Causal effect estimation in the presence of misclassified binary mediators

Kimberly A. H. Webb and Martin T. Wells

Women in Statistics and Data Science

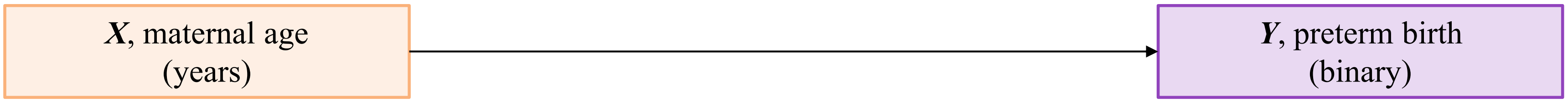
October 18, 2024

Mediation analysis

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

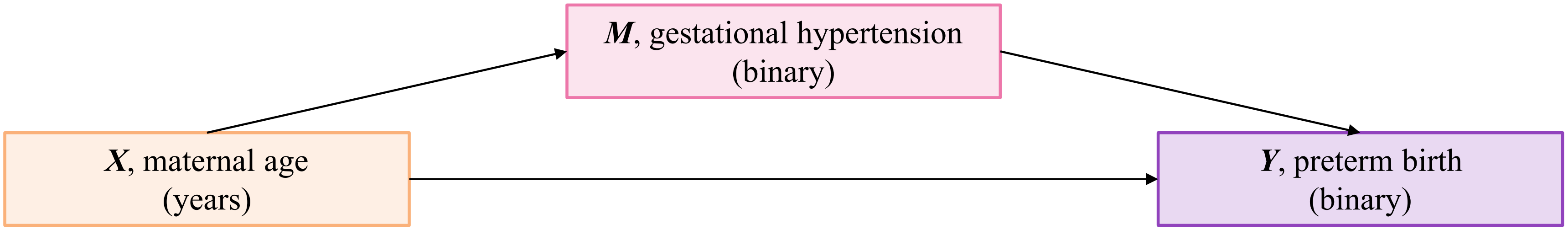
Mediation analysis

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



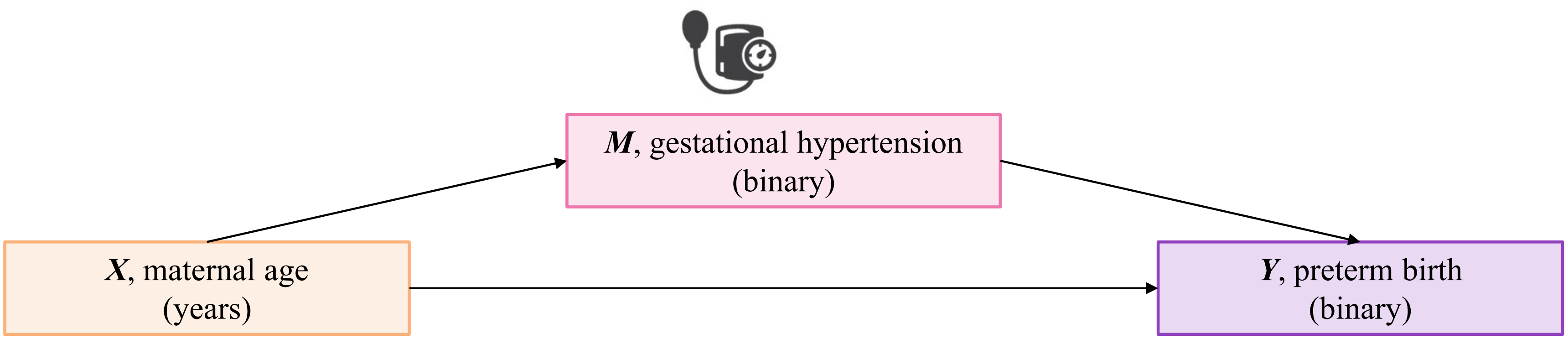
Mediation analysis

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



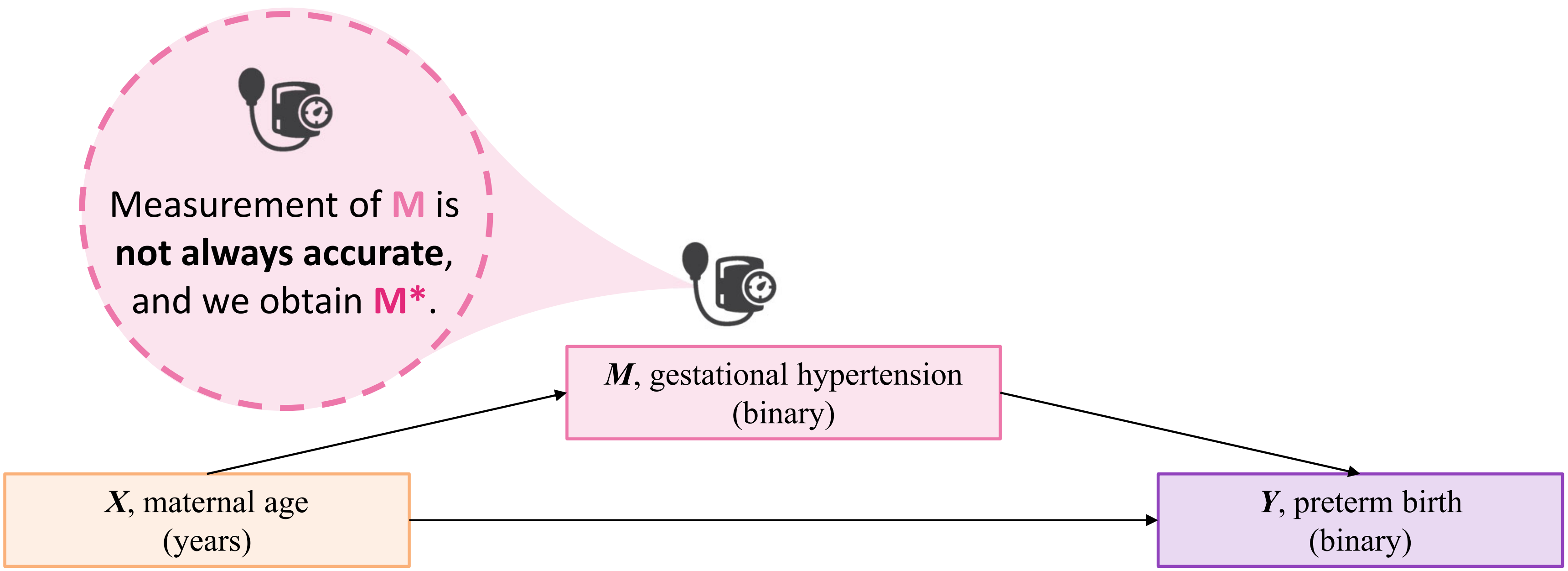
Mediation analysis

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



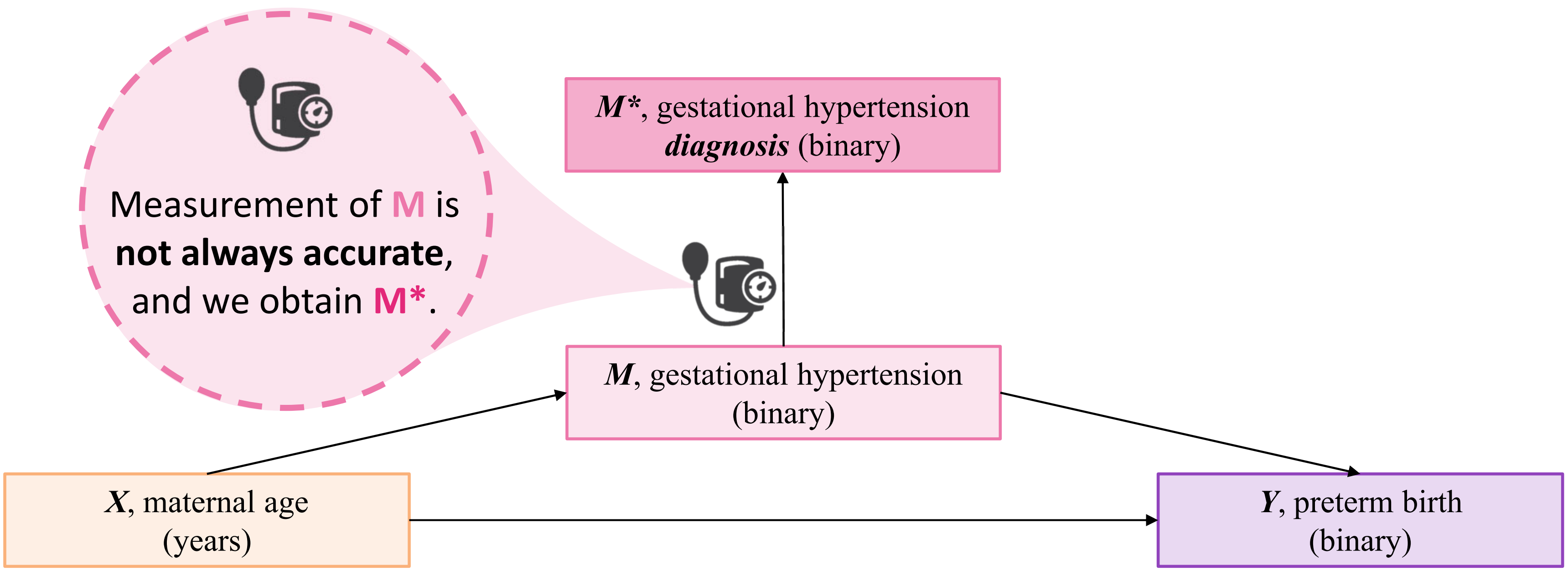
Mediation analysis

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



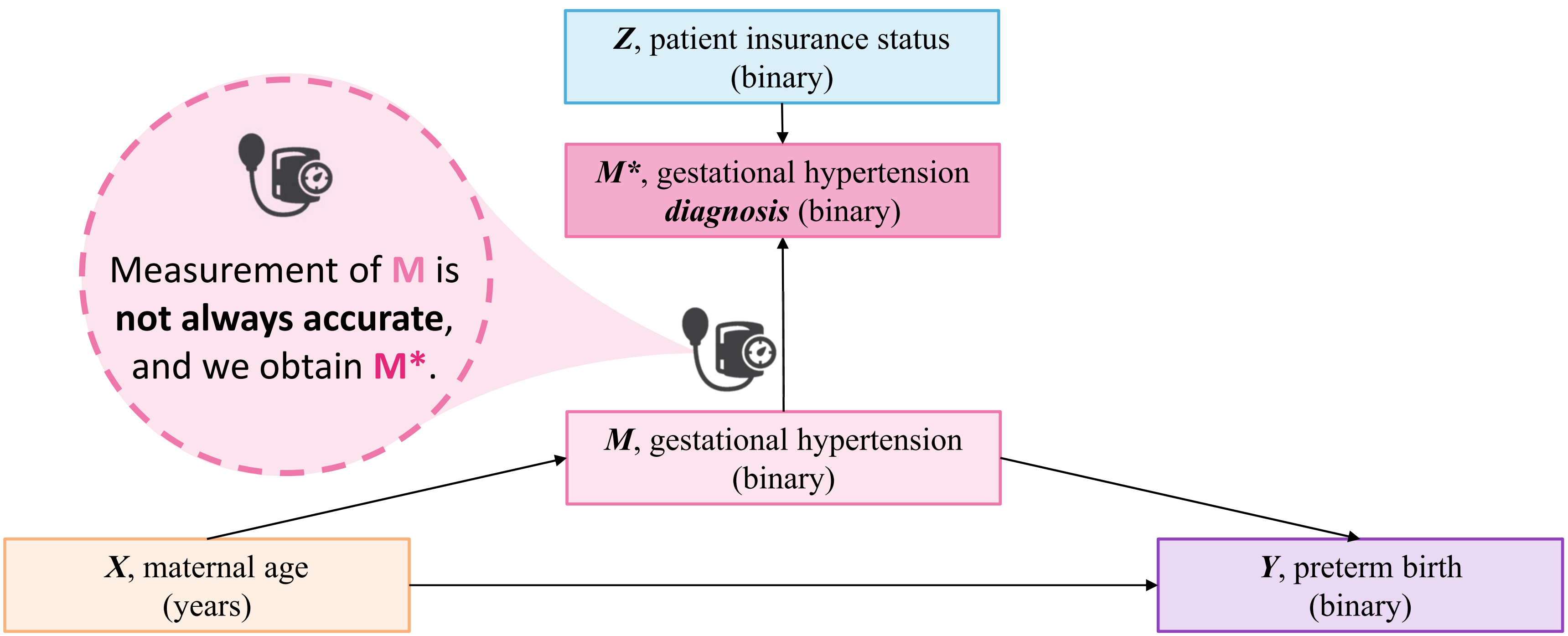
Mediation analysis

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



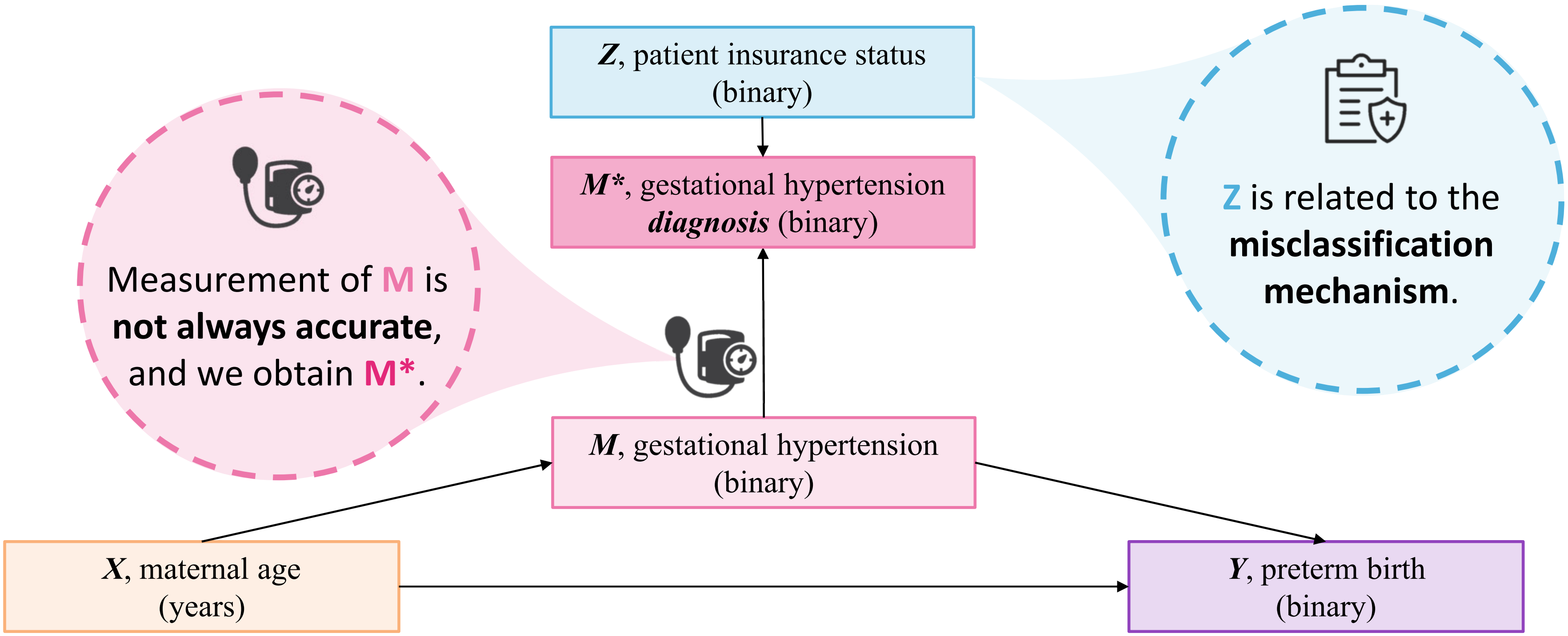
Mediation analysis

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

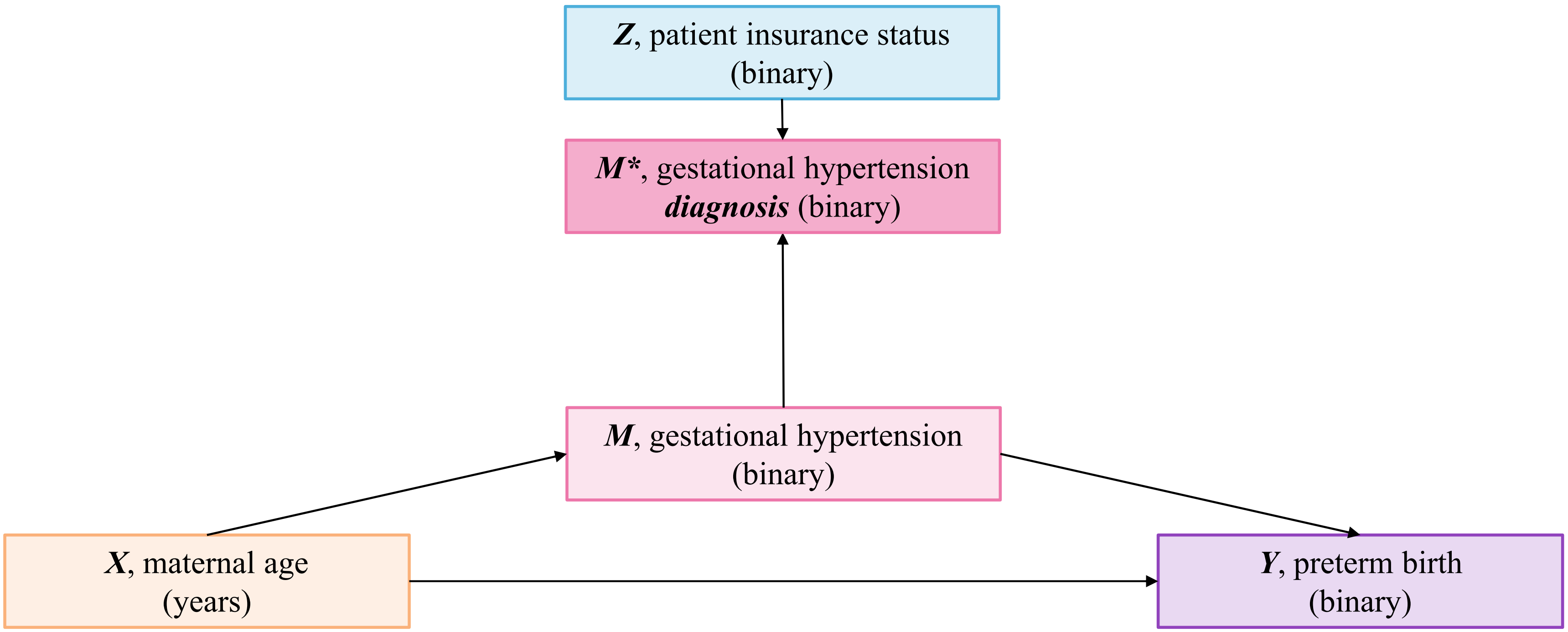


Mediation analysis

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

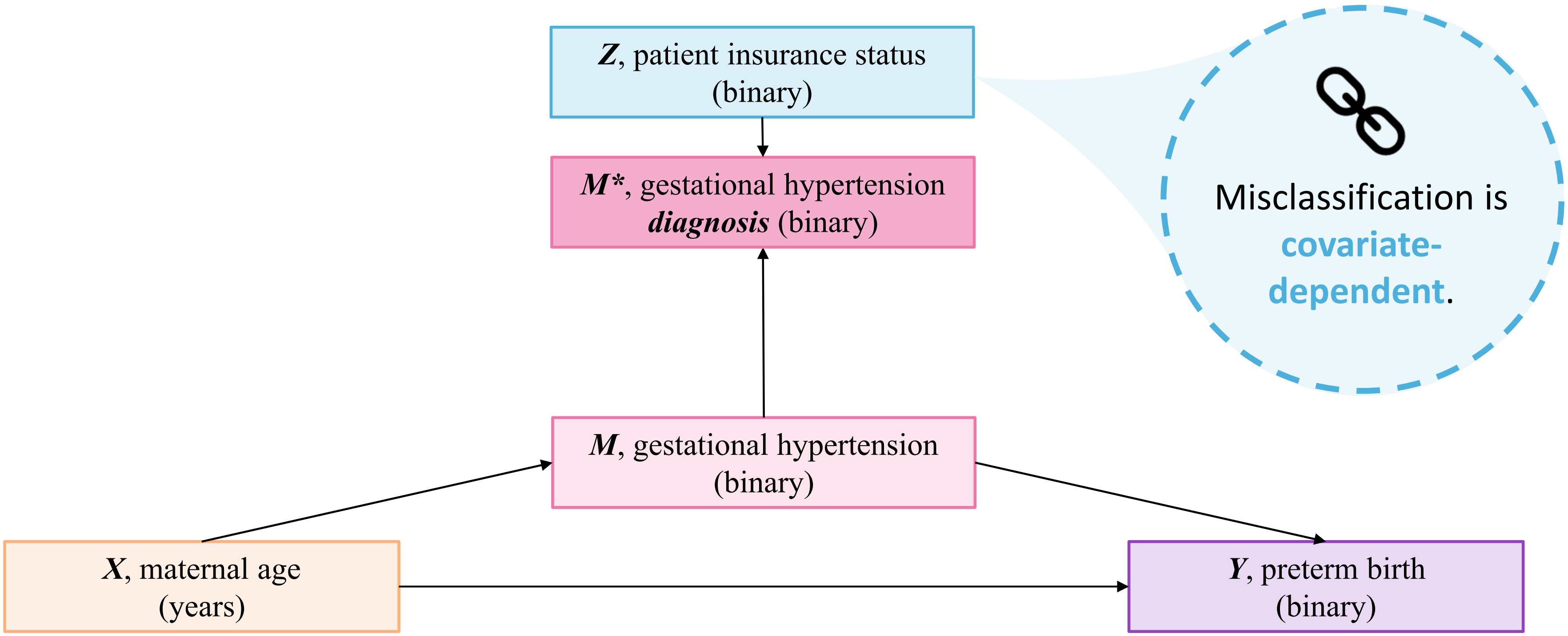


Challenges



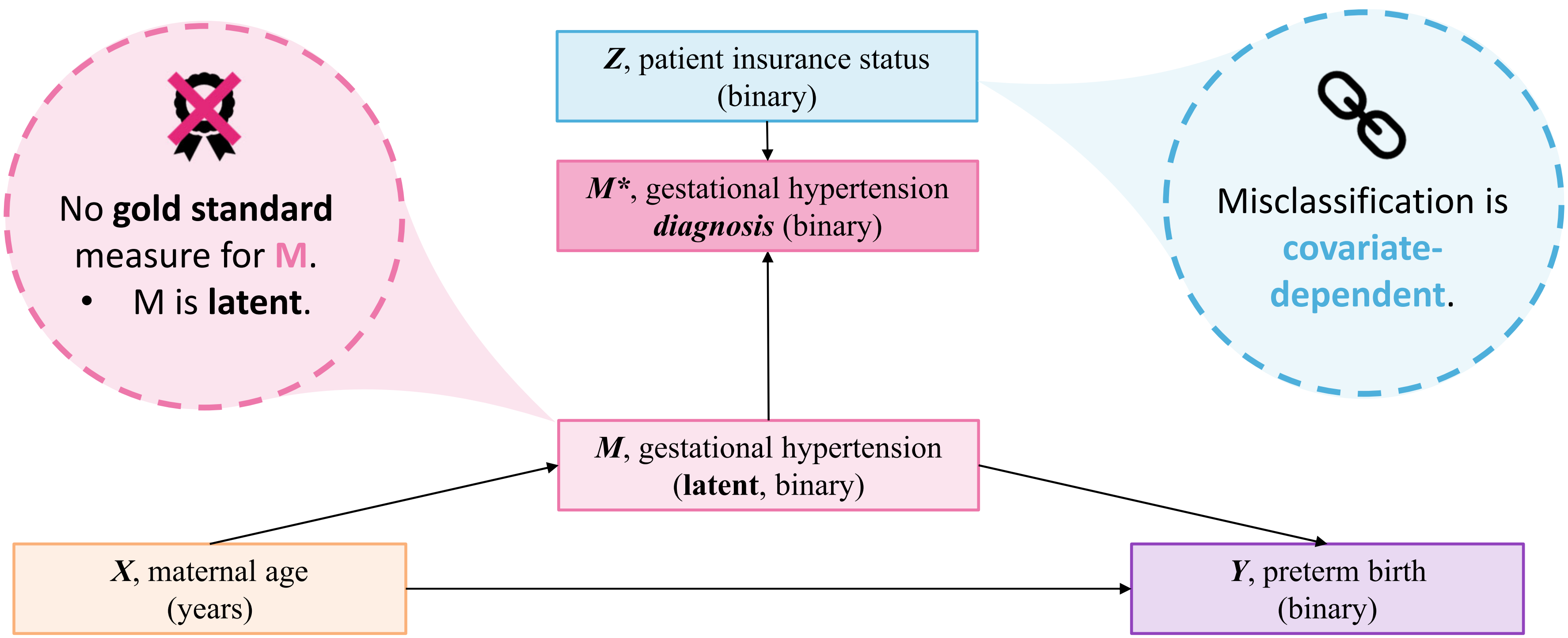
Z, patient insurance status
(binary)
*M**, gestational hypertension
diagnosis (binary)
M, gestational hypertension
(binary)
X, maternal age
(years)
Y, preterm birth
(binary)

Challenges



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

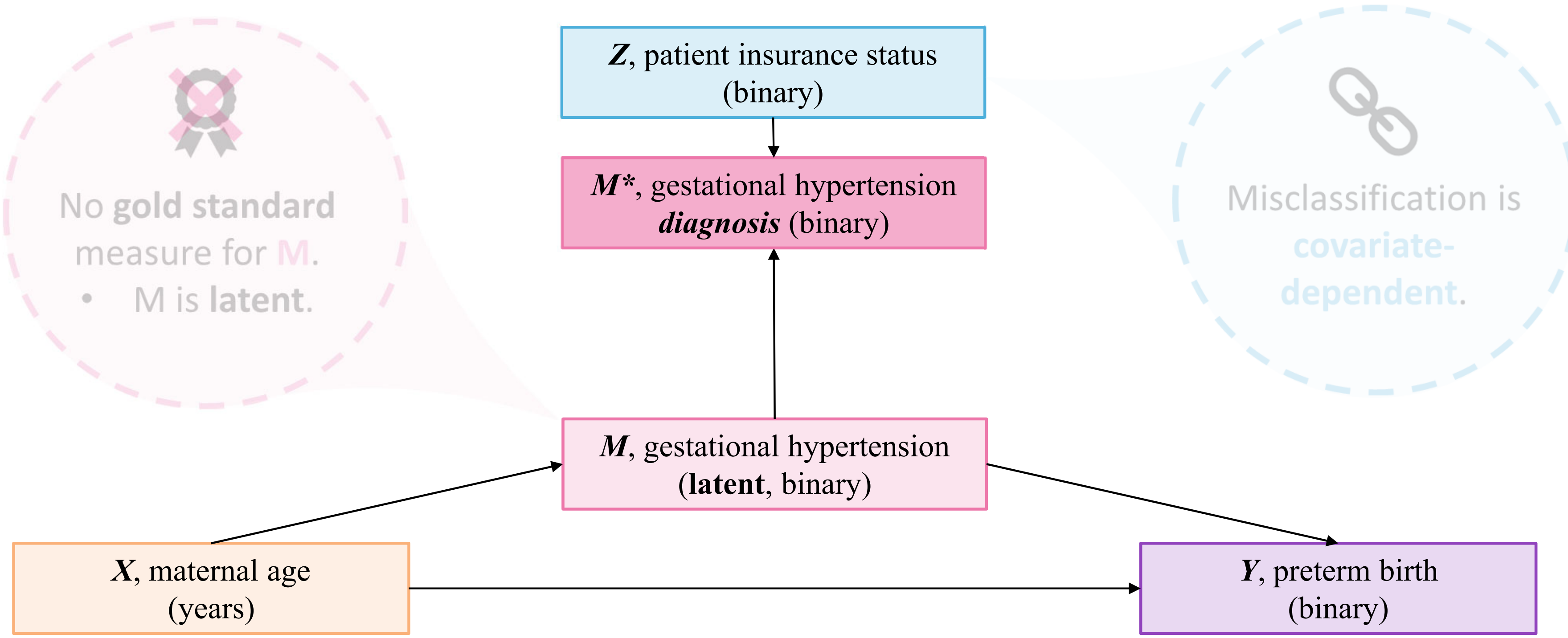
Challenges



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

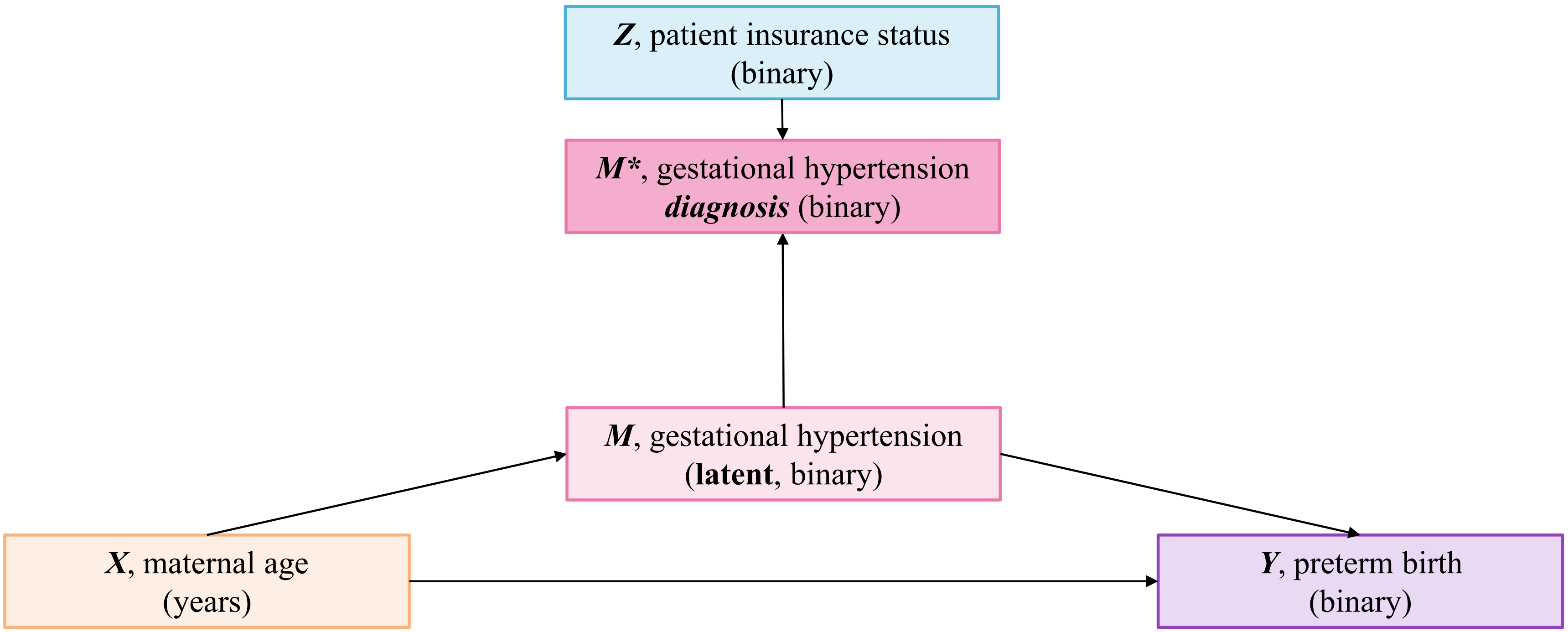
Ignoring misclassification in M* → Bias in parameter and effect estimates



Analysis Plan

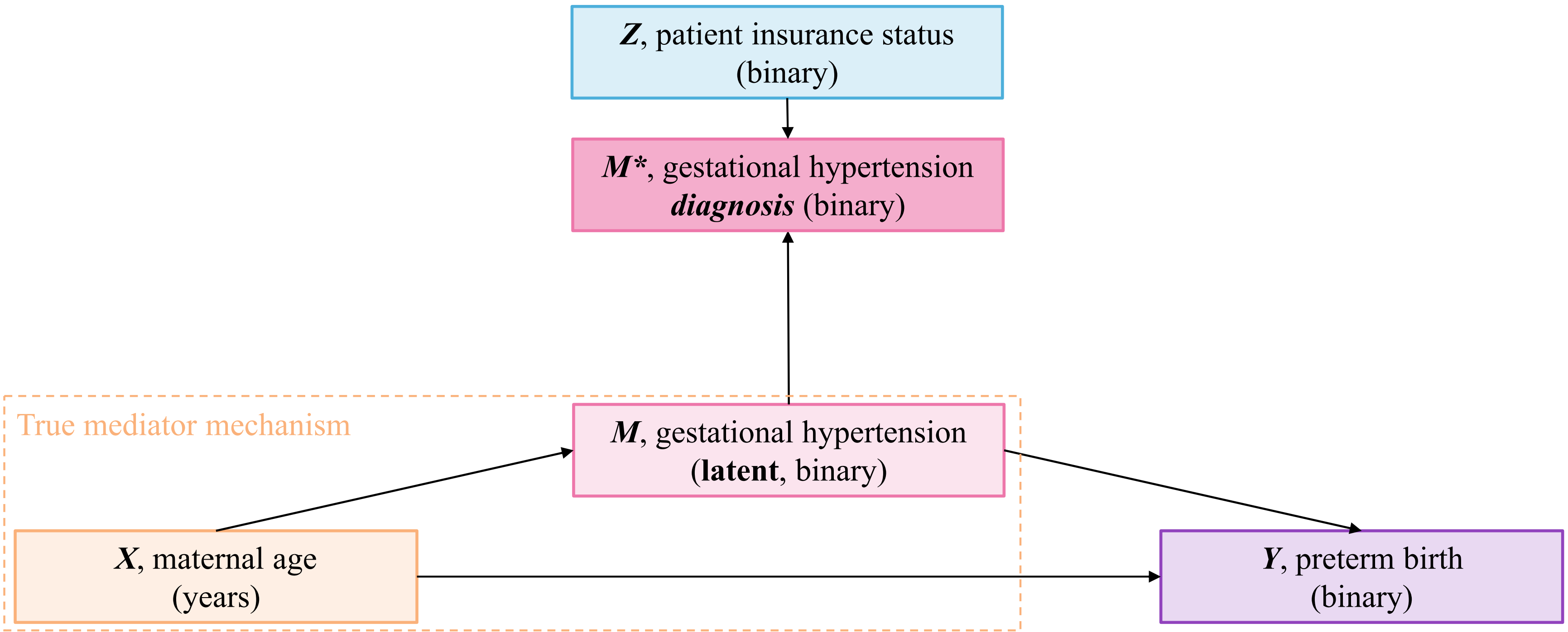
1. Define the **misclassification model**
2. Develop **estimation methods** for parameters of interest
3. Use parameter estimates to compute **(in)direct effects and misclassification rates**

Misclassification model



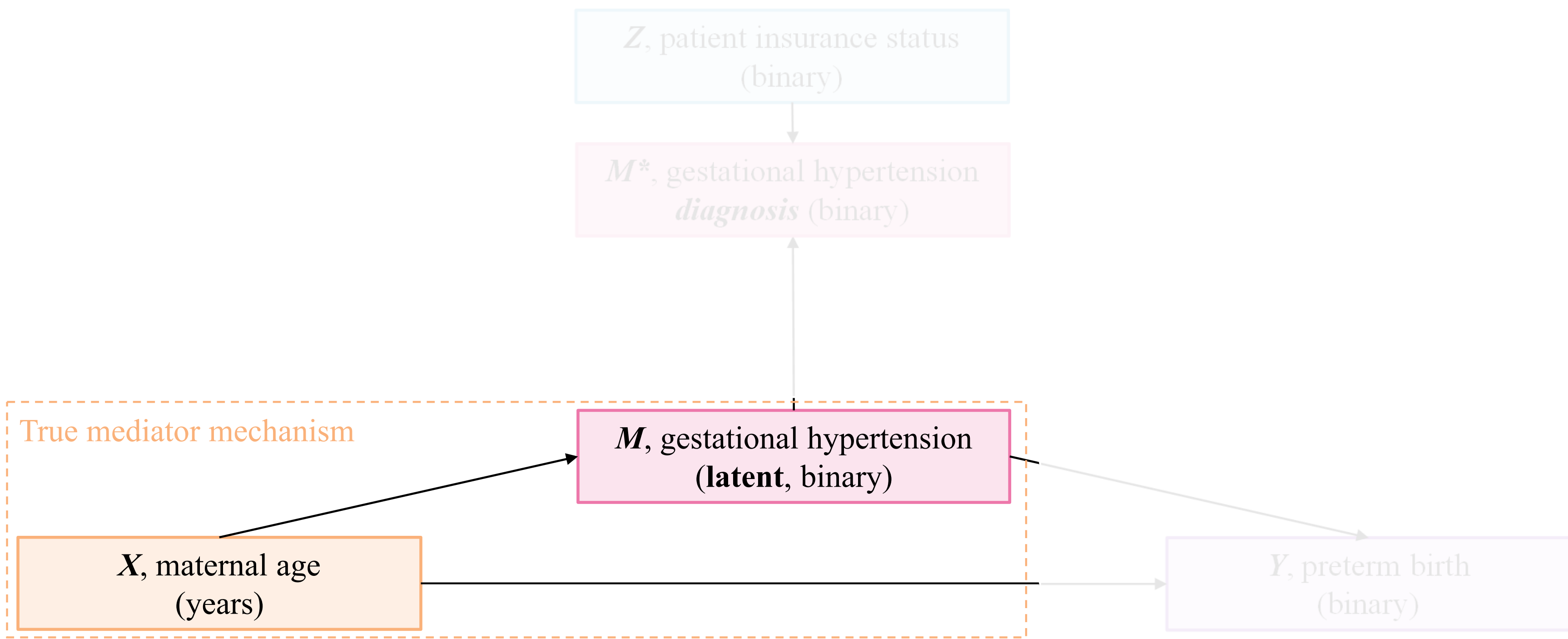
Misclassification model

1

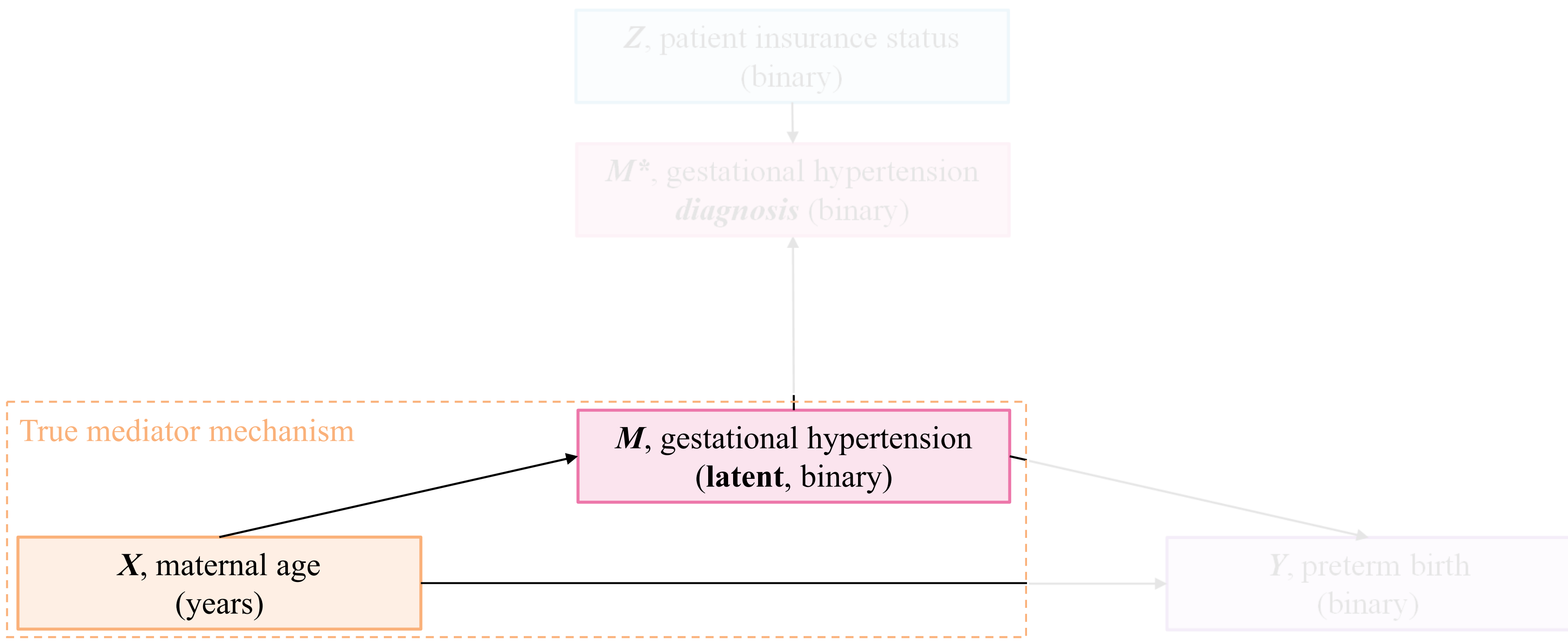


Misclassification model

1



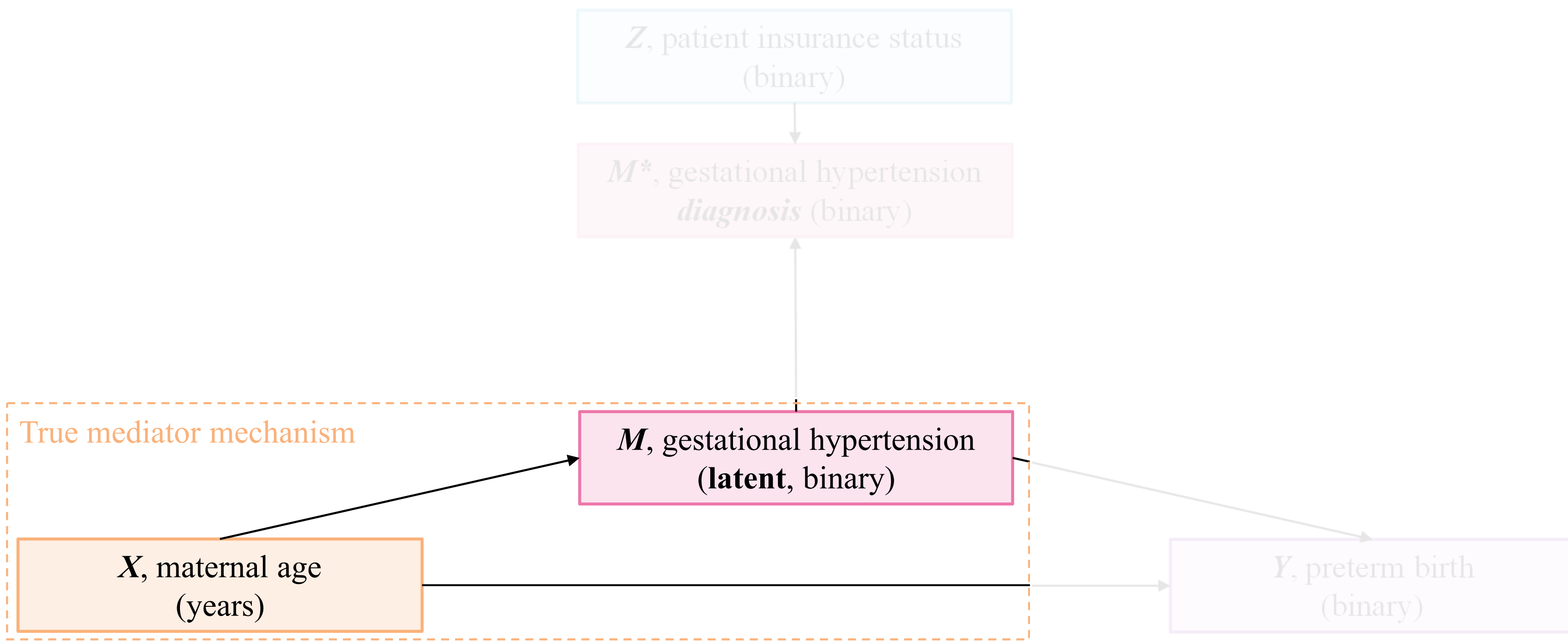
True mediator mechanism: $\text{logit}\{P(M = 1 | X, \boldsymbol{C}; \boldsymbol{\beta})\} = \beta_0 + \beta_X X + \boldsymbol{\beta}_C \boldsymbol{C}$ 1



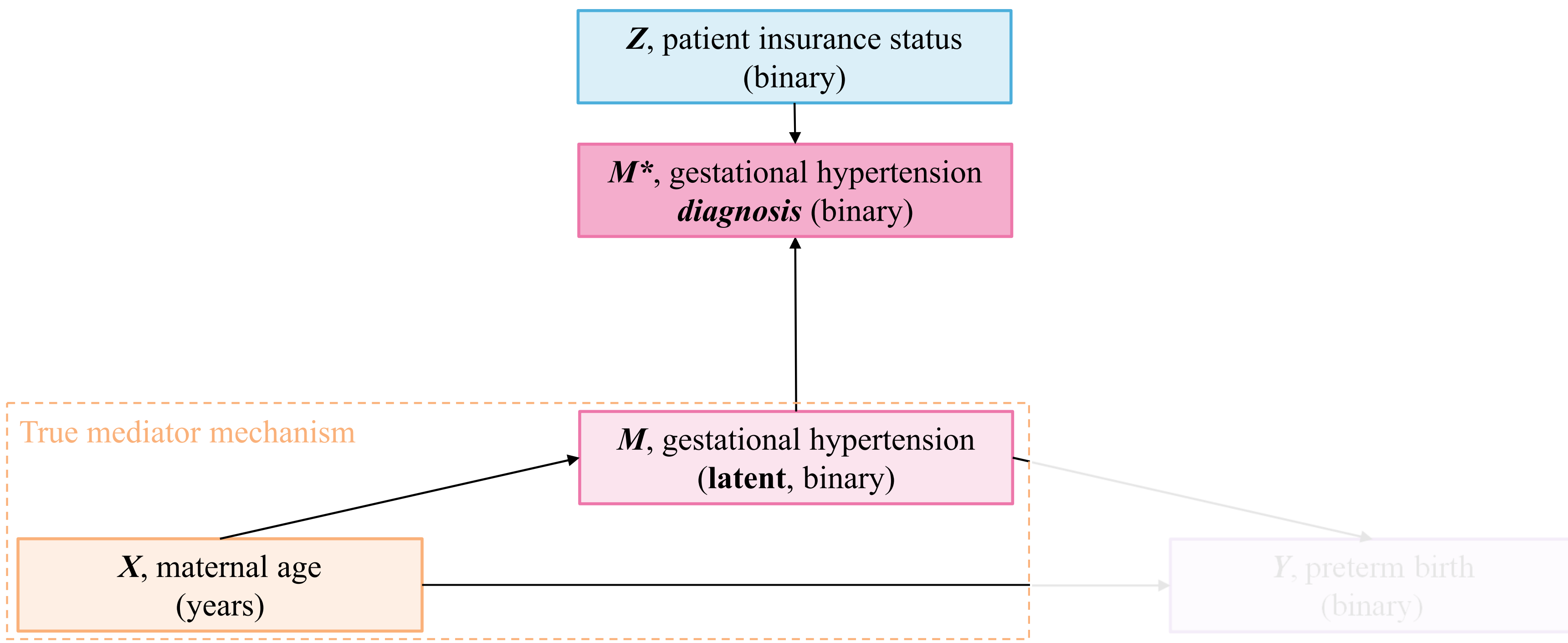
True mediator mechanism: $\operatorname{logit}\{P(M=1 \mid X, \boldsymbol{C}; \boldsymbol{\beta})\} = \beta_0 + \beta_X X + \beta_C \boldsymbol{C}$

1

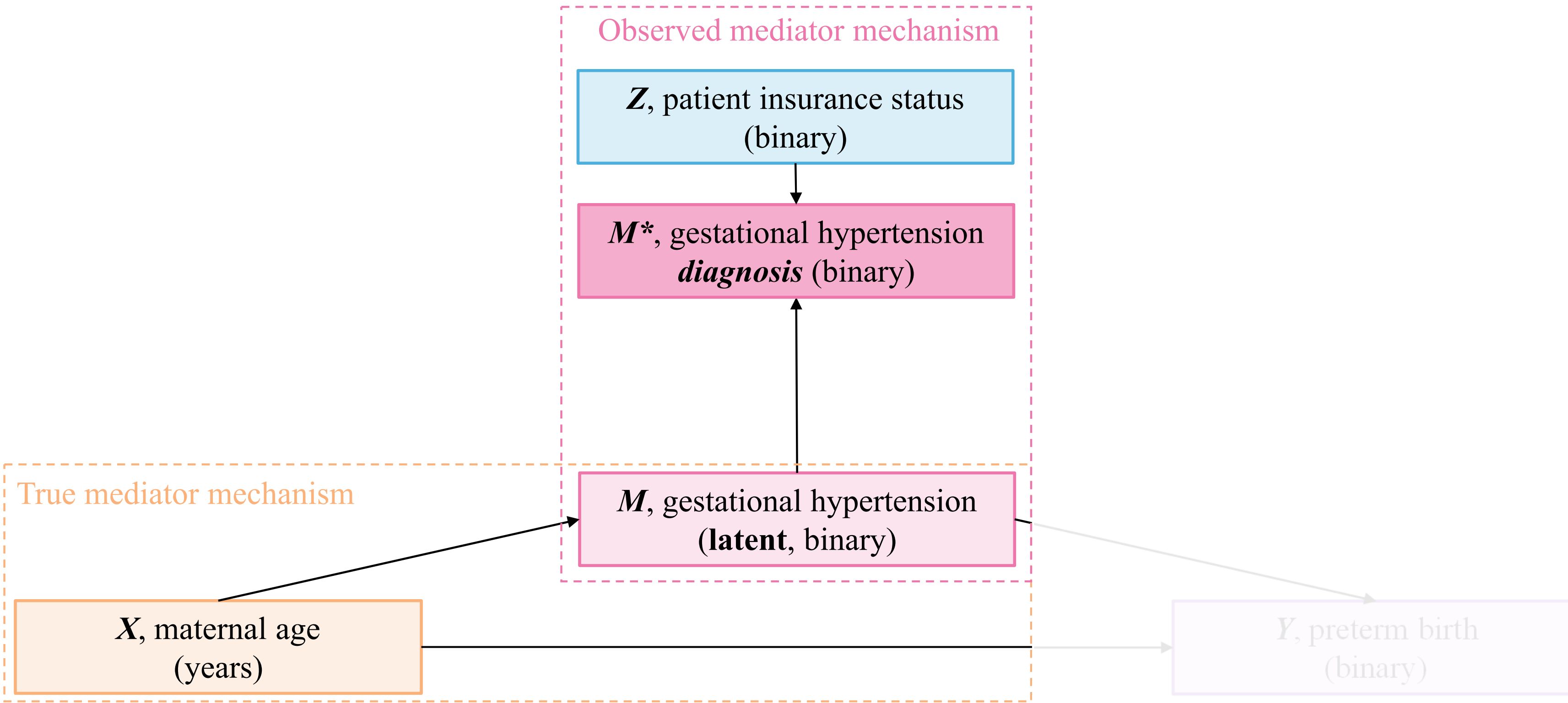
$\boldsymbol{C}$, covariates



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



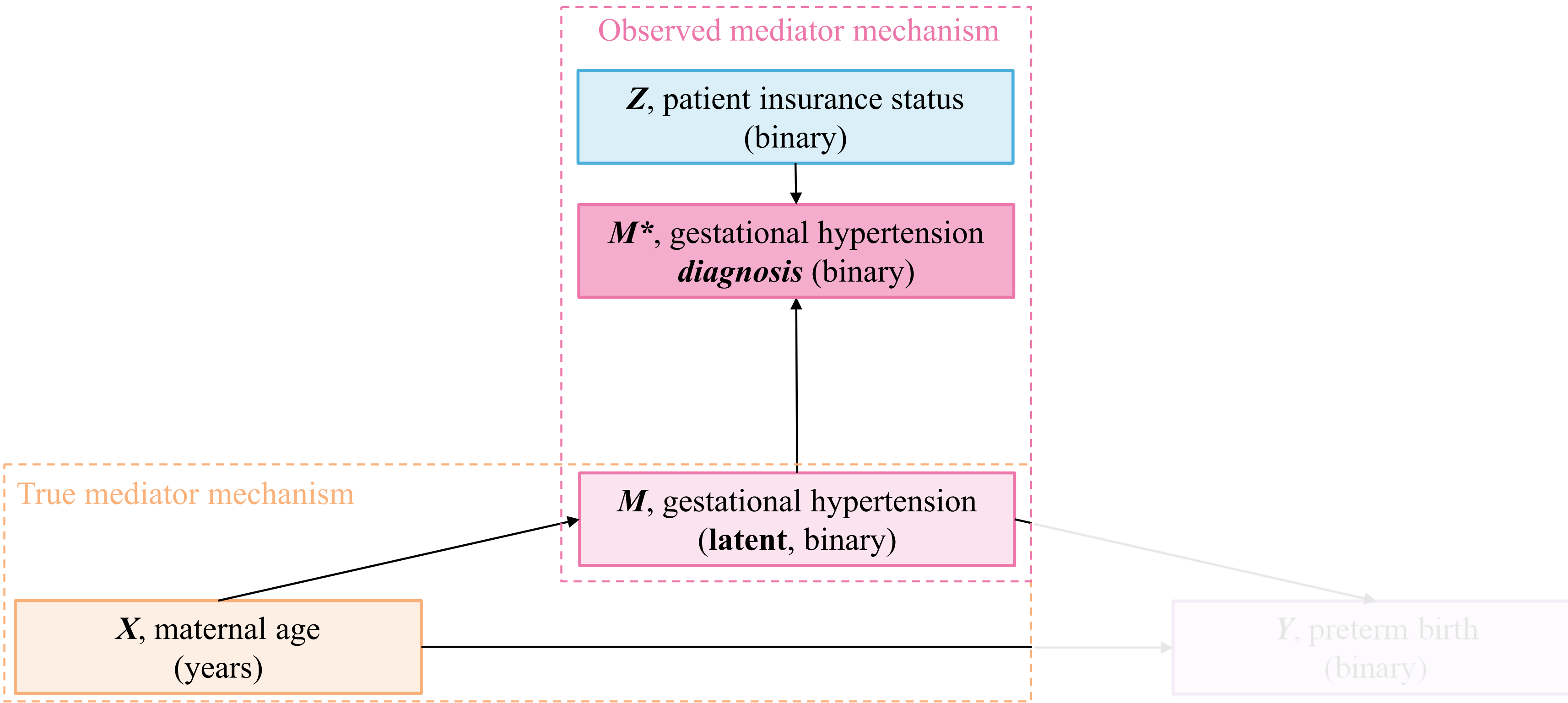
True mediator mechanism: $\text{logit}\{P(M = 1|X, \boldsymbol{C}; \boldsymbol{\beta})\} = \beta_0 + \beta_X X + \boldsymbol{\beta}_C \boldsymbol{C}$ 1



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

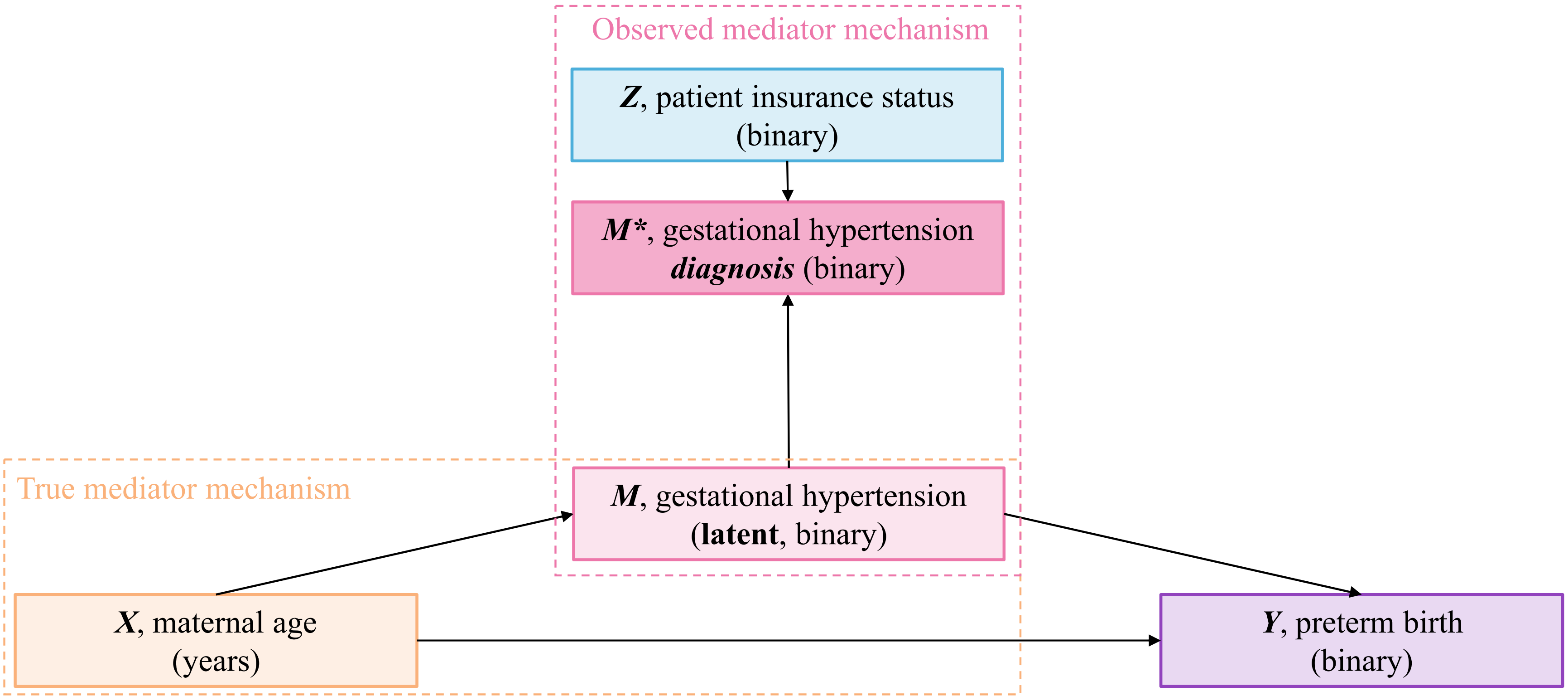
1

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



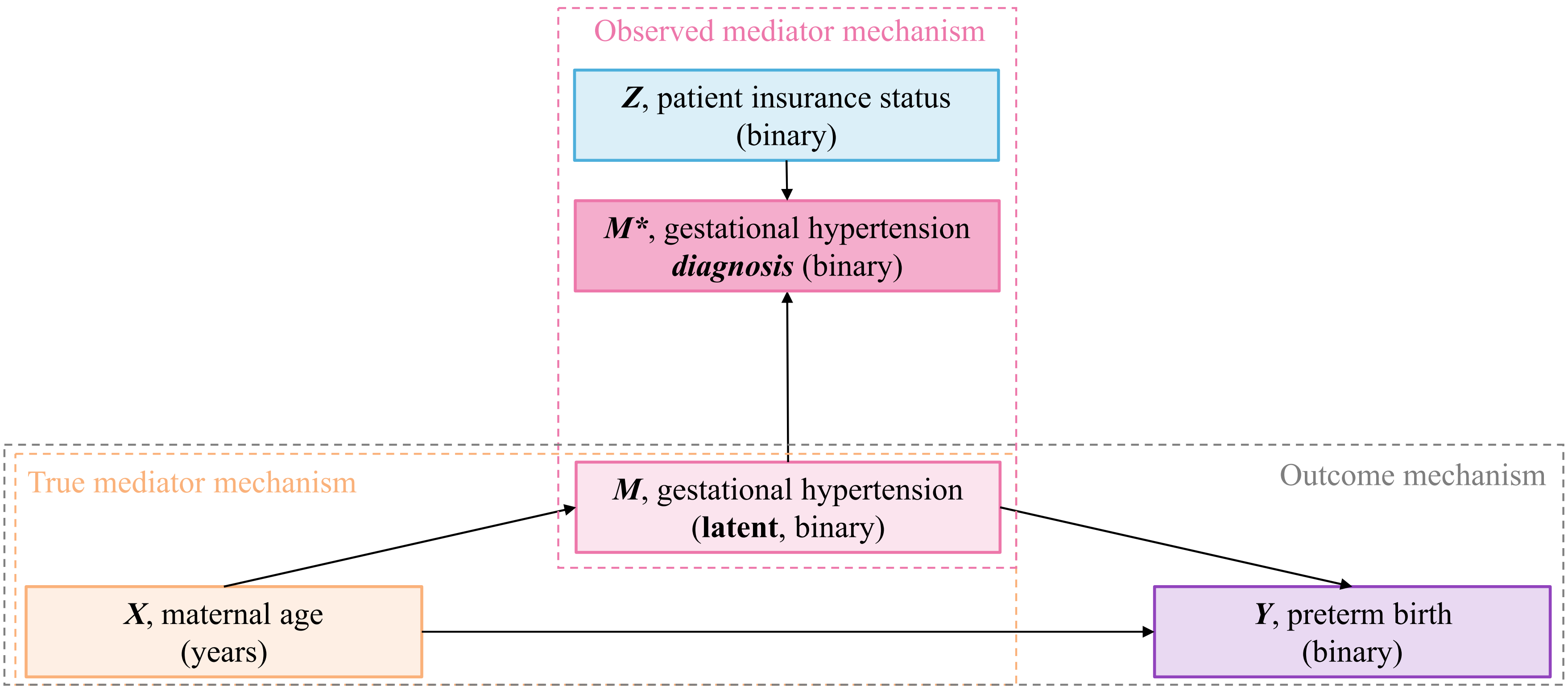
True mediator mechanism: $\text{logit}\{P(M=1|X,\boldsymbol{C};\boldsymbol{\beta})\} = \beta_0 + \beta_X X + \boldsymbol{\beta}_{\boldsymbol{C}}\boldsymbol{C}$ 1

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



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

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

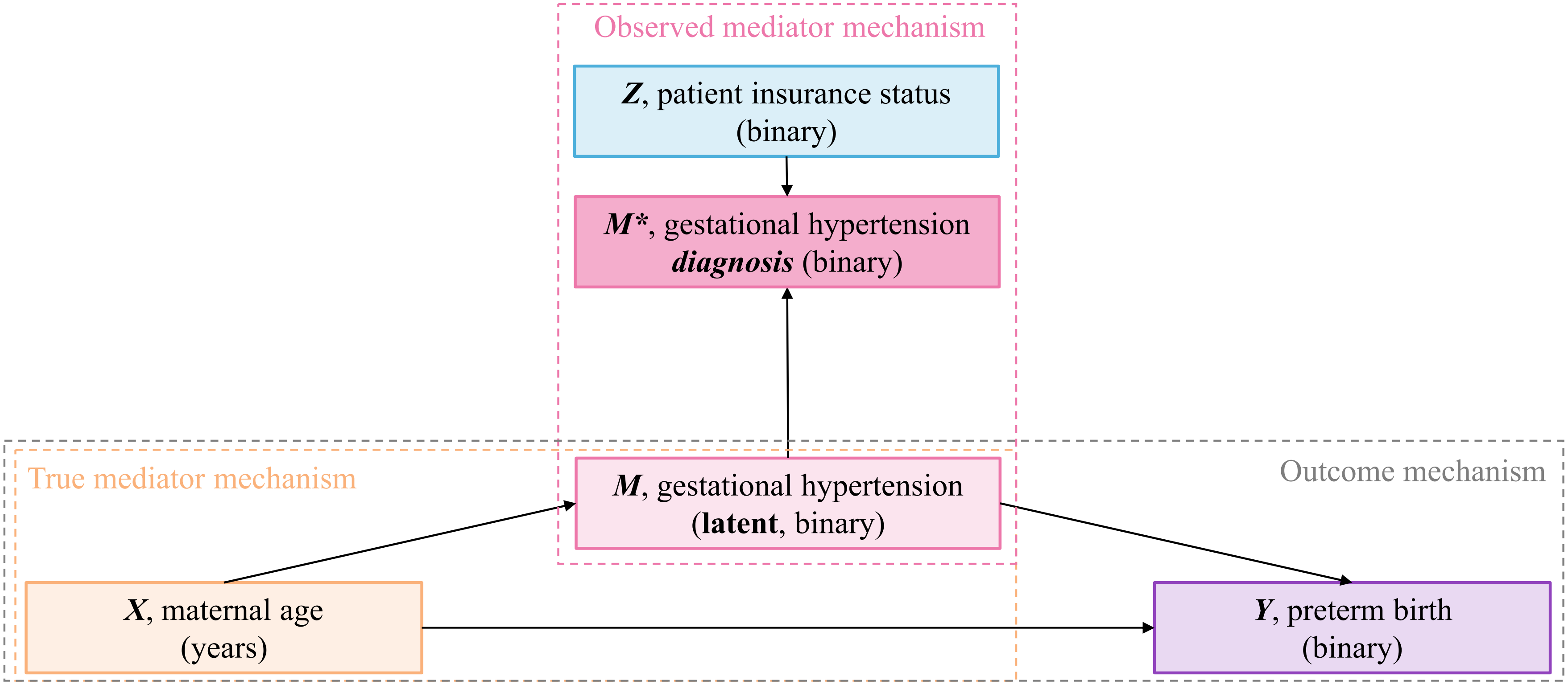


1

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

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

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

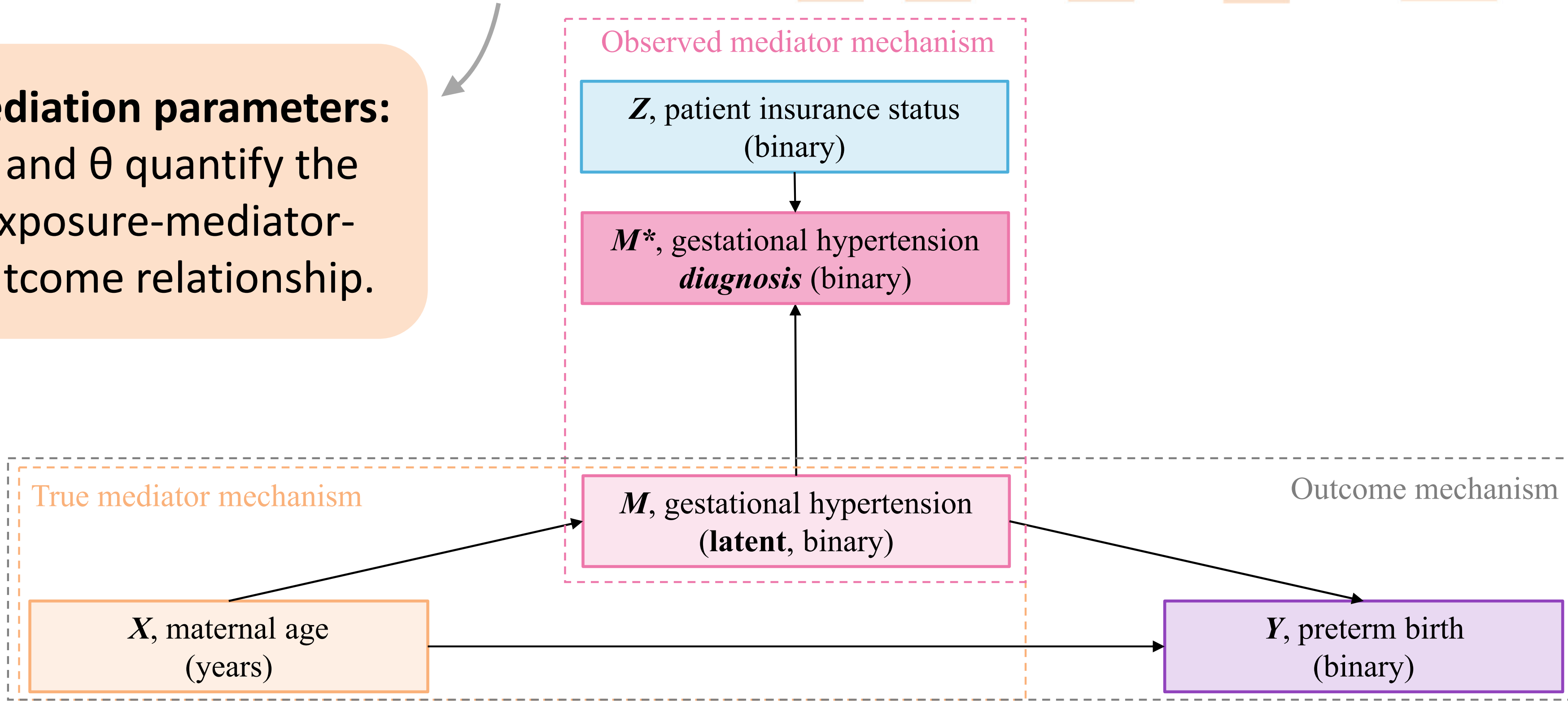


True mediator mechanism: $\text{logit}\{P(M=1|X,\boldsymbol{C};\boldsymbol{\beta})\} = \beta_0 + \beta_X X + \boldsymbol{\beta_C C}$ 1

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.



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

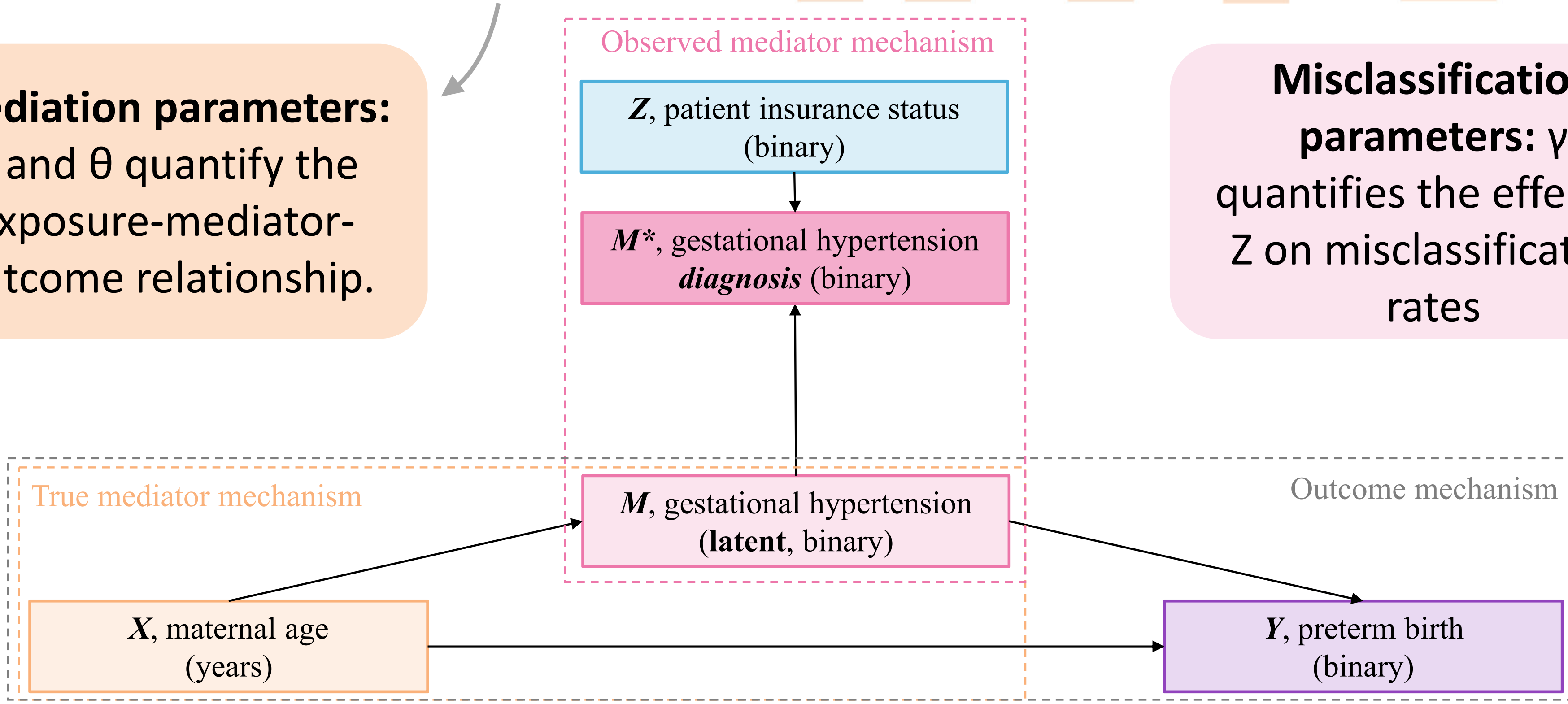
1

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

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

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

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

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

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

2

Estimation

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

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

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

#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

2

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

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

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

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

Complete data log-likelihood:

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

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

2

Estimation

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

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

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

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

Complete data log-likelihood:

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

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

Outcome

2

Estimation

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

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

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

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

Complete data log-likelihood:

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

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

2

Estimation

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

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

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

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

Complete data log-likelihood:

$$\ell_{complete}(\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\theta};X,\boldsymbol{C},\boldsymbol{Z},Y) = \sum_{i=1}^{N}\left[\underbrace{\ell_{Y|X,M,C}(\boldsymbol{\theta};X_i,M_i,\boldsymbol{C}_i,Y_i)}_{\text{Outcome}} + \underbrace{\sum_{j=1}^{2} \overset{\mathbb{I}(M_i = j)}{m_{ij}} \log\{\overset{P(M_i = j)}{\pi_{ij}}\}}_{\text{True mediator}} + \underbrace{\sum_{j=1}^{2}\sum_{\ell=1}^{2} m_{ij} \overset{\mathbb{I}(M_i^* = \ell)}{m^*_{i\ell}} \log\{\overset{P(M_i^* = \ell \mid M_i = j)}{\pi^*_{i\ell j}}\}}_{\text{Observed mediator}}\right]$$

2

Estimation

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

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

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

#1: OLS Correction

#2: Predictive value weighting

#3: An EM algorithm

Expectation Step ↔ **Maximization Step**

2

Estimation

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

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

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

#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**

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

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

Estimation

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

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

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

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

Expectation Step

Apply label switching correction from Webb and Wells (2023)

Maximization Step

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

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

2

Estimation

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

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

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

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

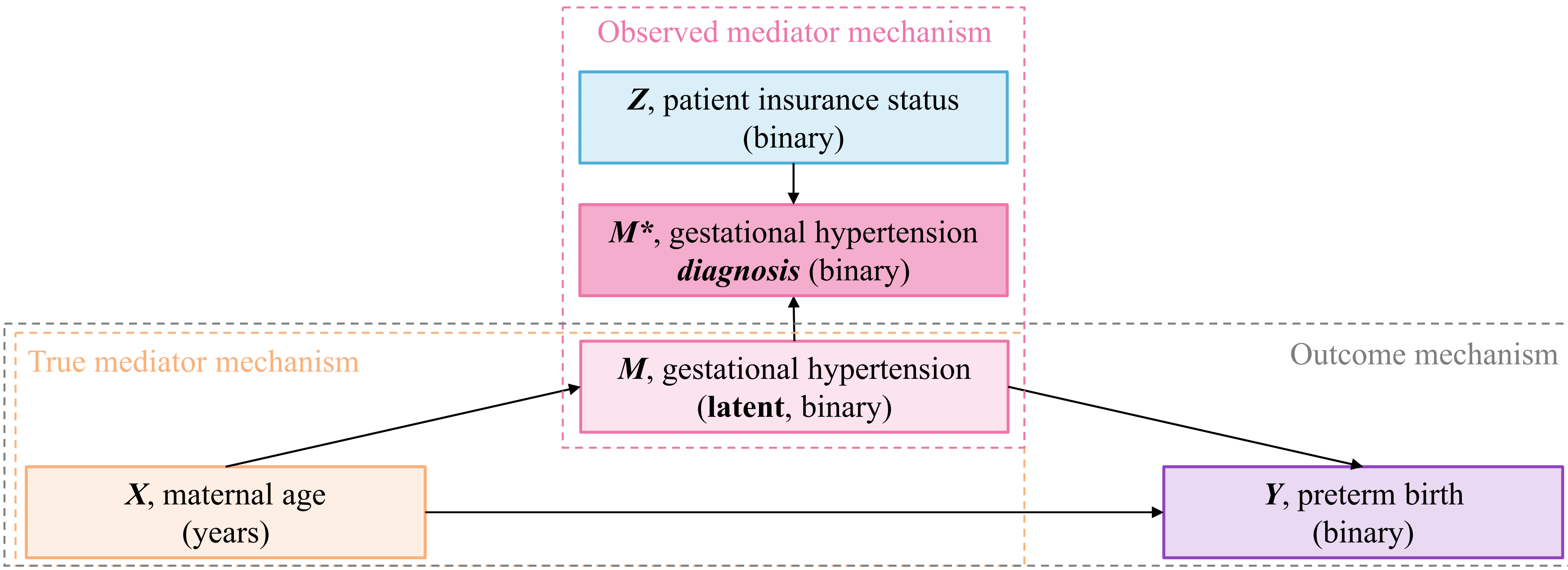
- Use the resulting bias-corrected parameter estimates to compute **(in)direct effects** for a change from **x̃ to x:**

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

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

Preterm birth study

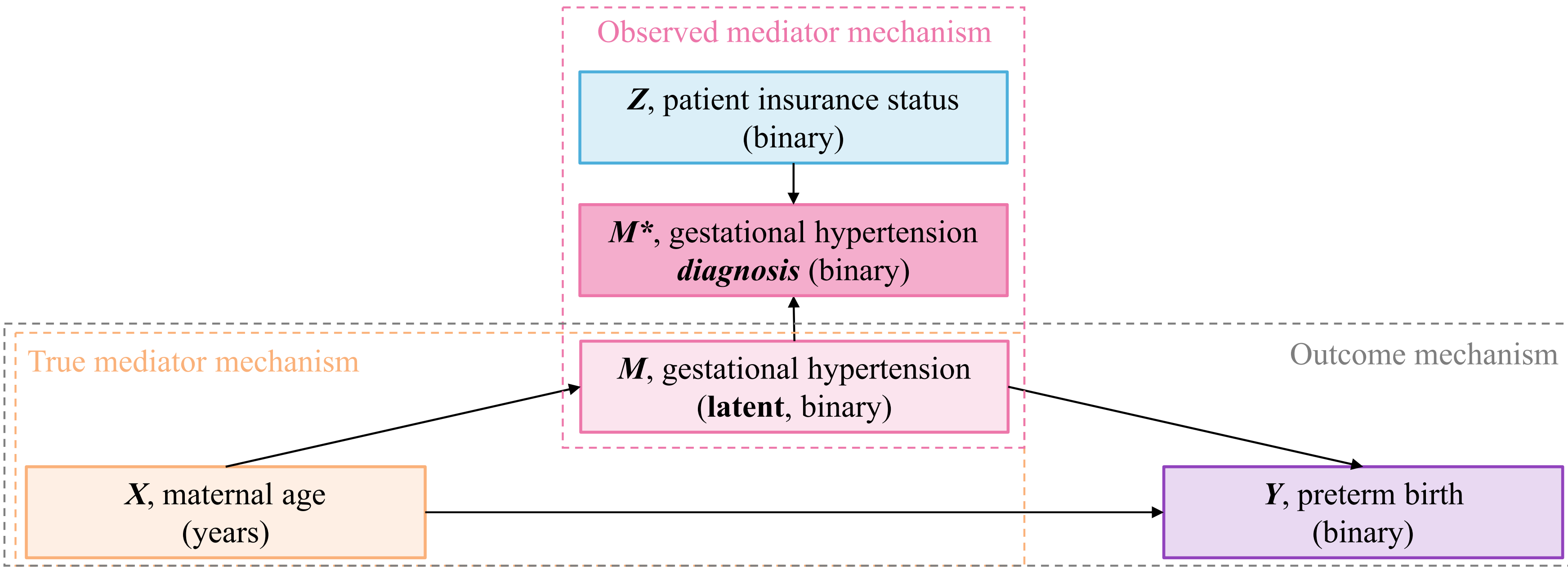
Does gestational hypertension mediate the association between maternal age and preterm birth, after accounting for potential misdiagnosis of gestational hypertension based on patient insurance status?



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**.

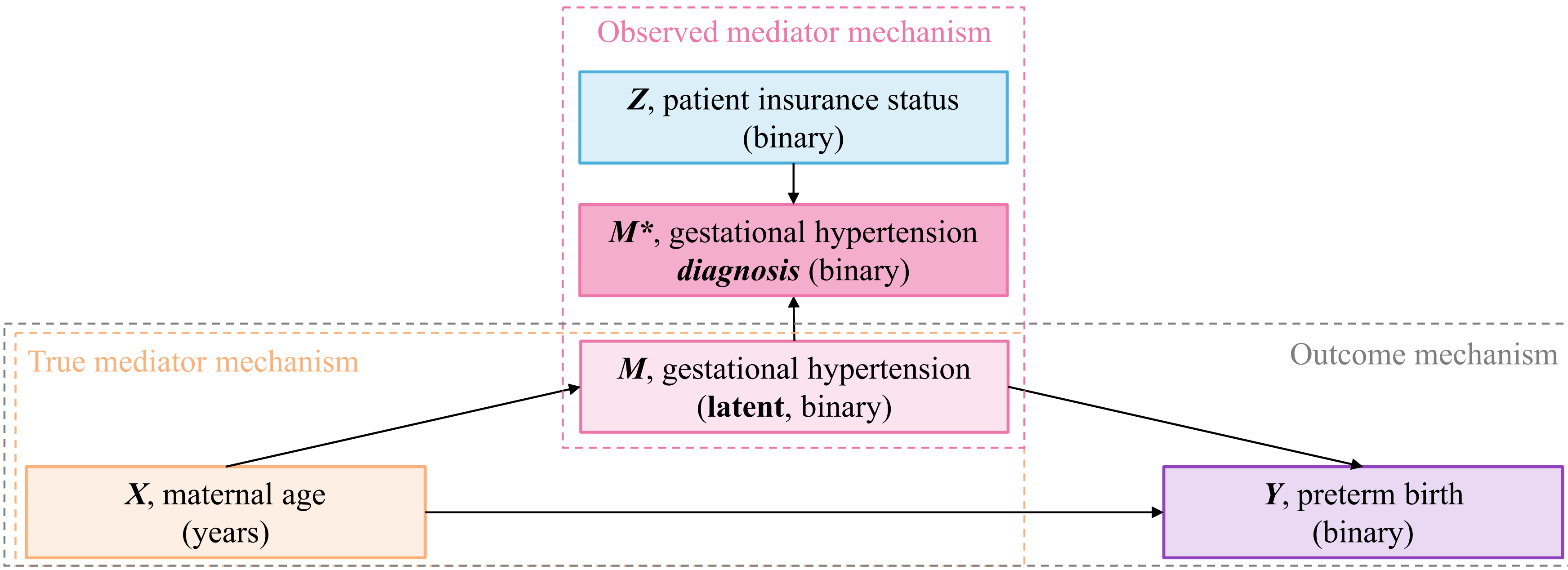


Preterm birth study

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**



3

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}				

3

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}				

3

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

3

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

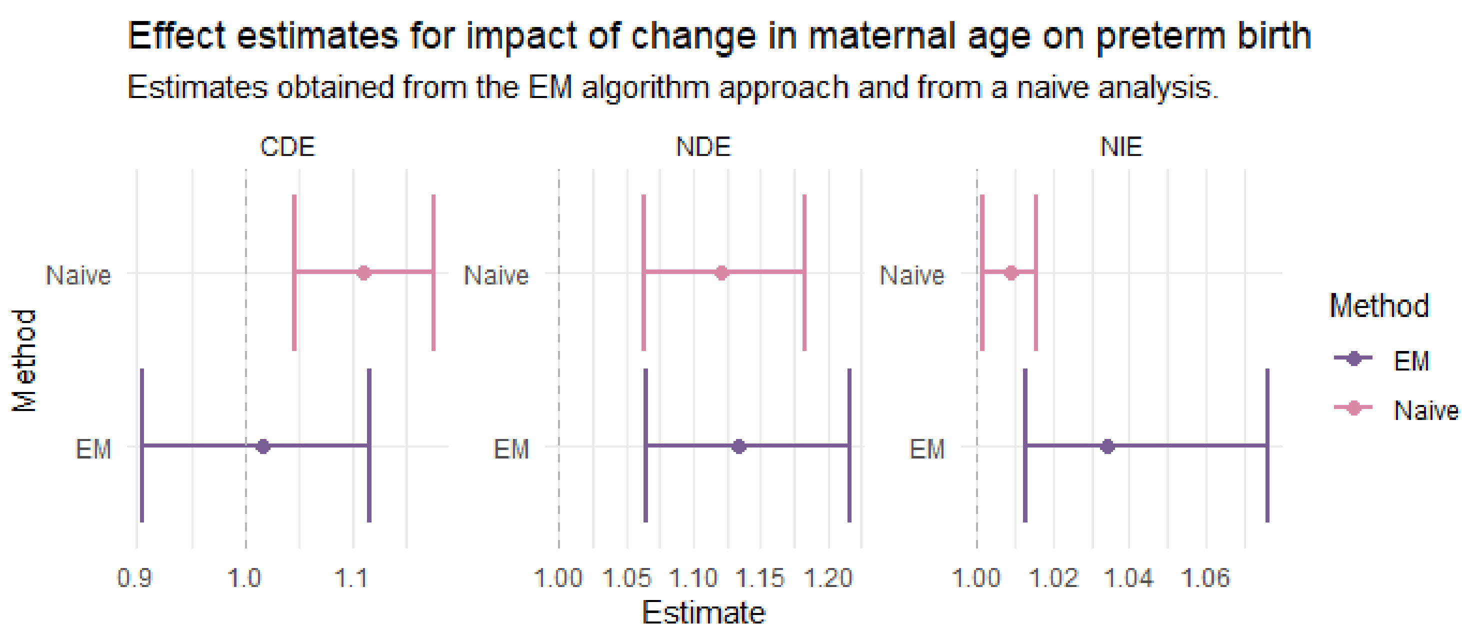
Use **β** and **θ** parameter estimates to compute **(in)direct effects**.

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.**

3

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**.

3

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%

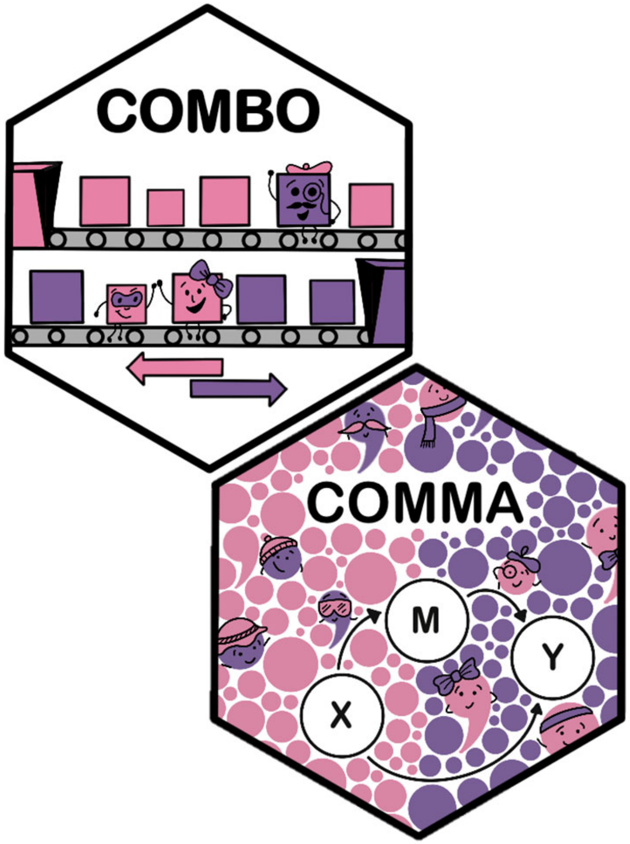
Key takeaways

- 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