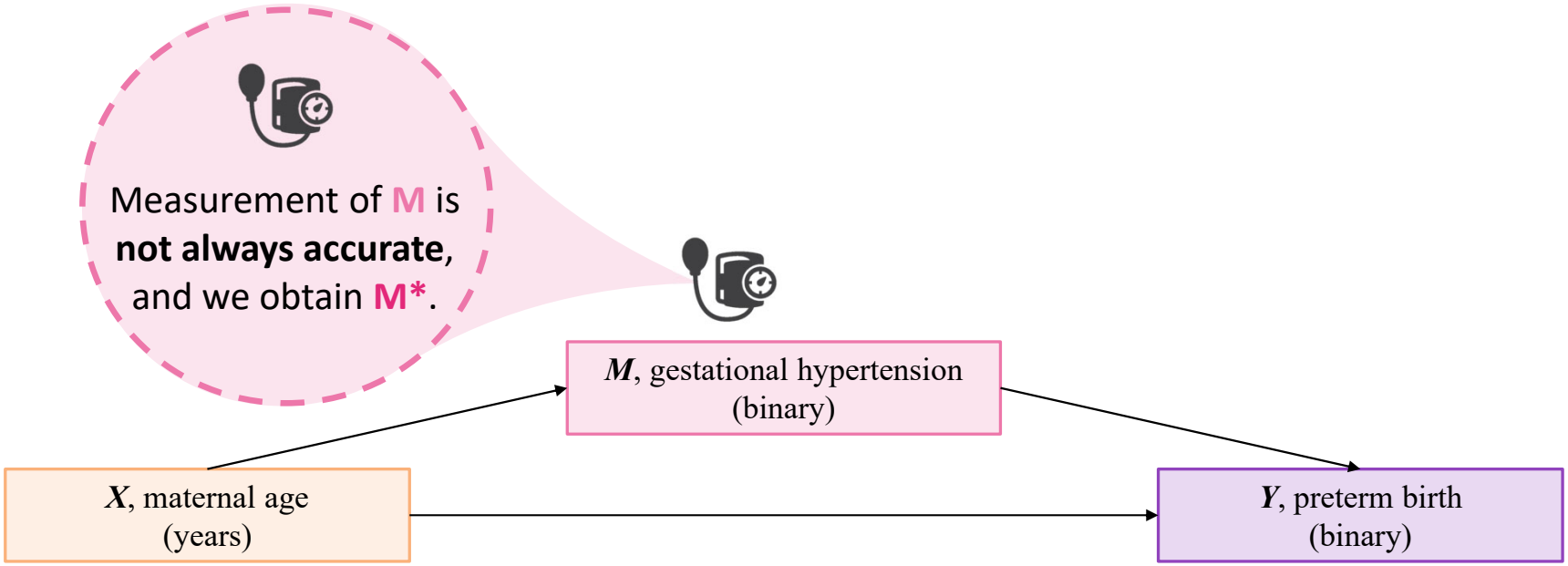
Mediation analysis

- **Mediation analysis** quantifies the effect of an **exposure (X)** on an **outcome (Y)**, mediated by some **intermediate (M)**.



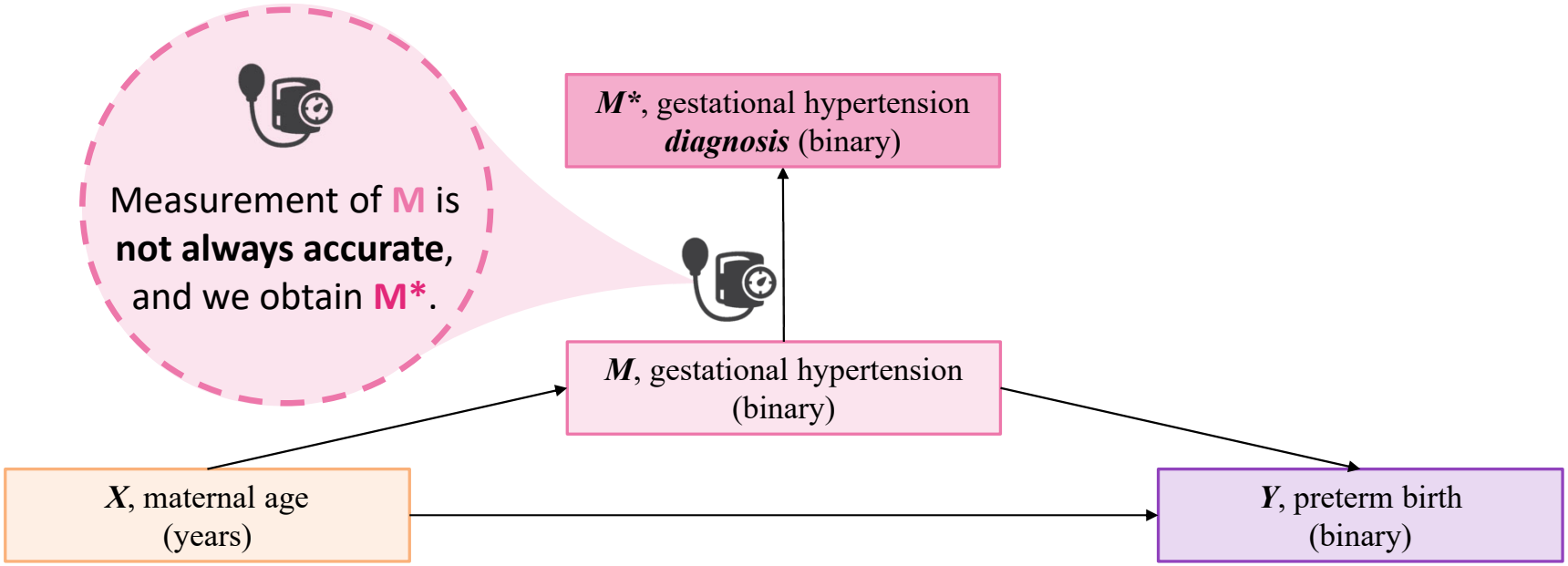
Mediation analysis

- **Mediation analysis** quantifies the effect of an **exposure (X)** on an **outcome (Y)**, mediated by some **intermediate (M)**.



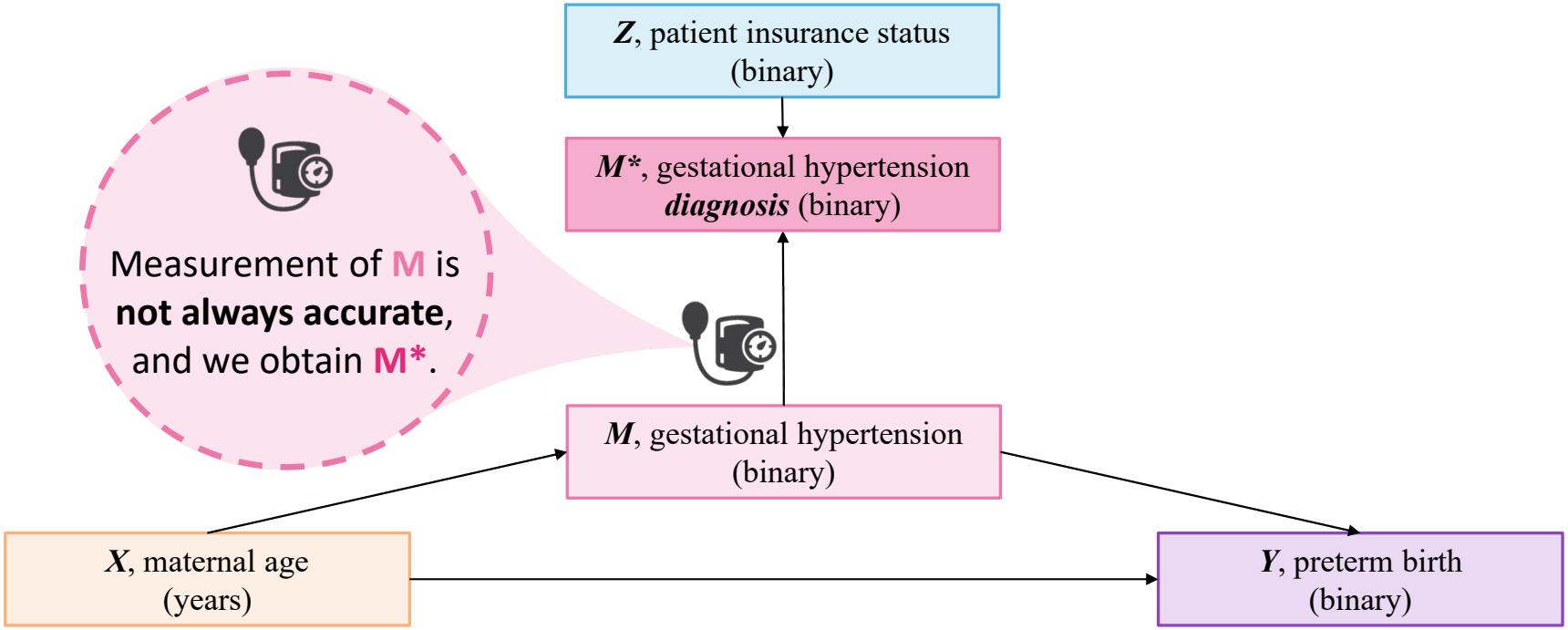
Mediation analysis

- **Mediation analysis** quantifies the effect of an **exposure (X)** on an **outcome (Y)**, mediated by some **intermediate (M)**.



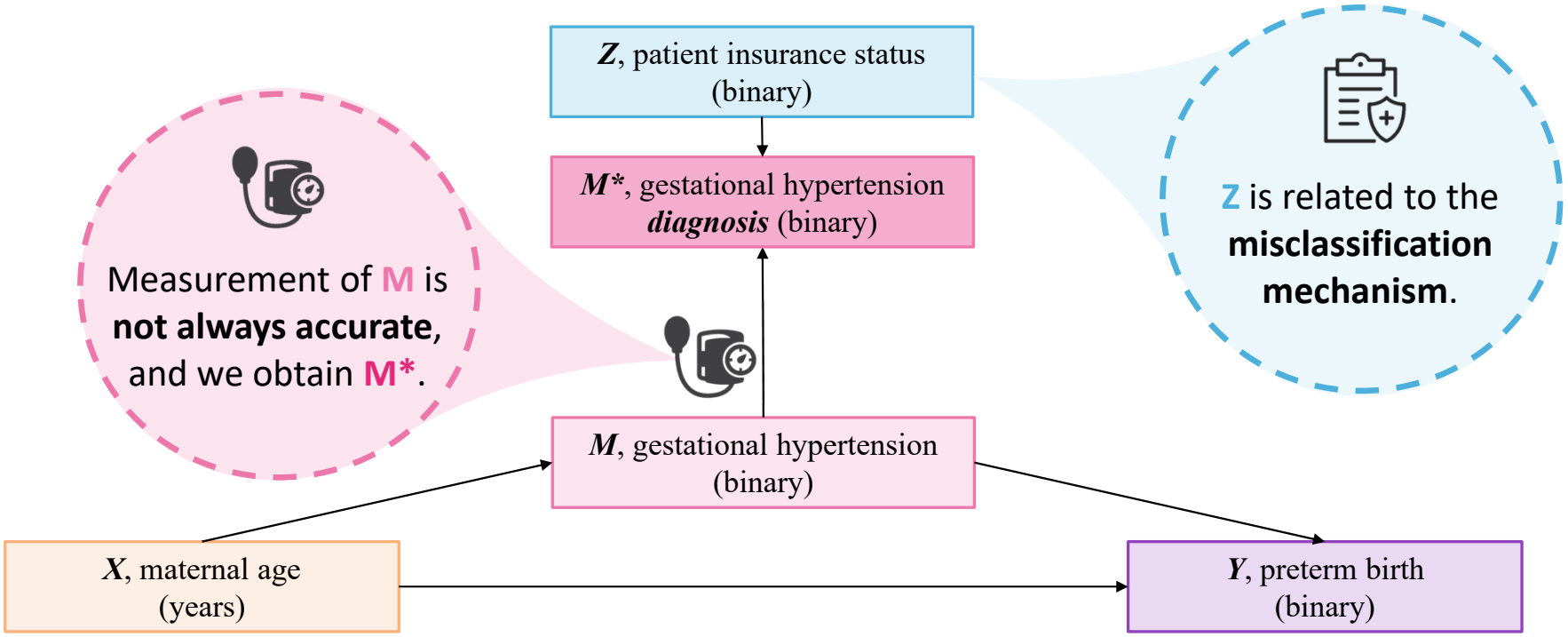
Mediation analysis

- **Mediation analysis** quantifies the effect of an **exposure (X)** on an **outcome (Y)**, mediated by some **intermediate (M)**.

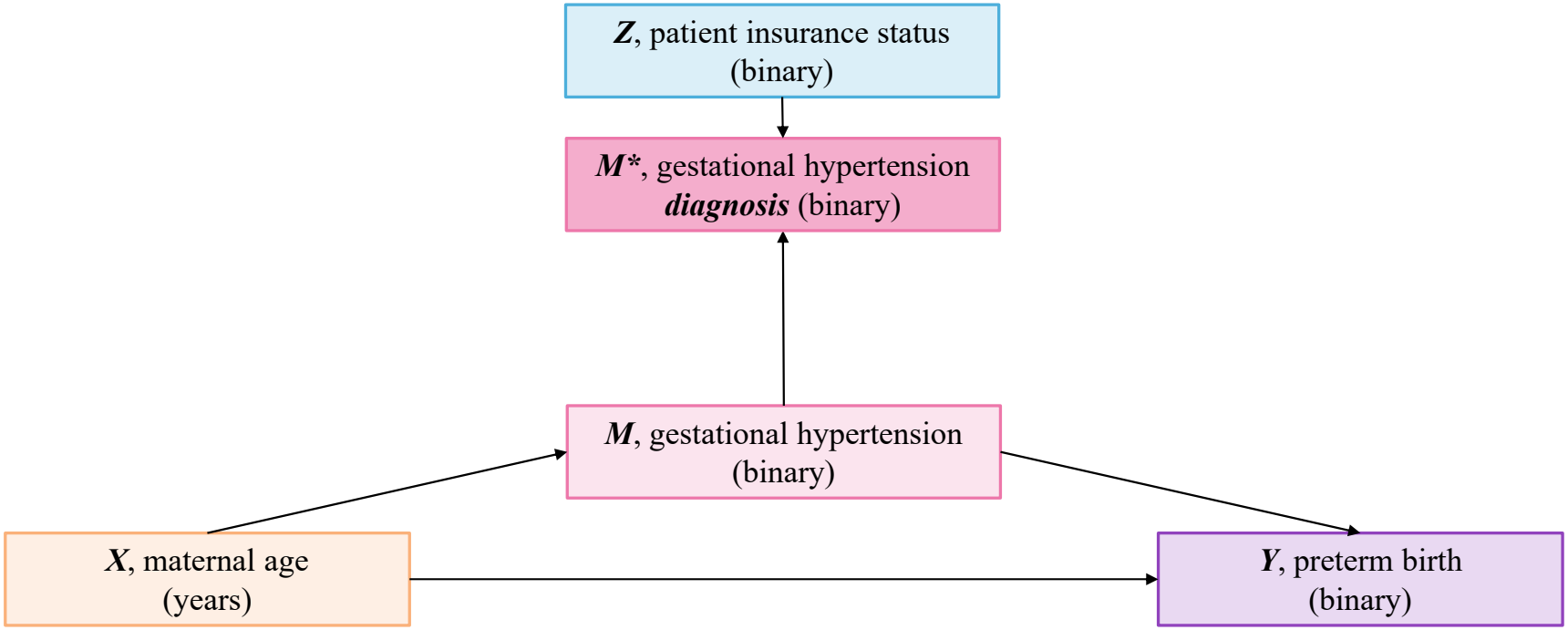


Mediation analysis

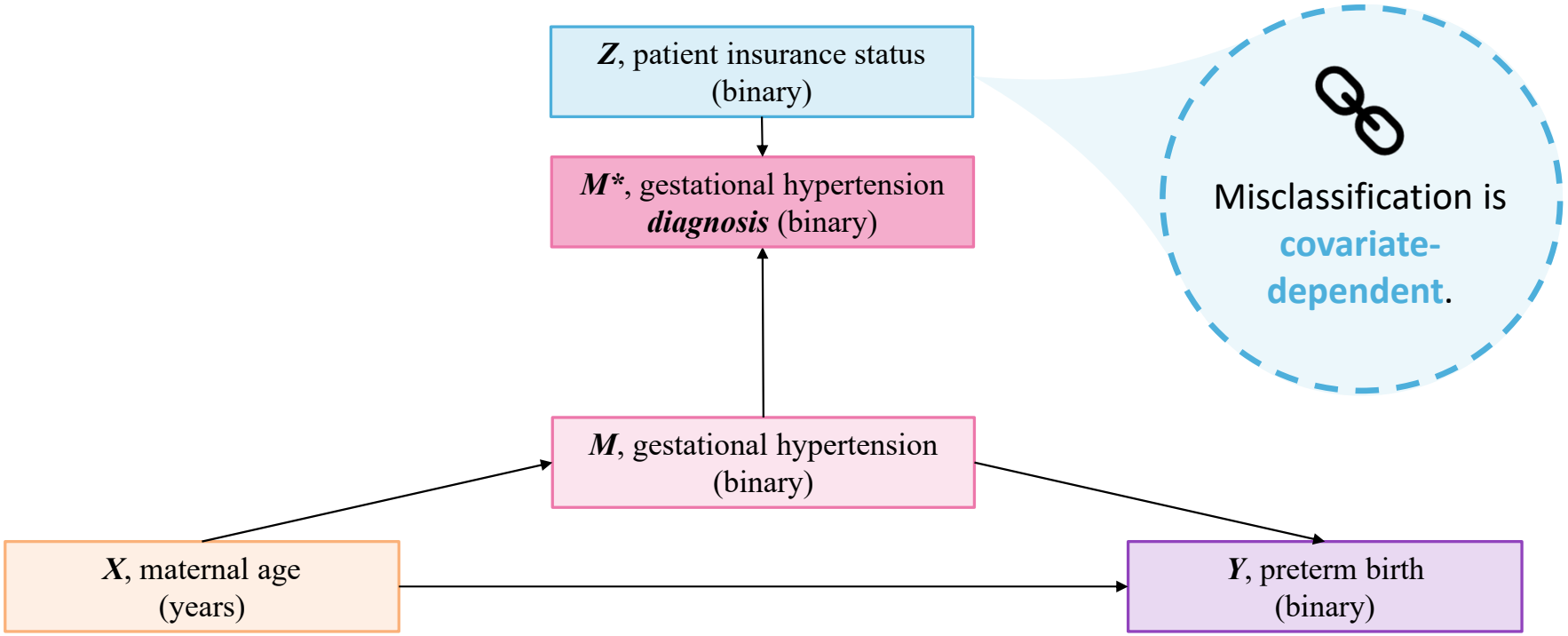
- **Mediation analysis** quantifies the effect of an **exposure (X)** on an **outcome (Y)**, mediated by some **intermediate (M)**.



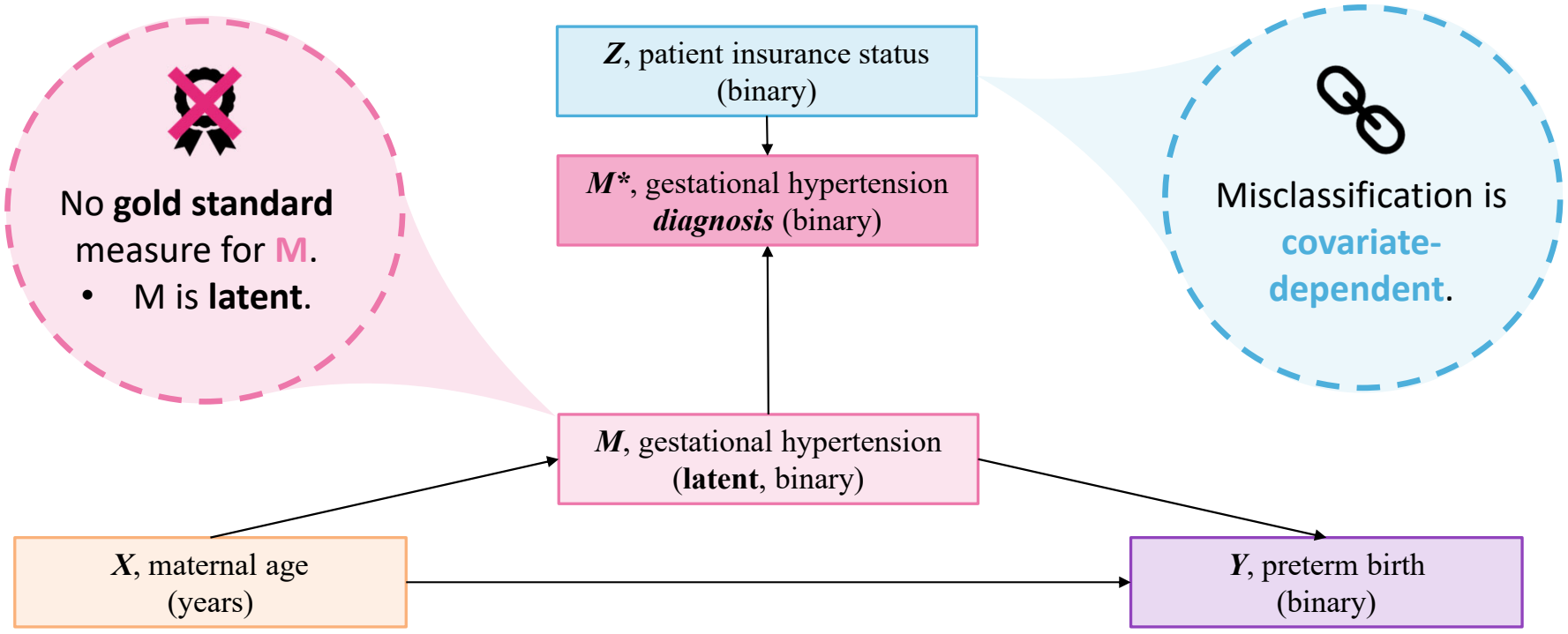
Challenges


Z, patient insurance status (binary)
M*, gestational hypertension ***diagnosis*** (binary)
M, gestational hypertension (binary)
X, maternal age (years)
Y, preterm birth (binary)

Challenges

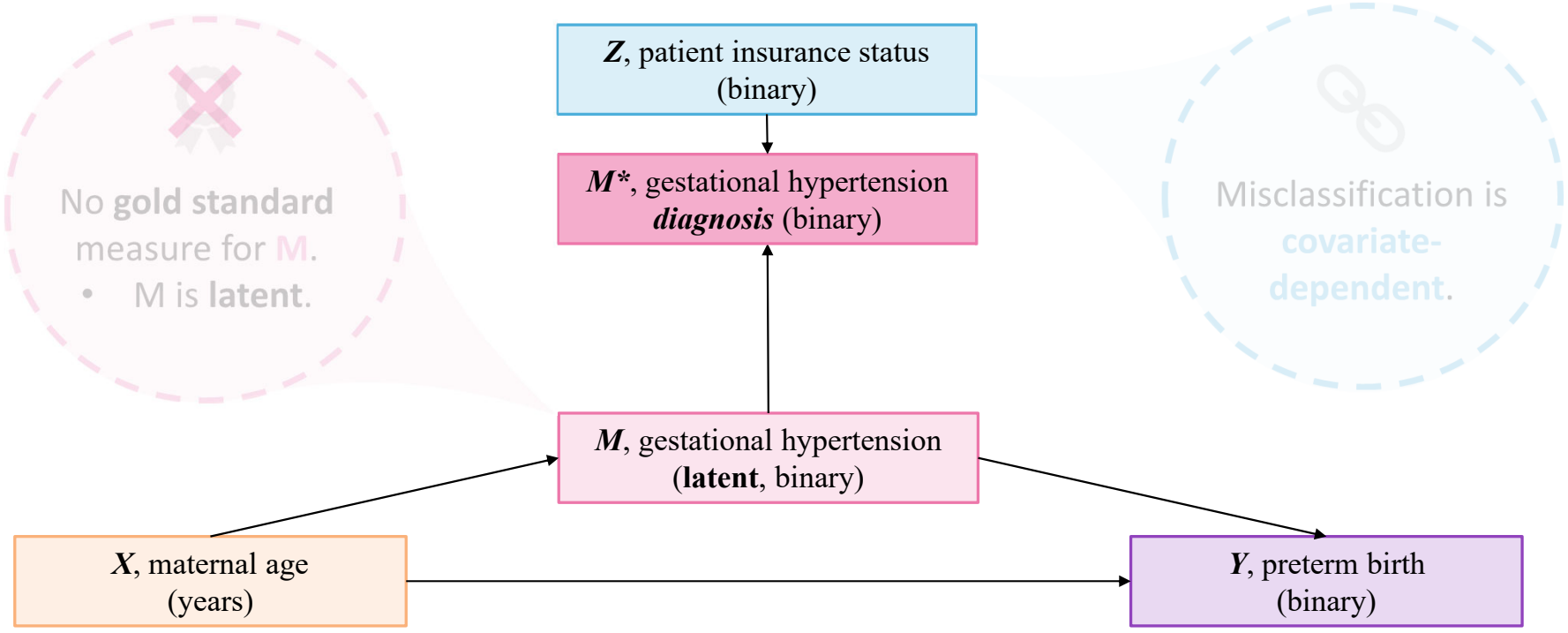

Z, patient insurance status (binary)
M*, gestational hypertension ***diagnosis*** (binary)
Misclassification is **covariate-dependent**.
M, gestational hypertension (binary)
X, maternal age (years)
Y, preterm birth (binary)

Challenges


No **gold standard** measure for M.
• M is **latent**.
Z, patient insurance status (binary)
M*, gestational hypertension ***diagnosis*** (binary)
Misclassification is **covariate-dependent**.
M, gestational hypertension (**latent**, binary)
X, maternal age (years)
Y, preterm birth (binary)

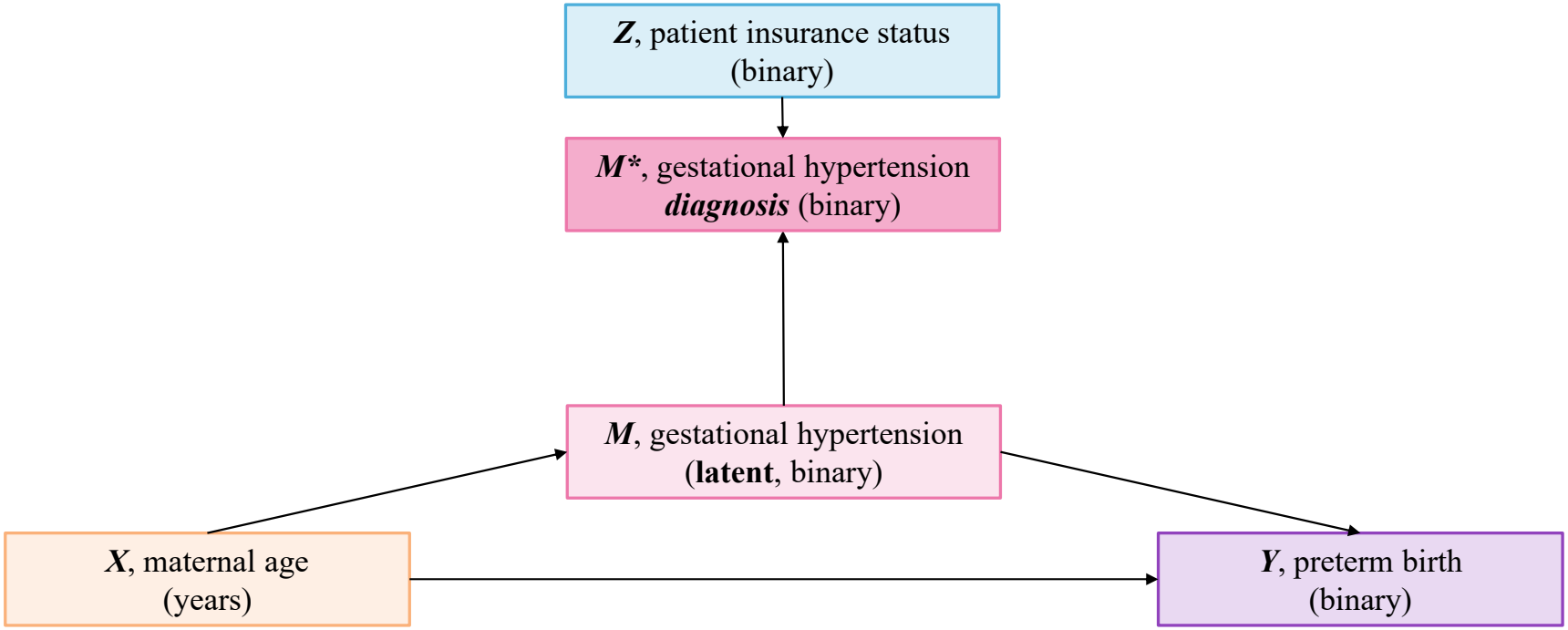
Challenges

Ignoring misclassification in M* → Bias in parameter and effect estimates

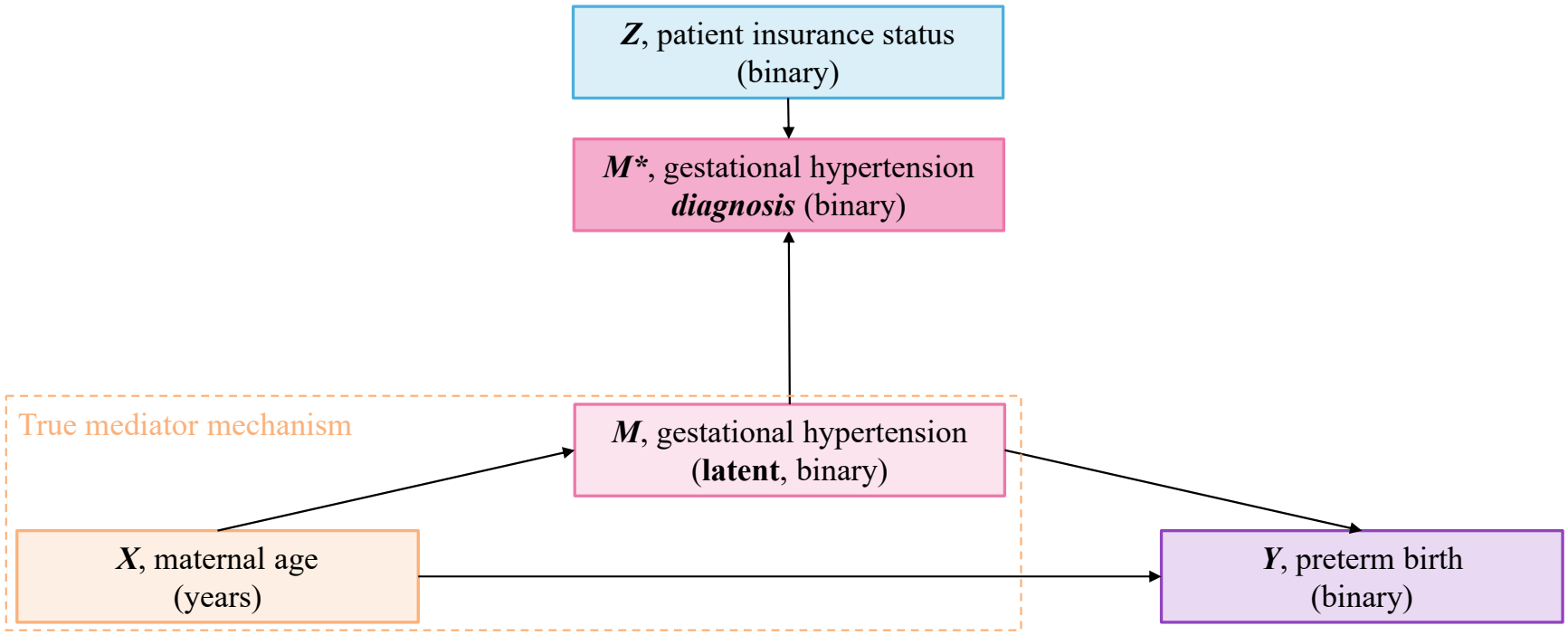


Analysis Plan

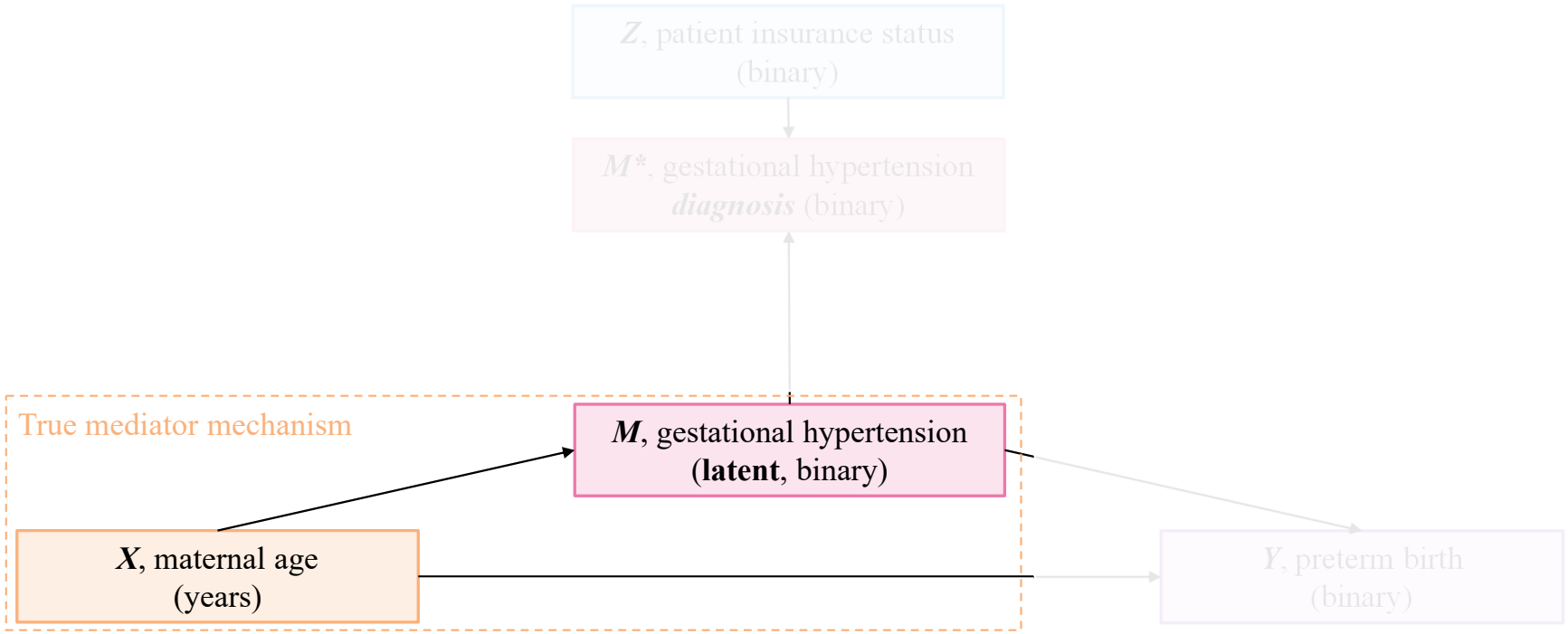
Misclassification model



Misclassification model

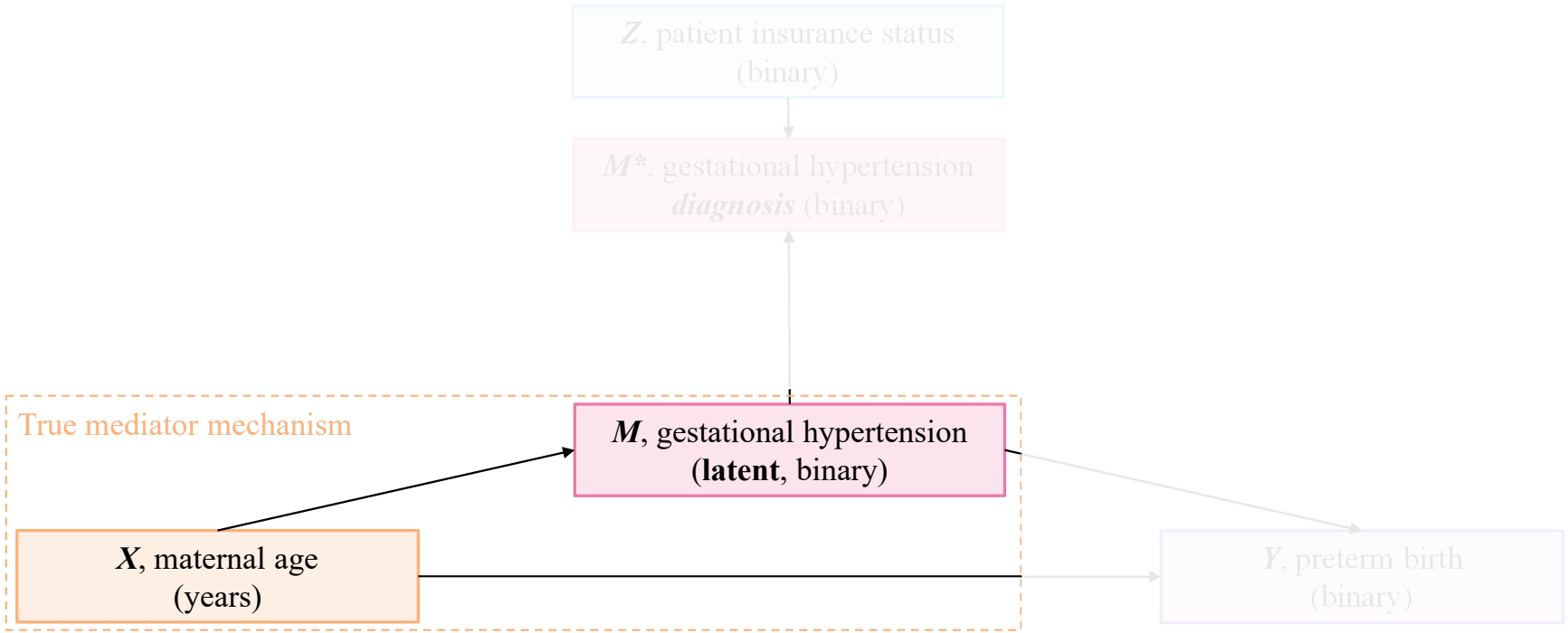


Misclassification model



True mediator mechanism: $\text{logit}\{P(M = 1|X, \boldsymbol{C}; \boldsymbol{\beta})\} = \beta_0 + \beta_X X + \boldsymbol{\beta_C C}$

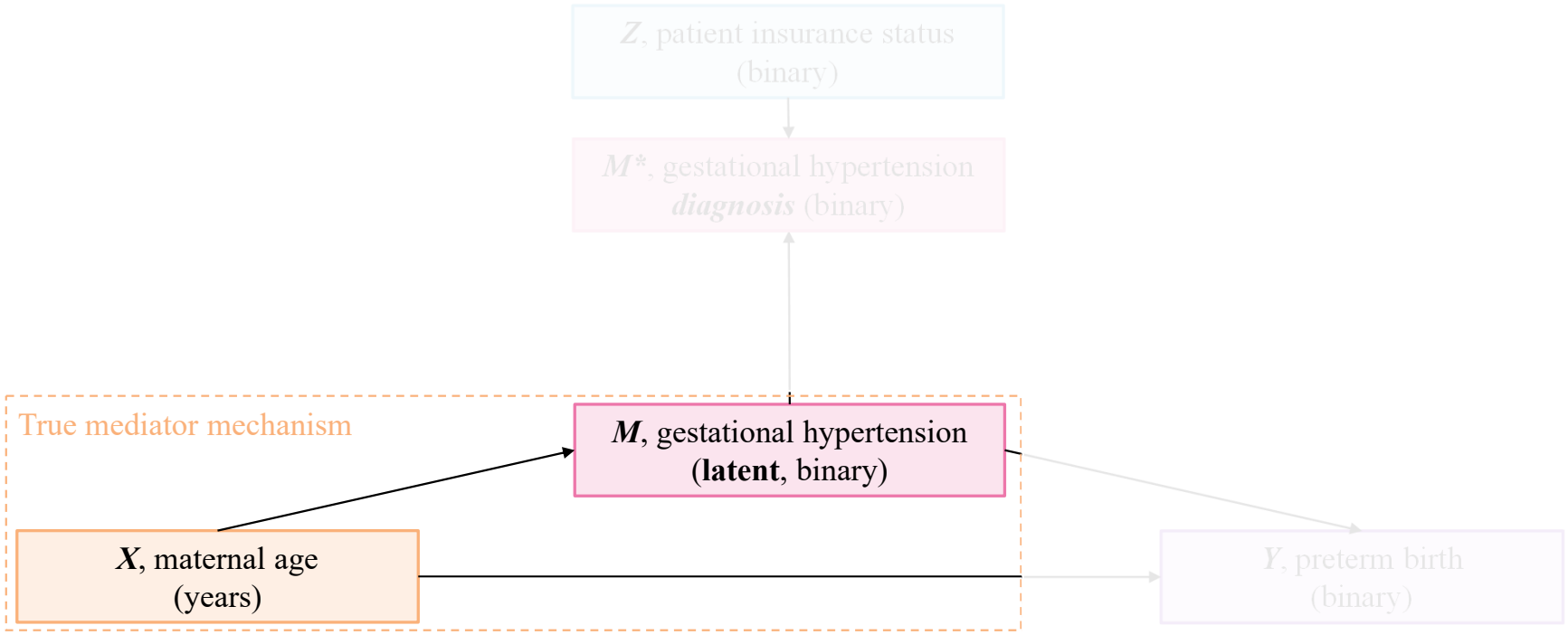
1



True mediator mechanism: $\text{logit}\{P(M = 1|X, \boldsymbol{C}; \boldsymbol{\beta})\} = \beta_0 + \beta_X X + \beta_C \boldsymbol{C}$

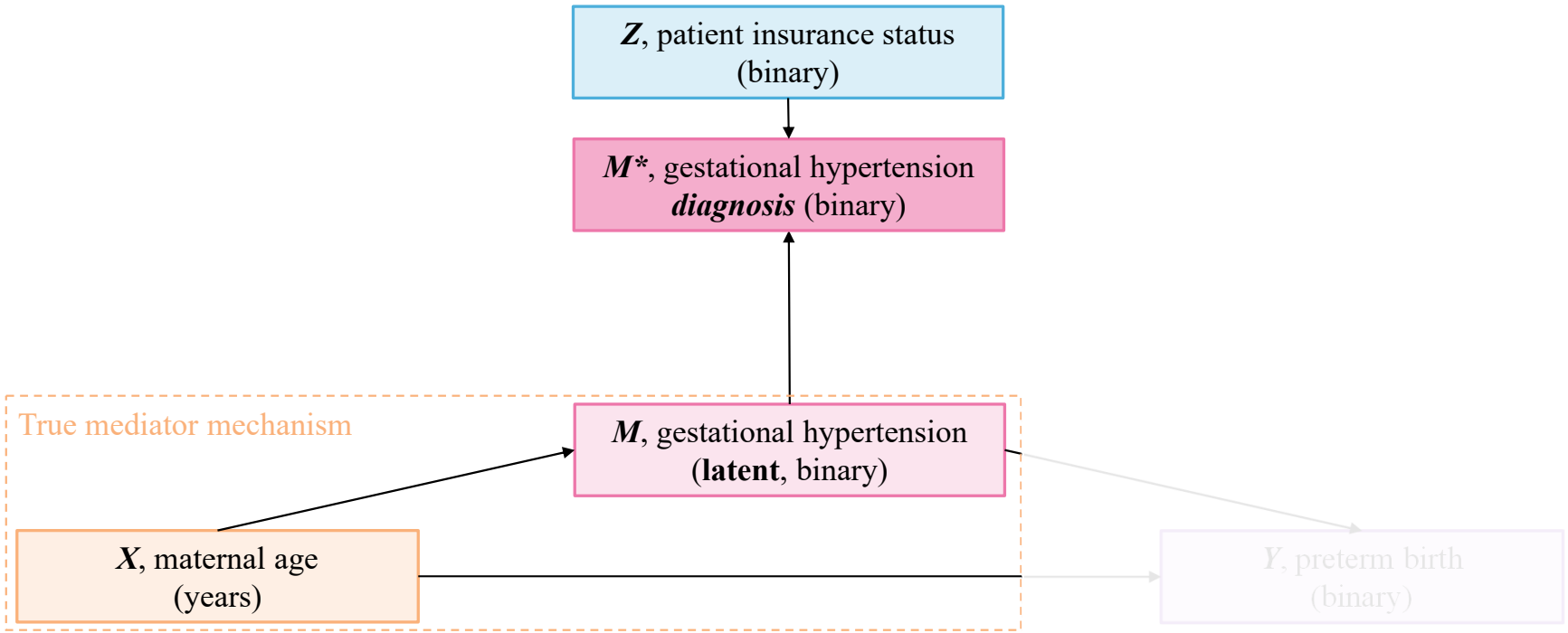
1

C, covariates

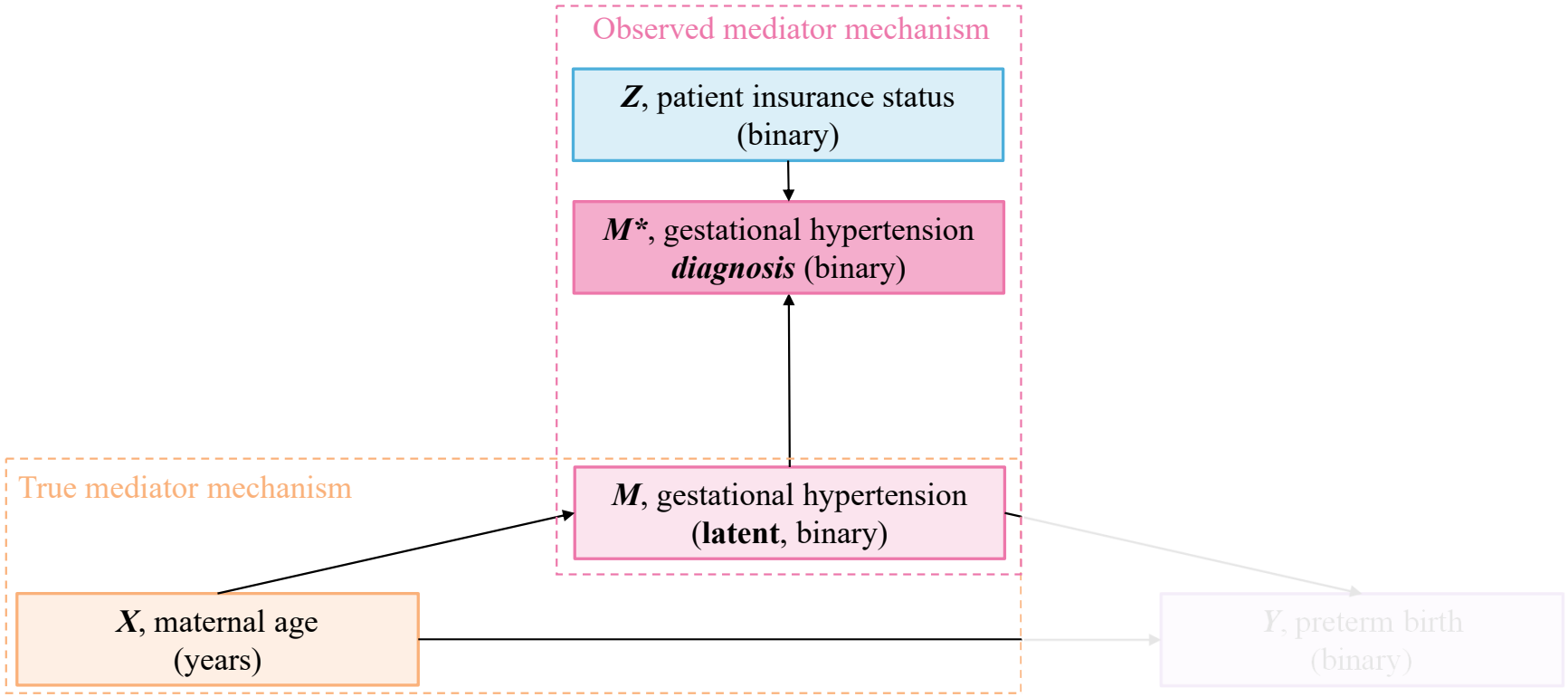


True mediator mechanism: $\text{logit}\{P(M = 1|X, \boldsymbol{C}; \boldsymbol{\beta})\} = \beta_0 + \beta_X X + \boldsymbol{\beta_C C}$

1

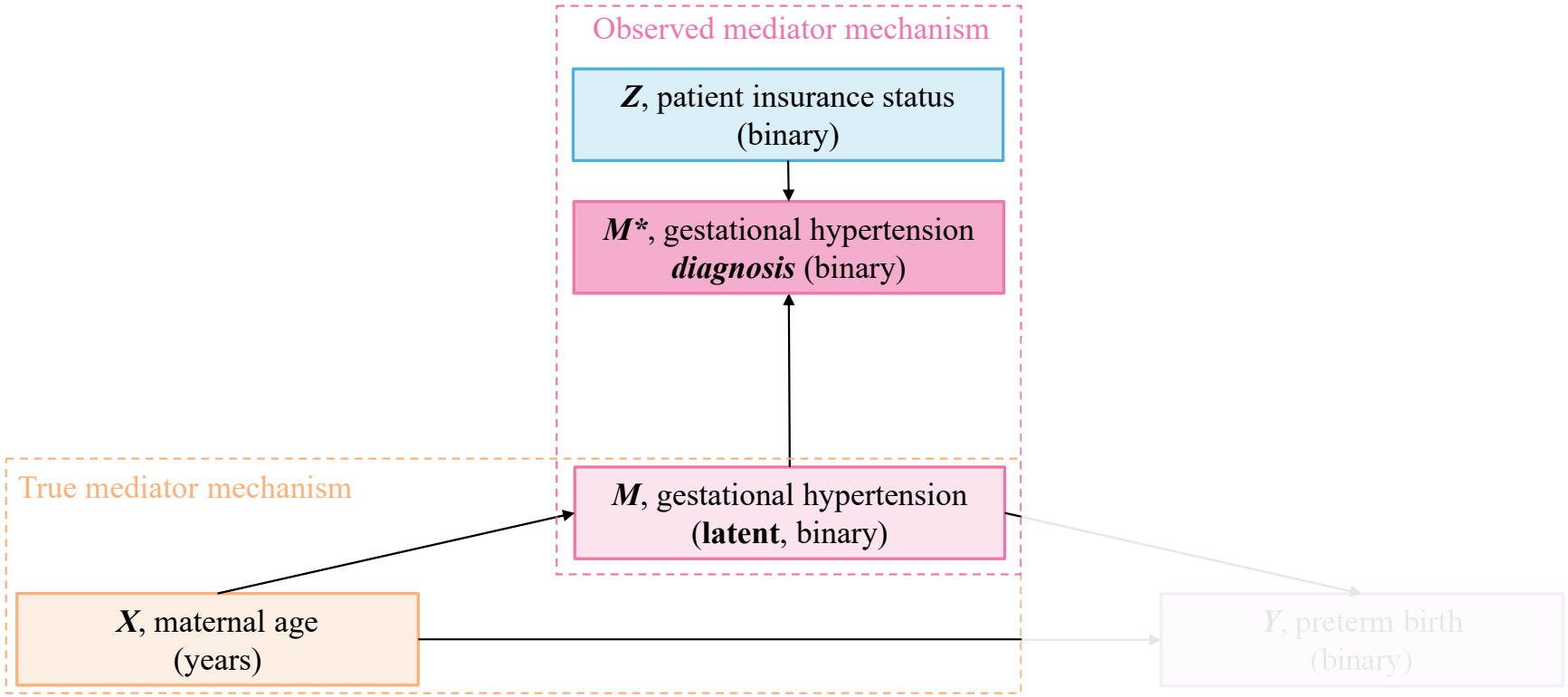


True mediator mechanism: $\text{logit}\{P(M = 1|X, \boldsymbol{C}; \boldsymbol{\beta})\} = \beta_0 + \beta_X X + \boldsymbol{\beta_C C}$ (1)



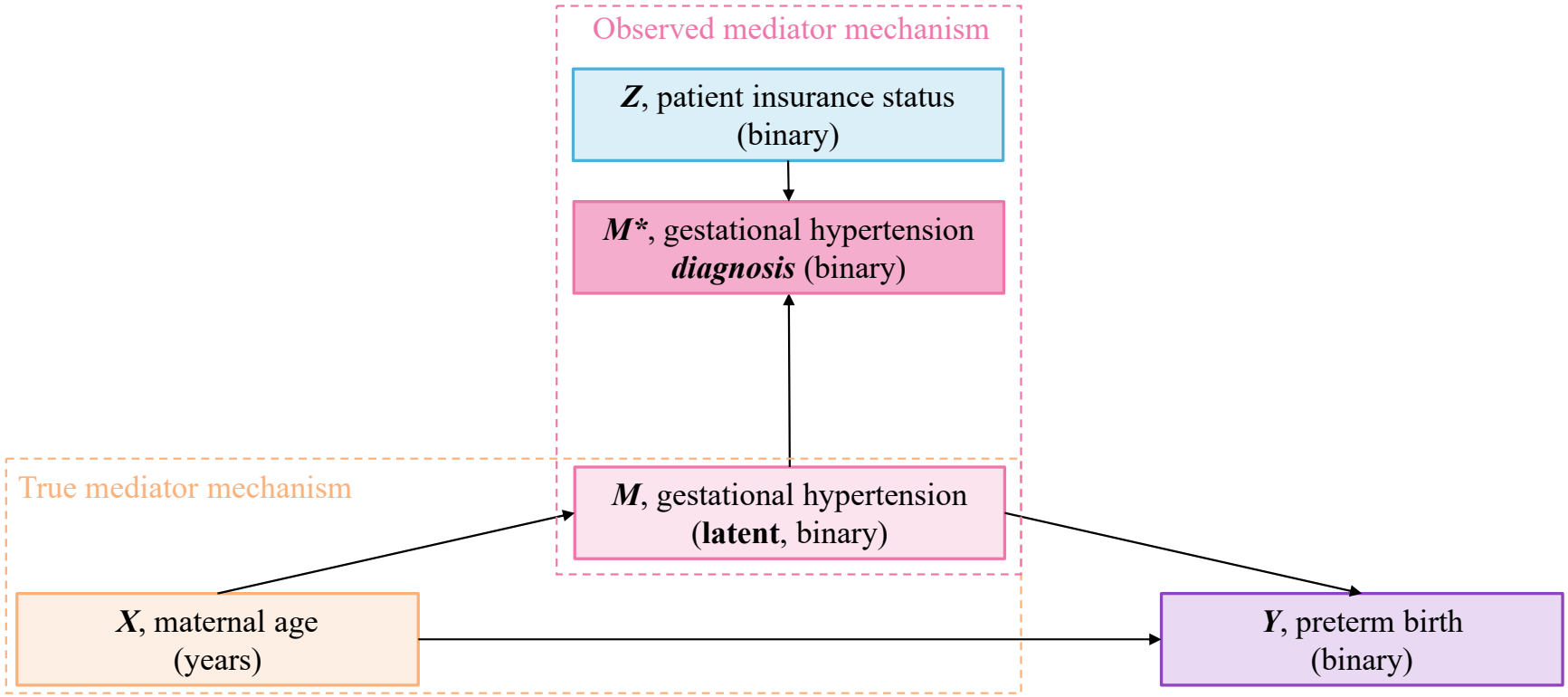
True mediator mechanism: $\text{logit}\{P(M = 1|X, \boldsymbol{C}; \boldsymbol{\beta})\} = \beta_0 + \beta_X X + \boldsymbol{\beta_C C}$ (1)

Observed mediator mechanism: $\text{logit}\{P(M^* = 1|M = m, \boldsymbol{Z}; \boldsymbol{\gamma})\} = \gamma_{1m0} + \boldsymbol{\gamma_{1mZ} Z}$



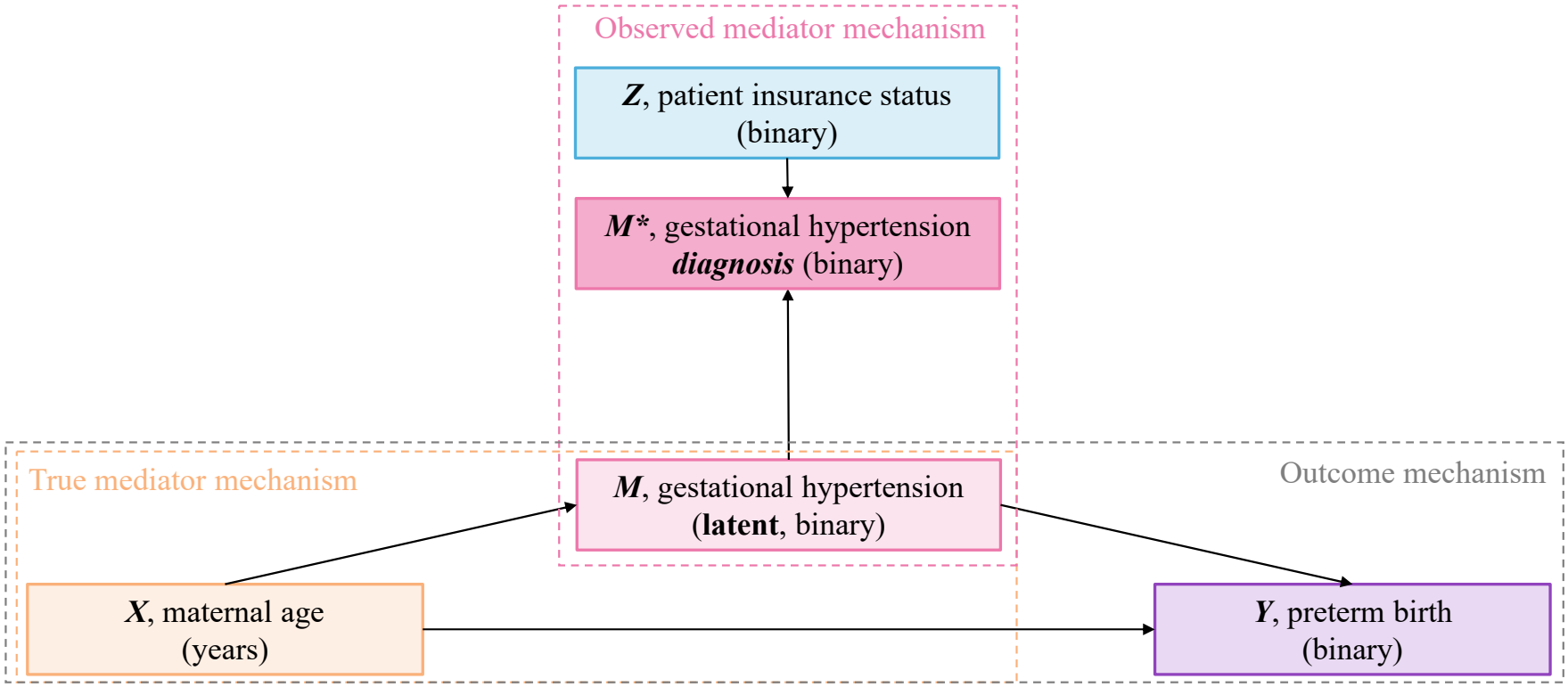
True mediator mechanism: $\text{logit}\{P(M = 1|X, \boldsymbol{C}; \boldsymbol{\beta})\} = \beta_0 + \beta_X X + \boldsymbol{\beta_C C}$ (1)

Observed mediator mechanism: $\text{logit}\{P(M^* = 1|M = m, \boldsymbol{Z}; \boldsymbol{\gamma})\} = \gamma_{1m0} + \boldsymbol{\gamma_{1mZ} Z}$



True mediator mechanism: $\text{logit}\{P(M = 1|X, \boldsymbol{C}; \boldsymbol{\beta})\} = \beta_0 + \beta_X X + \boldsymbol{\beta_C C}$ (1)

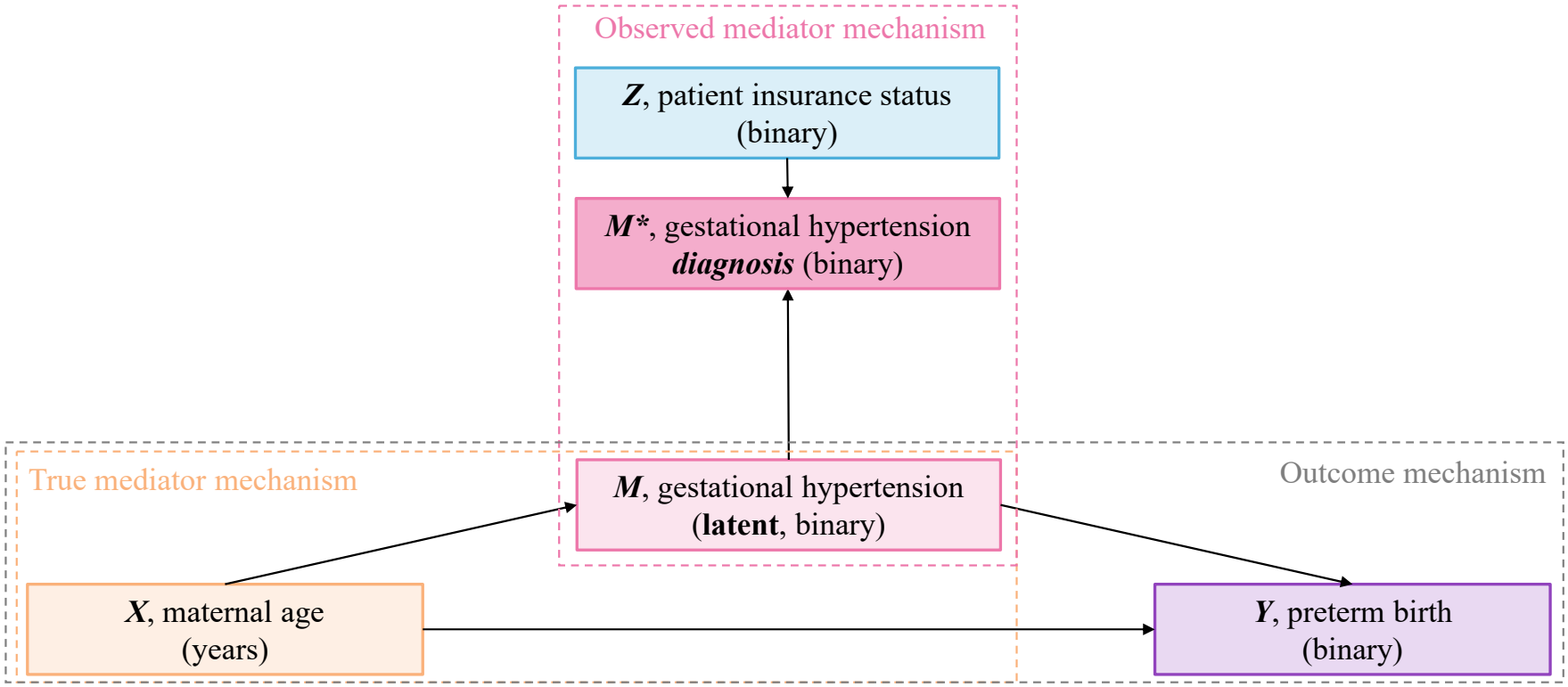
Observed mediator mechanism: $\text{logit}\{P(M^* = 1|M = m, \boldsymbol{Z}; \boldsymbol{\gamma})\} = \gamma_{1m0} + \boldsymbol{\gamma_{1mZ} Z}$



True mediator mechanism: $\text{logit}\{P(M = 1|X, \boldsymbol{C}; \boldsymbol{\beta})\} = \beta_0 + \beta_X X + \boldsymbol{\beta_C C}$ (1)

Observed mediator mechanism: $\text{logit}\{P(M^* = 1|M = m, \boldsymbol{Z}; \boldsymbol{\gamma})\} = \gamma_{1m0} + \boldsymbol{\gamma_{1mZ} Z}$

Outcome mechanism: $E(Y|X, \boldsymbol{C}, M) = \theta_0 + \theta_X X + \boldsymbol{\theta_C C} + \theta_M M + \theta_{XM} XM$



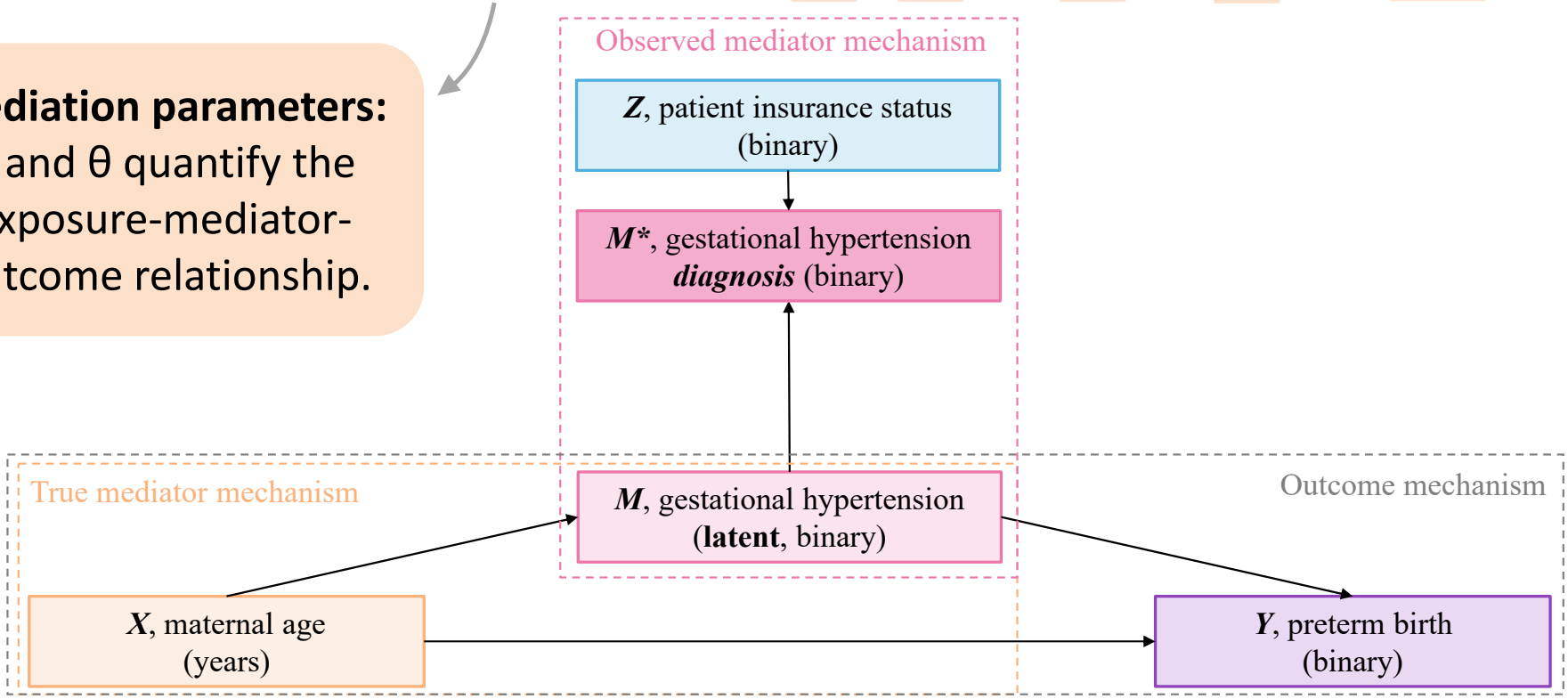
1

True mediator mechanism: $\text{logit}\{P(M = 1|X, \boldsymbol{C}; \boldsymbol{\beta})\} = \beta_0 + \beta_X X + \boldsymbol{\beta_C C}$

Observed mediator mechanism: $\text{logit}\{P(M^* = 1|M = m, \boldsymbol{Z}; \boldsymbol{\gamma})\} = \gamma_{1m0} + \boldsymbol{\gamma_{1mZ} Z}$

Outcome mechanism: $E(Y|X, \boldsymbol{C}, M) = \theta_0 + \theta_X X + \boldsymbol{\theta_C C} + \theta_M M + \theta_{XM} XM$

Mediation parameters:
β and θ quantify the exposure-mediator-outcome relationship.



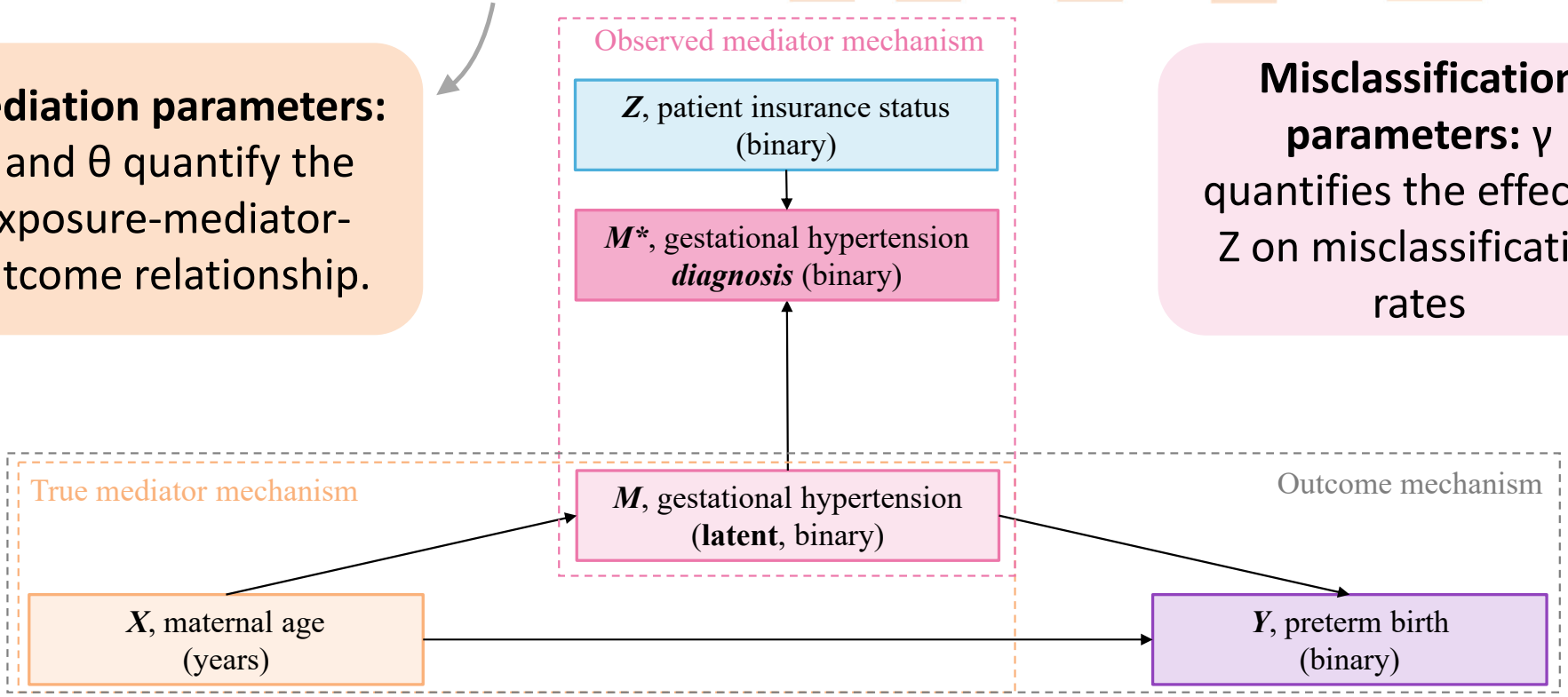
True mediator mechanism: $\text{logit}\{P(M = 1|X, \boldsymbol{C}; \boldsymbol{\beta})\} = \beta_0 + \beta_X X + \boldsymbol{\beta_C C}$ (1)

Observed mediator mechanism: $\text{logit}\{P(M^* = 1|M = m, \boldsymbol{Z}; \boldsymbol{\gamma})\} = \gamma_{1m0} + \boldsymbol{\gamma_{1mZ} Z}$

Outcome mechanism: $E(Y|X, \boldsymbol{C}, M) = \theta_0 + \theta_X X + \boldsymbol{\theta_C C} + \theta_M M + \theta_{XM} XM$

Mediation parameters: β and θ quantify the exposure-mediator-outcome relationship.

Misclassification parameters: γ quantifies the effect of Z on misclassification rates



2

Estimation

True mediator mechanism: $\text{logit}\{P(M = 1|X, \boldsymbol{C}; \boldsymbol{\beta})\} = \beta_0 + \beta_X X + \boldsymbol{\beta_C C}$

Observed mediator mechanism: $\text{logit}\{P(M^* = 1|M = m, \boldsymbol{Z}; \boldsymbol{\gamma})\} = \gamma_{1m0} + \boldsymbol{\gamma_{1mZ} Z}$

Outcome mechanism: $E(Y|X, \boldsymbol{C}, M) = \theta_0 + \theta_X X + \boldsymbol{\theta_C C} + \theta_M M + \theta_{XM} XM$

#1: OLS Correction	#2: Predictive value weighting	#3: An EM algorithm

Estimation

True mediator mechanism: $\text{logit}\{P(M = 1|X, \boldsymbol{C}; \boldsymbol{\beta})\} = \beta_0 + \beta_X X + \boldsymbol{\beta_C C}$

Observed mediator mechanism: $\text{logit}\{P(M^* = 1|M = m, \boldsymbol{Z}; \boldsymbol{\gamma})\} = \gamma_{1m0} + \boldsymbol{\gamma_{1mZ} Z}$

Outcome mechanism: $E(Y|X, \boldsymbol{C}, M) = \theta_0 + \theta_X X + \boldsymbol{\theta_C C} + \theta_M M + \theta_{XM} XM$

#1: OLS Correction	#2: Predictive value weighting	#3: An EM algorithm

Key point: We can use ***COMBO*** to estimate subject-level sensitivity and specificity, and then plug these values into existing misclassification correction procedures.

- Existing procedures relied on *known* sensitivity and specificity.

1. Extended from Nguimkeu, Rosenman, and Tennekoon (2021), "Regression with a misclassified binary regressor: Correcting for hidden bias".
2. Extended from Lyles and Lin (2010), "Sensitivity analysis for misclassification in logistic regression via likelihood methods and PVW".

Estimation

True mediator mechanism: $\text{logit}\{P(M = 1|X, \boldsymbol{C}; \boldsymbol{\beta})\} = \beta_0 + \beta_X X + \boldsymbol{\beta_C C}$

Observed mediator mechanism: $\text{logit}\{P(M^* = 1|M = m, \boldsymbol{Z}; \boldsymbol{\gamma})\} = \gamma_{1m0} + \boldsymbol{\gamma_{1mZ} Z}$

Outcome mechanism: $E(Y|X, \boldsymbol{C}, M) = \theta_0 + \theta_X X + \boldsymbol{\theta_C C} + \theta_M M + \theta_{XM} XM$

#1: OLS Correction | #2: Predictive value weighting | #3: An EM algorithm

Complete data log-likelihood:

$$\ell_{complete}(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\gamma}; X, \boldsymbol{C}, \boldsymbol{Z}, Y)$$

$$= \sum_{i=1}^{N} \left[\ell_{Y|X,M,\boldsymbol{C}}(\boldsymbol{\theta}; X_i, M_i, \boldsymbol{C}_i, Y_i) + \sum_{j=1}^{2} m_{ij} \log\{\pi_{ij}\} + \sum_{j=1}^{2} \sum_{\ell=1}^{2} m_{ij} m^*_{i\ell} \log\{\pi^*_{i\ell j}\} \right]$$

Estimation

True mediator mechanism: $\text{logit}\{P(M = 1|X, \boldsymbol{C}; \boldsymbol{\beta})\} = \beta_0 + \beta_X X + \boldsymbol{\beta_C C}$

Observed mediator mechanism: $\text{logit}\{P(M^* = 1|M = m, \boldsymbol{Z}; \boldsymbol{\gamma})\} = \gamma_{1m0} + \boldsymbol{\gamma_{1mZ} Z}$

Outcome mechanism: $E(Y|X, \boldsymbol{C}, M) = \theta_0 + \theta_X X + \boldsymbol{\theta_C C} + \theta_M M + \theta_{XM} XM$

#1: OLS Correction | #2: Predictive value weighting | #3: An EM algorithm

Complete data log-likelihood:

$$\ell_{complete}(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\gamma}; X, \boldsymbol{C}, \boldsymbol{Z}, Y)$$

$$= \sum_{i=1}^{N} \left[\ell_{Y|X,M,\boldsymbol{C}}(\boldsymbol{\theta}; X_i, M_i, \boldsymbol{C}_i, Y_i) + \sum_{j=1}^{2} m_{ij} \log\{\pi_{ij}\} + \sum_{j=1}^{2} \sum_{\ell=1}^{2} m_{ij} m^*_{i\ell} \log\{\pi^*_{i\ell j}\} \right]$$

Outcome

Estimation

True mediator mechanism: $\text{logit}\{P(M = 1|X, \boldsymbol{C}; \boldsymbol{\beta})\} = \beta_0 + \beta_X X + \boldsymbol{\beta_C C}$

Observed mediator mechanism: $\text{logit}\{P(M^* = 1|M = m, \boldsymbol{Z}; \boldsymbol{\gamma})\} = \gamma_{1m0} + \boldsymbol{\gamma_{1mZ} Z}$

Outcome mechanism: $E(Y|X, \boldsymbol{C}, M) = \theta_0 + \theta_X X + \boldsymbol{\theta_C C} + \theta_M M + \theta_{XM} XM$

#1: OLS Correction | #2: Predictive value weighting | #3: An EM algorithm

Complete data log-likelihood:

$$\ell_{complete}(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\gamma}; X, \boldsymbol{C}, \boldsymbol{Z}, Y)$$

$$= \sum_{i=1}^{N} \left[\ell_{Y|X,M,\boldsymbol{C}}(\boldsymbol{\theta}; X_i, M_i, \boldsymbol{C}_i, Y_i) + \sum_{j=1}^{2} \mathrm{I}(M_i = j) \log\{P(M_i = j)\} + \sum_{j=1}^{2} \sum_{\ell=1}^{2} m_{ij} m^*_{i\ell} \log\{\pi^*_{i\ell j}\} \right]$$

Outcome | True mediator

Estimation

True mediator mechanism: $\text{logit}\{P(M = 1|X, \boldsymbol{C}; \boldsymbol{\beta})\} = \beta_0 + \beta_X X + \boldsymbol{\beta_C C}$

Observed mediator mechanism: $\text{logit}\{P(M^* = 1|M = m, \boldsymbol{Z}; \boldsymbol{\gamma})\} = \gamma_{1m0} + \boldsymbol{\gamma_{1mZ} Z}$

Outcome mechanism: $E(Y|X, \boldsymbol{C}, M) = \theta_0 + \theta_X X + \boldsymbol{\theta_C C} + \theta_M M + \theta_{XM} XM$

#1: OLS Correction | #2: Predictive value weighting | **#3: An EM algorithm**

Complete data log-likelihood:

$$\ell_{complete}(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\gamma}; X, \boldsymbol{C}, \boldsymbol{Z}, Y)$$

$$= \sum_{i=1}^{N} \left[\ell_{Y|X,M,\boldsymbol{C}}(\theta; X_i, M_i, \boldsymbol{C}_i, Y_i) + \sum_{j=1}^{2} m_{ij} \log\{\pi_{ij}\} + \sum_{j=1}^{2} \sum_{\ell=1}^{2} m_{ij} m^*_{i\ell} \log\{\pi^*_{i\ell j}\} \right]$$

$P(M_i = j)$ — $\mathbb{I}(M_i = j)$ — $P(M_i^* = \ell \mid M_i = j)$ — $\mathbb{I}(M_i^* = \ell)$

Outcome | True mediator | Observed mediator

Estimation

True mediator mechanism: $\text{logit}\{P(M = 1|X, \boldsymbol{C}; \boldsymbol{\beta})\} = \beta_0 + \beta_X X + \boldsymbol{\beta_C C}$

Observed mediator mechanism: $\text{logit}\{P(M^* = 1|M = m, \boldsymbol{Z}; \boldsymbol{\gamma})\} = \gamma_{1m0} + \boldsymbol{\gamma_{1mZ} Z}$

Outcome mechanism: $E(Y|X, \boldsymbol{C}, M) = \theta_0 + \theta_X X + \boldsymbol{\theta_C C} + \theta_M M + \theta_{XM} XM$

#1: OLS Correction | #2: Predictive value weighting | #3: An EM algorithm

Expectation Step ⟷ **Maximization Step**

Estimation

True mediator mechanism: $\text{logit}\{P(M = 1|X, \boldsymbol{C}; \boldsymbol{\beta})\} = \beta_0 + \beta_X X + \boldsymbol{\beta_C C}$

Observed mediator mechanism: $\text{logit}\{P(M^* = 1|M = m, \boldsymbol{Z}; \boldsymbol{\gamma})\} = \gamma_{1m0} + \boldsymbol{\gamma_{1mZ} Z}$

Outcome mechanism: $E(Y|X, \boldsymbol{C}, M) = \theta_0 + \theta_X X + \boldsymbol{\theta_C C} + \theta_M M + \theta_{XM} XM$

#1: OLS Correction | #2: Predictive value weighting | #3: An EM algorithm

Expectation Step ⟷ **Maximization Step**

$$w_{ij} = P(M_i = j|M_i^*, X_i, \boldsymbol{C}_i, \boldsymbol{Z}_i, Y_i)$$

$$= \sum_{\ell=1}^{2} \frac{m_{i\ell}^* \pi_{i\ell j}^* \pi_{ij} E[Y_i|X_i, M_i = j, \boldsymbol{C}_i, \boldsymbol{\theta}^{(t)}]}{\sum_{k=1}^{2} \pi_{i\ell k}^* \pi_{ik} E[Y_i|X_i, M_i = k, \boldsymbol{C}_i, \boldsymbol{\theta}^{(t)}]}$$

Estimation

True mediator mechanism: $\text{logit}\{P(M = 1|X, \boldsymbol{C}; \boldsymbol{\beta})\} = \beta_0 + \beta_X X + \boldsymbol{\beta_C C}$

Observed mediator mechanism: $\text{logit}\{P(M^* = 1|M = m, \boldsymbol{Z}; \boldsymbol{\gamma})\} = \gamma_{1m0} + \boldsymbol{\gamma_{1mZ} Z}$

Outcome mechanism: $E(Y|X, \boldsymbol{C}, M) = \theta_0 + \theta_X X + \boldsymbol{\theta_C C} + \theta_M M + \theta_{XM} XM$

#1: OLS Correction | #2: Predictive value weighting | #3: An EM algorithm

Expectation Step ⟷ **Maximization Step**

$$w_{ij} = P(M_i = j|M_i^*, X_i, \boldsymbol{C}_i, \boldsymbol{Z}_i, Y_i)$$

$$= \sum_{\ell=1}^{2} \frac{m_{i\ell}^* \pi_{i\ell j}^* \pi_{ij} E[Y_i|X_i, M_i = j, \boldsymbol{C}_i, \boldsymbol{\theta}^{(t)}]}{\sum_{k=1}^{2} \pi_{i\ell k}^* \pi_{ik} E[Y_i|X_i, M_i = k, \boldsymbol{C}_i, \boldsymbol{\theta}^{(t)}]}$$

$$Q = \sum_{i=1}^{N} [\sum_{j=1}^{2} \ell_{Y|X,M,\boldsymbol{C}}(\boldsymbol{\theta}; X_i, M_i = w_{ij}, \boldsymbol{C}_i, Y_i)$$

$$+ \sum_{j=1}^{2} w_{ij} \log\{\pi_{ij}\} + \sum_{j=1}^{2} \sum_{\ell=1}^{2} w_{ij} m_{i\ell}^* \log\{\pi_{i\ell j}^*\}]$$

Estimation

True mediator mechanism: $\text{logit}\{P(M = 1|X, \boldsymbol{C}; \boldsymbol{\beta})\} = \beta_0 + \beta_X X + \boldsymbol{\beta_C C}$

Observed mediator mechanism: $\text{logit}\{P(M^* = 1|M = m, \boldsymbol{Z}; \boldsymbol{\gamma})\} = \gamma_{1m0} + \boldsymbol{\gamma_{1mZ} Z}$

Outcome mechanism: $E(Y|X, \boldsymbol{C}, M) = \theta_0 + \theta_X X + \boldsymbol{\theta_C C} + \theta_M M + \theta_{XM} XM$

#1: OLS Correction | #2: Predictive value weighting | #3: An EM algorithm

Expectation Step ⟷ **Maximization Step**

Apply label switching correction from Webb and Wells (2023)

$$w_{ij} = P(M_i = j|M_i^*, X_i, \boldsymbol{C}_i, \boldsymbol{Z}_i, Y_i)$$

$$= \sum_{\ell=1}^{2} \frac{m_{i\ell}^* \pi_{i\ell j}^* \pi_{ij} E[Y_i|X_i, M_i = j, \boldsymbol{C}_i, \boldsymbol{\theta}^{(t)}]}{\sum_{k=1}^{2} \pi_{i\ell k}^* \pi_{ik} E[Y_i|X_i, M_i = k, \boldsymbol{C}_i, \boldsymbol{\theta}^{(t)}]}$$

$$Q = \sum_{i=1}^{N} \Big[\sum_{j=1}^{2} \ell_{Y|X,M,\boldsymbol{C}}(\boldsymbol{\theta}; X_i, M_i = w_{ij}, \boldsymbol{C}_i, Y_i)$$

$$+ \sum_{j=1}^{2} w_{ij} \log\{\pi_{ij}\} + \sum_{j=1}^{2} \sum_{\ell=1}^{2} w_{ij} m_{i\ell}^* \log\{\pi_{i\ell j}^*\} \Big]$$

2

Estimation

True mediator mechanism: $\text{logit}\{P(M = 1|X, \boldsymbol{C}; \boldsymbol{\beta})\} = \beta_0 + \beta_X X + \boldsymbol{\beta_C C}$

Observed mediator mechanism: $\text{logit}\{P(M^* = 1|M = m, \boldsymbol{Z}; \boldsymbol{\gamma})\} = \gamma_{1m0} + \boldsymbol{\gamma_{1mZ} Z}$

Outcome mechanism: $E(Y|X, \boldsymbol{C}, M) = \theta_0 + \theta_X X + \boldsymbol{\theta_C C} + \theta_M M + \theta_{XM} XM$

#1: OLS Correction	#2: Predictive value weighting	#3: An EM algorithm

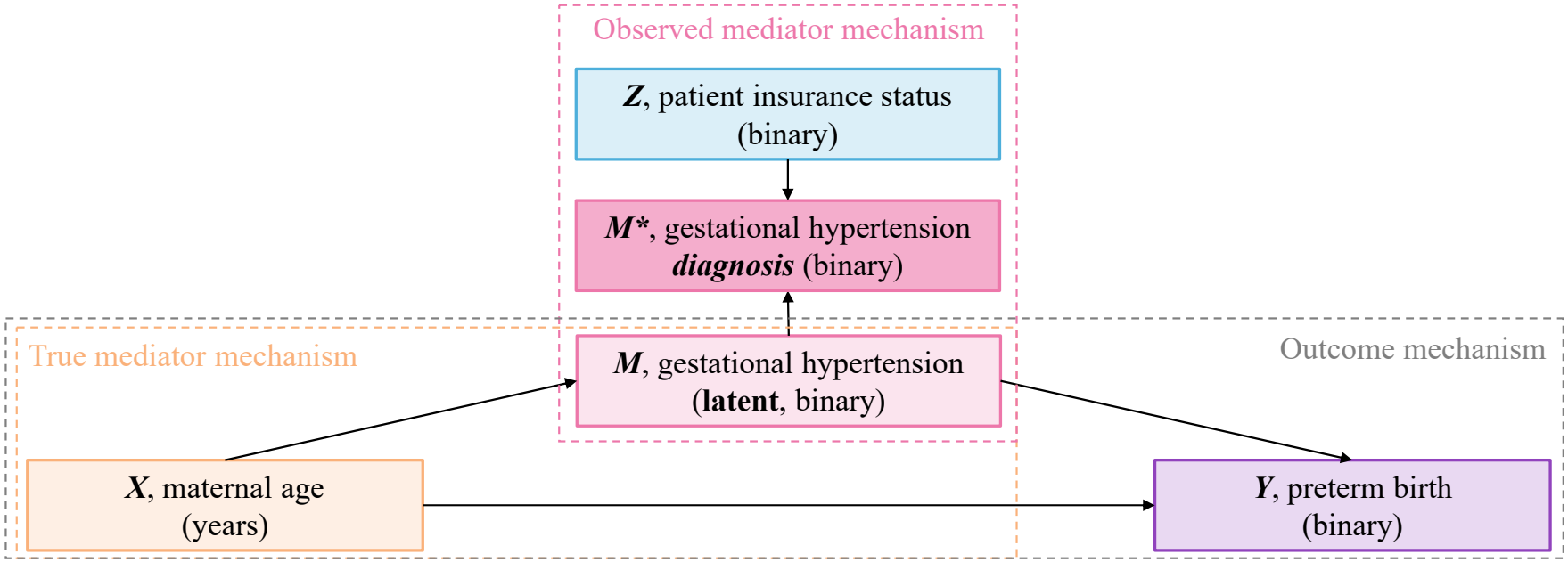
- Use the resulting bias-corrected parameter estimates to compute **(in)direct effects** for a change from **x̃ to x:**

$$OR^{NDE} \cong \frac{\exp(\theta_X x)\left\{1 + \exp\left(\theta_M + \theta_{XM} x + \beta_0 + \beta_X \tilde{x} + \beta_C c\right)\right\}}{\exp(\theta_X \tilde{x})\left\{1 + \exp\left(\theta_M + \theta_{XM} \tilde{x} + \beta_0 + \beta_X \tilde{x} + \beta_C c\right)\right\}}$$

$$OR^{NIE} \cong \frac{\left\{1 + \exp\left(\beta_0 + \beta_X \tilde{x} + \beta_C c\right)\right\} 1 + \exp\left(\theta_X + \theta_{XM} x + \beta_0 + \beta_X x + \beta_C c\right)\right\}}{\left\{1 + \exp\left(\beta_0 + \beta_X x + \beta_C c\right)\right\}\left\{1 + \exp\left(\theta_M + \theta_{XM} x + \beta_0 + \beta_X \tilde{x} + \beta_C c\right)\right\}}$$

Preterm birth study

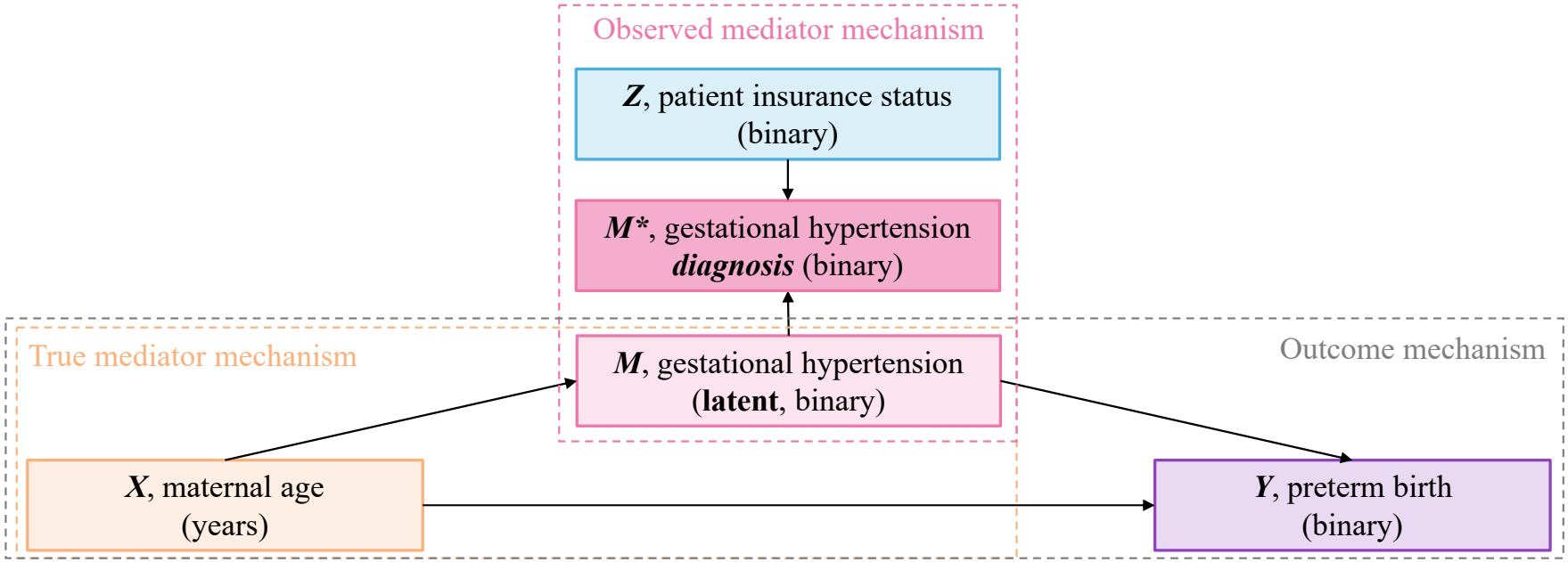
? Does **gestational hypertension** mediate the association between **maternal age** and **preterm birth**, after accounting for potential **misdiagnosis of gestational hypertension** based on **patient insurance status**?



Preterm birth study

Data: National Vital Statistics System

- Provides demographic and health data for all births in a year in the US.
- Random subsample from calendar year 2021, **N = 20,000**.

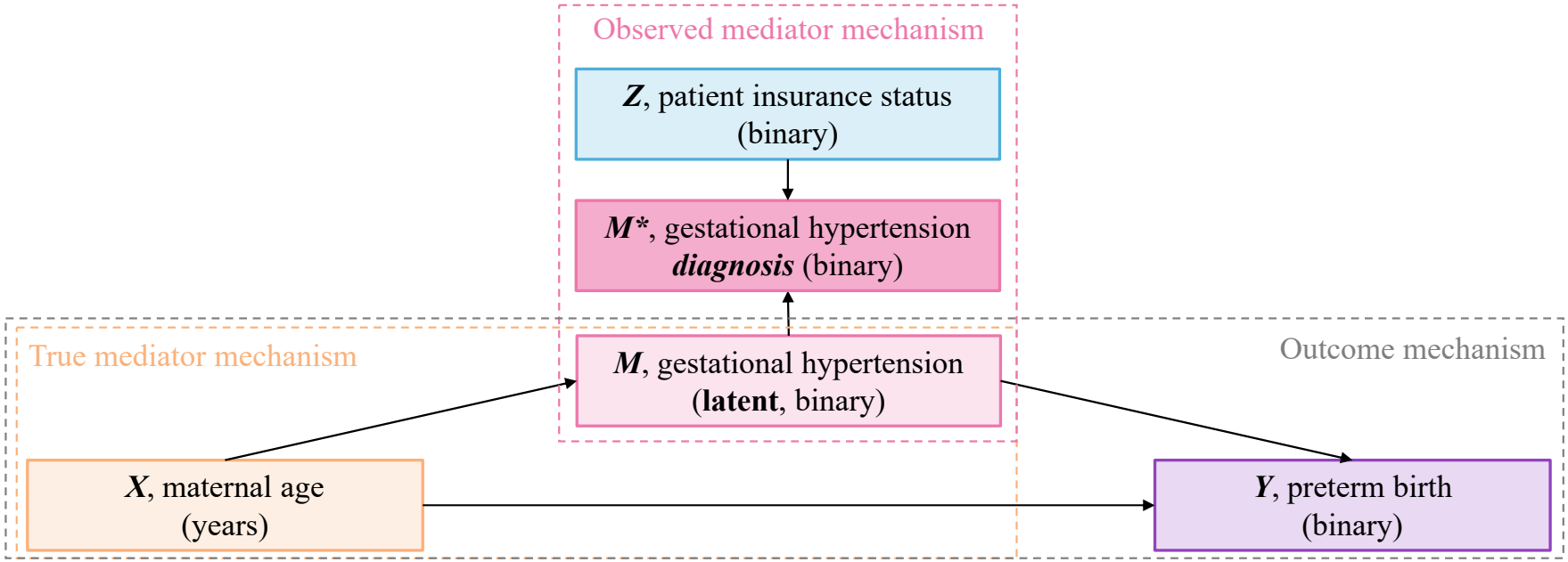


Preterm birth study

True mediator mechanism: **M** ~ **X** + Race + Education + Parity + Smoking + BMI

Observed mediator mechanism: **M*** | **M** ~ Race + **Z**

Outcome mechanism: **Y** ~ **X** + Race + Education + Parity + Smoking + BMI + **M** + **M** * **X**



Preterm birth study

True mediator mechanism: **M** ~ **X** + Race + Education + Parity + Smoking + BMI

Observed mediator mechanism: **M*** **|** **M** ~ Race + **Z**

Outcome mechanism: **Y** ~ **X** + Race + Education + Parity + Smoking + BMI + **M** + **M** * **X**

	EM Algorithm		**Naïve Analysis**	
	Est.	SE	Est.	SE
β_X				
$\gamma_{Z,\, G = 1}$				
$\gamma_{Z,\, G = 2}$				
θ_X				
θ_M				
θ_{XM}				

Preterm birth study

True mediator mechanism: **M** ~ **X** + Race + Education + Parity + Smoking + BMI

Observed mediator mechanism: **M*** | **M** ~ Race + **Z**

Outcome mechanism: **Y** ~ **X** + Race + Education + Parity + Smoking + BMI + **M** + **M** * **X**

Association between **age** & **gestational hypertension** is unchanged after accounting for misdiagnosis.

	EM Algorithm		**Naïve Analysis**	
	Est.	SE	Est.	SE
β_X	0.10	0.04	0.08	0.03
$\gamma_{Z,\, G=1}$				
$\gamma_{Z,\, G=2}$				
θ_X				
θ_M				
θ_{XM}				

Preterm birth study

True mediator mechanism: **M** ~ **X** + Race + Education + Parity + Smoking + BMI

Observed mediator mechanism: **M*** | **M** ~ Race + **Z**

Outcome mechanism: **Y** ~ **X** + Race + Education + Parity + Smoking + BMI + **M** + **M** * **X**

Association between **age** & **gestational hypertension** is unchanged after accounting for misdiagnosis.

Association between **gestational hypertension** & **preterm birth** strengthens.

	EM Algorithm		**Naïve Analysis**	
	Est.	SE	Est.	SE
β_X	0.10	0.04	0.08	0.03
$\gamma_{Z,\,G=1}$				
$\gamma_{Z,\,G=2}$				
θ_X	0.02	0.05	0.10	0.03
θ_M	1.19	0.17	0.88	0.06
θ_{XM}	0.19	0.09	0.06	0.06

Preterm birth study

True mediator mechanism: **M** ~ **X** + Race + Education + Parity + Smoking + BMI

Observed mediator mechanism: **M*** | **M** ~ Race + **Z**

Outcome mechanism: **Y** ~ **X** + Race + Education + Parity + Smoking + BMI + **M** + **M** * **X**

Association between **age** & **gestational hypertension** is unchanged after accounting for misdiagnosis.

Association between **gestational hypertension** & **preterm birth** strengthens.

	EM Algorithm		**Naïve Analysis**	
	Est.	SE	Est.	SE
β_X	0.10	0.04	0.08	0.03
θ_X	0.02	0.05	0.10	0.03
θ_M	1.19	0.17	0.88	0.06
θ_{XM}	0.19	0.09	0.06	0.06

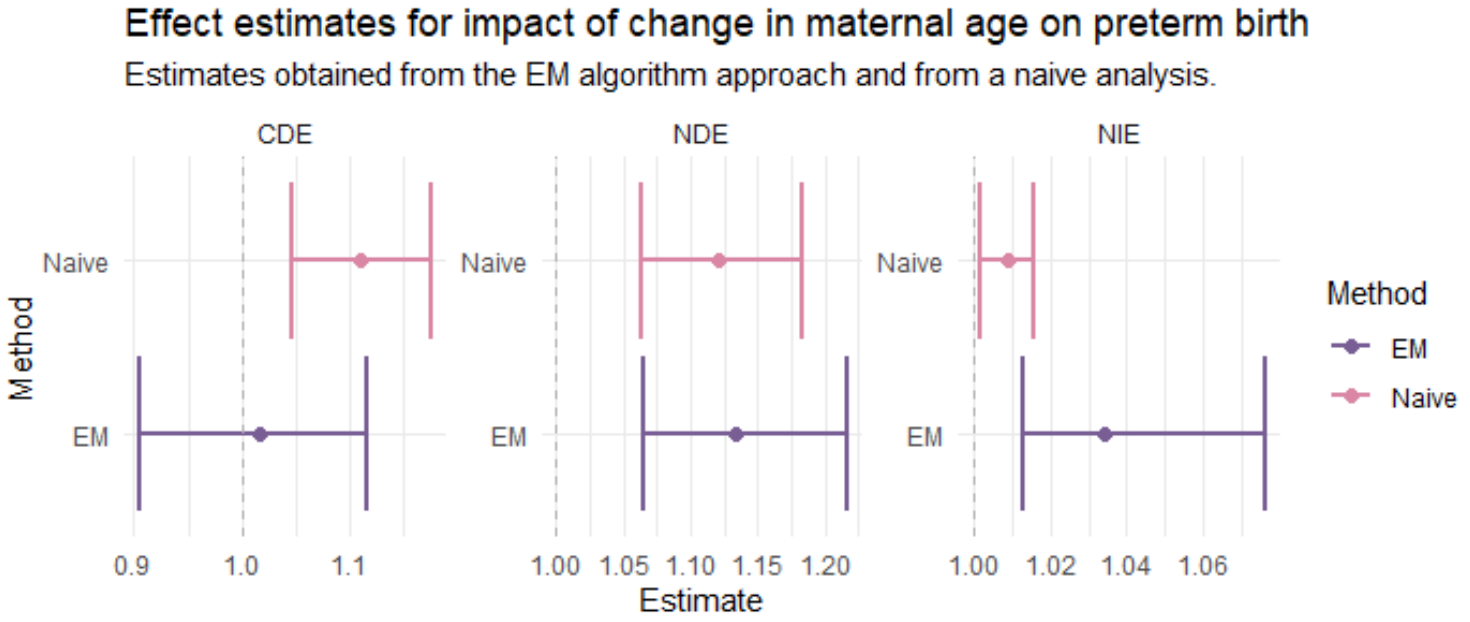
Use **β** and **θ** parameter estimates to compute **(in)direct effects**.

Preterm birth study

True mediator mechanism: **M** ~ **X** + Race + Education + Parity + Smoking + BMI

Observed mediator mechanism: **M*** | **M** ~ Race + **Z**

Outcome mechanism: **Y** ~ **X** + Race + Education + Parity + Smoking + BMI + **M** + **M** * **X**



Use **β** and **θ** parameter estimates to compute **(in)direct effects**.

Preterm birth study

True mediator mechanism: **M** ~ **X** + Race + Education + Parity + Smoking + BMI

Observed mediator mechanism: **M*** | **M** ~ Race + **Z**

Outcome mechanism: **Y** ~ **X** + Race + Education + Parity + Smoking + BMI + **M** + **M** * **X**

Association between **age** & **gestational hypertension** is unchanged after accounting for misdiagnosis.

Association between **gestational hypertension** & **preterm birth** strengthens.

	EM Algorithm		**Naïve Analysis**	
	Est.	SE	Est.	SE
β_X	0.10	0.04	0.08	0.03
$\gamma_{Z,\,G=1}$	-1.01	0.40	-	-
$\gamma_{Z,\,G=2}$	2.09	8.81	-	-
θ_X	0.02	0.05	0.10	0.03
θ_M	1.19	0.17	0.88	0.06
θ_{XM}	0.19	0.09	0.06	0.06

Use γ estimates to compute **sensitivity and specificity**.

Preterm birth study

True mediator mechanism: **M** ~ **X** + Race + Education + Parity + Smoking + BMI

Observed mediator mechanism: **M*** | **M** ~ Race + **Z**

Outcome mechanism: **Y** ~ **X** + Race + Education + Parity + Smoking + BMI + **M** + **M** * **X**

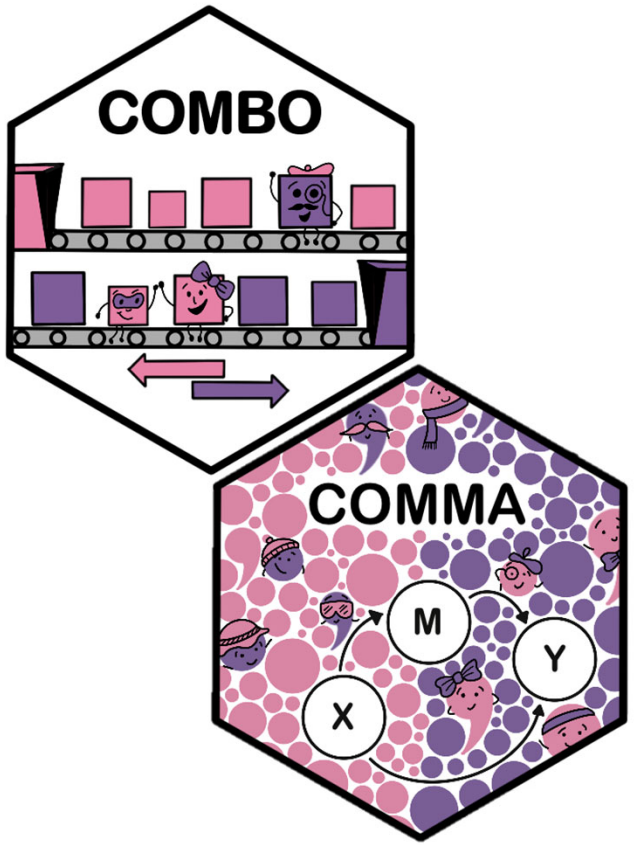
	Estimated Specificity P(no M* | no M)	**Estimated Sensitivity** P(M* | M)
Insured	99.9%	43.1%
Self-Pay	99.4%	21.7%

Key takeaways

- Developed new methods for handling misclassified binary mediator variables.
- Computed (in)direct effects using bias-corrected parameter estimates.
- Quantified gestational hypertension misdiagnosis rates based on insurance status.

Software

- Estimation methods for **misclassified outcomes** are available in the *COMBO* R Package on CRAN.
  - **Co**rrecting **M**isclassified **B**inary **O**utcomes
- Estimation methods for **misclassified mediators** are available in the *COMMA* R Package on CRAN.
  - **Co**rrecting **M**isclassified **M**ediation **A**nalysis



Thank you!

Kimberly A. H. Webb

kiw58@pitt.edu

kimhwebb.com → My "webb-site" ☺

Scan for the arXiv preprint

bit.ly/webb-mediation

Cornell Bowers CIS
Statistics and Data Science

University of Pittsburgh