Supplementary Methods

1. The calculation process of $Cov(f_{11D}+f_{12D}, f_{11D}+f_{21D})$, $Cov(f_{11d}+f_{12d}, f_{11d}+f_{21d})$ and $E(\hat{\beta}_D)$

From the nature of covariance, mathematical expectation definition and the nature of covariance, we can get

$$\begin{aligned}
&Cov(f_{11D}+f_{12D}, f_{11D}+f_{21D}) \\
&= E\{[\hat{f}_{11D}+\hat{f}_{12D}-E(f_{11D}+f_{12D})][\hat{f}_{11D}+\hat{f}_{21D}-E(f_{11D}+f_{21D})]\} \\
&= E\{[\hat{f}_{11D}+\hat{f}_{12D}-(f_{11D}+f_{12D})][\hat{f}_{11D}+\hat{f}_{21D}-(f_{11D}+f_{21D})]\} \\
&= E[(\hat{f}_{11D}+\hat{f}_{12D})(\hat{f}_{11D}+\hat{f}_{21D})]-(f_{11D}+f_{12D})\cdot(f_{11D}+f_{21D}) \\
&= E[\hat{f}_{11D}{}^2+\hat{f}_{12D}\cdot\hat{f}_{11D}+\hat{f}_{11D}\hat{f}_{21D}+\hat{f}_{12D}\hat{f}_{21D}]-(f_{11D}+f_{12D})\cdot(f_{11D}+f_{21D})
\end{aligned}, \quad \text{(a)}$$

$$E(\hat{f}_{11D}{}^2)=Var(\hat{f}_{11D})+E^2(\hat{f}_{11D})=\frac{f_{11D}(1-f_{11D})}{n_D}+f_{11D}{}^2, \quad \text{(b)}$$

$$\begin{aligned}
&E(\hat{f}_{12D}\cdot\hat{f}_{11D})=Cov(\hat{f}_{12D},\hat{f}_{11D})+E(\hat{f}_{12D})E(\hat{f}_{11D}) \\
&=-\frac{f_{12D}\cdot f_{11D}}{n_D}+f_{12D}\cdot f_{11D}
\end{aligned}, \quad \text{(c)}$$

$$\begin{aligned}
&E(\hat{f}_{11D}\cdot\hat{f}_{21D})=Cov(\hat{f}_{11D},\hat{f}_{21D})+E(\hat{f}_{11D})E(\hat{f}_{21D}) \\
&=-\frac{f_{11D}\cdot f_{21D}}{n_D}+f_{11D}\cdot f_{21D}
\end{aligned}, \quad \text{(d)}$$

and

$$\begin{aligned}
&E(\hat{f}_{12D}\cdot\hat{f}_{21D})=Cov(\hat{f}_{12D},\hat{f}_{21D})+E(\hat{f}_{12D})E(\hat{f}_{21D}) \\
&=-\frac{f_{12D}\cdot f_{21D}}{n_D}+f_{12D}\cdot f_{21D}
\end{aligned}. \quad \text{(e)}$$

From Equation a, Equation b, Equation c, Equation d and Equation e in the supplementary methods, we can get

$$\begin{aligned}
&Cov(f_{11D}+f_{12D},f_{11D}+f_{21D})\\
&=E[\hat{f}_{11D}{}^{2}+\hat{f}_{12D}\cdot\hat{f}_{11D}+\hat{f}_{11D}\cdot\hat{f}_{21D}+\hat{f}_{12D}\cdot\hat{f}_{21D}]-(f_{11D}+f_{12D})\cdot(f_{11D}+f_{21D})\\
&=\frac{f_{11D}\cdot(1-f_{11D})}{n_D}+f_{11D}{}^{2}+(-\frac{f_{12D}\cdot f_{11D}}{n_D}+f_{12D}\cdot f_{11D})+\ldots\\
&(-\frac{f_{11D}\cdot f_{21D}}{n_D}+f_{11D}\cdot f_{21D})+(-\frac{f_{12D}\cdot f_{21D}}{n_D}+f_{12D}\cdot f_{21D})-(f_{11D}+f_{12D})\cdot(f_{11D}+f_{21D})\\
&=\frac{f_{11D}\cdot(1-f_{11D})-f_{12D}\cdot f_{11D}-f_{11D}\cdot f_{21D}-f_{12D}\cdot f_{21D}}{n_D}+\ldots\\
&f_{11D}{}^{2}+f_{12D}\cdot f_{11D}+f_{11D}\cdot f_{21D}+f_{12D}\cdot f_{21D}-(f_{11D}+f_{12D})\cdot(f_{11D}+f_{21D})\\
&=\frac{f_{11D}\cdot(1-f_{11D})-f_{12D}\cdot f_{11D}-f_{11D}\cdot f_{21D}-f_{12D}\cdot f_{21D}}{n_D}\\
&=\frac{f_{11D}\cdot(1-f_{11D}-f_{12D}-f_{21D})-f_{12D}\cdot f_{21D}}{n_D}\\
&=\frac{f_{11D}\cdot f_{22D}-f_{12D}\cdot f_{21D}}{n_D}
\end{aligned} \quad . \qquad (f)$$

From

$$\begin{aligned}
&f_{11D}-(f_{11D}+f_{12D})\cdot(f_{11D}+f_{21D})\\
&=f_{11D}-(f_{11D}{}^{2}+f_{11D}\cdot f_{21D}+f_{12D}\cdot f_{11D}+f_{12D}\cdot f_{21D})\\
&=f_{11D}\cdot(1-f_{11D}-f_{21D}-f_{12D})-f_{12D}\cdot f_{21D}\\
&=f_{11D}\cdot f_{22D}-f_{12D}\cdot f_{21D}
\end{aligned}, \qquad (g)$$

we can get

$$Cov(f_{11D}+f_{12D},f_{11D}+f_{21D})=\frac{f_{11D}-(f_{11D}+f_{12D})\cdot(f_{11D}+f_{21D})}{n_D}. \qquad (h)$$

In the same way, $Cov(f_{11\mathrm{d}}+f_{12d},f_{11d}+f_{21d})$ can be calculated as

$$Cov(f_{11d}+f_{12d},f_{11d}+f_{21d})=\frac{f_{11d}-(f_{11d}+f_{12d})\cdot(f_{11d}+f_{21d})}{n_d}. \qquad (i)$$

From Equation 5, Equation 6, Equation 7 in the main body and the nature of the mathematical expectation, $E(\hat{\beta}_D)$ can be calculated as

$$\begin{aligned}
&E(\hat{\beta}_D) = E(\hat{f}_{11D}) - E[(\hat{f}_{11D} + \hat{f}_{12D}) \cdot (\hat{f}_{11D} + \hat{f}_{21D})] \\
&= f_{11D} - [E(\hat{f}_{11D} + \hat{f}_{12D}) \cdot E(\hat{f}_{11D} + \hat{f}_{21D}) + Cov(f_{11D} + f_{12D}, f_{11D} + f_{21D})] \\
&= f_{11D} - [(f_{11D} + f_{12D}) \cdot (f_{11D} + f_{21D}) + Cov(f_{11D} + f_{12D}, f_{11D} + f_{21D})] \\
&= f_{11D} \cdot f_{22D} - f_{12D} \cdot f_{21D} - Cov(f_{11D} + f_{12D}, f_{11D} + f_{21D})
\end{aligned} \quad . \tag{j}$$

2. The calculation process of $f_{11D}, f_{12D}, f_{22D}, f_{21D}, f_{11d}, f_{12d}, f_{22d}$ and f_{21d}

As X1, X2, M1 and M2 have two possible values: 0 and 1 respectively, f_{11D} can be calculated using Equation 7 in the main body and Multiplicative Theorem of Conditional Probability as follows

$$\begin{aligned}
& f_{11D} = P(M_1 = 1, M_2 = 1 \mid D) \\
& = P(M_1 = 1, M_2 = 1, X_1 = 1, X_2 = 1 \mid D) + P(M_1 = 1, M_2 = 1, X_1 = 0, X_2 = 1 \mid D) + ... \\
& P(M_1 = 1, M_2 = 1, X_1 = 1, X_2 = 0 \mid D) + P(M_1 = 1, M_2 = 1, X_1 = 0, X_2 = 0 \mid D) \\
& = P(M_1 = 1, M_2 = 1 \mid X_1 = 1, X_2 = 1) \cdot P(X_1 = 1, X_2 = 1 \mid D) + ... \\
& P(M_1 = 1, M_2 = 1 \mid X_1 = 0, X_2 = 1) \cdot P(X_1 = 0, X_2 = 1 \mid D) + ... \\
& P(M_1 = 1, M_2 = 1 \mid X_1 = 1, X_2 = 0) \cdot P(X_1 = 1, X_2 = 0 \mid D) + ... \\
& P(M_1 = 1, M_2 = 1 \mid X_1 = 0, X_2 = 0) \cdot P(X_1 = 0, X_2 = 0 \mid D)
\end{aligned} \quad . \quad \text{(k)}$$

Because X1 and M2 are unlinked and X2 and M1 are unlinked, we can conclude that

$$\begin{aligned}
& P(M_1 = 1, M_2 = 1 \mid X_1 = 1, X_2 = 1) = P(M_1 = 1 \mid X_1 = 1) P(M_2 = 1 \mid X_2 = 1) \\
& P(M_1 = 1, M_2 = 1 \mid X_1 = 0, X_2 = 1) = P(M_1 = 1 \mid X_1 = 0) P(M_2 = 1 \mid X_2 = 1) \\
& P(M_1 = 1, M_2 = 1 \mid X_1 = 1, X_2 = 0) = P(M_1 = 1 \mid X_1 = 1) P(M_2 = 1 \mid X_2 = 0) \\
& P(M_1 = 1, M_2 = 1 \mid X_1 = 0, X_2 = 0) = P(M_1 = 1 \mid X_1 = 0) P(M_2 = 1 \mid X_2 = 0)
\end{aligned} \quad . \quad \text{(l)}$$

According to Bayesian conditional probability formula and Equation 6 in the main body, eight conditional probabilities can be calculated as

$$P(M_1=1|X_1=1)=\frac{P(M_1=1,X_1=1)}{P(X_1=1)}=\frac{P(M_1=1)\cdot P(X_1=1)+LD_{X1,M1}}{P(X_1=1)}$$
$$P(M_2=1|X_2=1)=\frac{P(M_2=1,X_2=1)}{P(X_2=1)}=\frac{P(M_2=1)\cdot P(X_2=1)+LD_{X2,M2}}{P(X_2=1)}$$
$$P(M_1=1|X_1=0)=\frac{P(M_1=1,X_1=0)}{P(X_1=0)}=\frac{P(M_1=1)\cdot[1-P(X_1=1)]+LD_{X1,M1}}{1-P(X_1=1)}$$
$$P(M_2=1|X_2=0)=\frac{P(M_2=1,X_2=0)}{P(X_2=0)}=\frac{P(M_2=1)\cdot[1-P(X_2=1)]+LD_{X2,M2}}{1-P(X_2=1)}$$
$$P(M_1=0|X_1=1)=\frac{P(M_1=0,X_1=1)}{P(X_1=1)}=\frac{[1-P(M_1=1)]\cdot P(X_1=1)+LD_{X1,M1}}{P(X_1=1)}$$
$$P(M_2=0|X_2=1)=\frac{P(M_2=0,X_2=1)}{P(X_2=1)}=\frac{[1-P(M_2=1)]\cdot P(X_2=1)+LD_{X2,M2}}{P(X_2=1)}$$
$$P(M_1=0|X_1=0)=\frac{P(M_1=0,X_1=0)}{P(X_1=0)}=\frac{[1-P(M_1=1)]\cdot[1-P(X_1=1)]+LD_{X1,M1}}{1-P(X_1=1)}$$
$$P(M_2=0|X_2=0)=\frac{P(M_2=0,X_2=0)}{P(X_2=0)}=\frac{[1-P(M_2=1)]\cdot[1-P(X_2=1)]+LD_{X2,M2}}{1-P(X_2=1)}. \quad \text{(m)}$$

Thus, $P(M_1=1,M_2=1|X_1=1,X_2=1)$, $P(M_1=1,M_2=1|X_1=0,X_2=1)$, $P(M_1=1,M_2=1|X_1=1,X_2=0)$ and $P(M_1=1,M_2=1|X_1=0,X_2=1)$ can be calculated according to Equation m in the supplementary methods.

Assume that $P(X_1=1,X_2=1|D)=(1+K)P(X_1=1|D)P(X_2=1|D)$, from Equation 6 in the main body we can conclude that

$$\begin{aligned}
&P(X_1=1,X_2=0|D)=P(X_1=1|D)-P(X_1=1,X_2=1|D)\\
&P(X_1=0,X_2=1|D)=P(X_2=1|D)-P(X_1=1,X_2=1|D)\\
&P(X_1=0,X_2=0|D)=1-P(X_1=1,X_2=1|D)-P(X_1=1,X_2=0|D)-P(X_1=0,X_2=1|D)
\end{aligned} \quad . \quad \text{(n)}$$

From Bayesian conditional probability formula, Equation 6 and Equation 10 in the main body we can conclude that

$$
\begin{aligned}
RR_{X1} &= \frac{P(D \mid X_1 = 1)}{P(D \mid X_1 = 0)} \\
&= \frac{P(D, X_1 = 1) / P(X_1 = 1)}{P(D, X_1 = 0) / P(X_1 = 0)} \\
&= \frac{P(D, X_1 = 1) \cdot P(X_1 = 0)}{P(D, X_1 = 0) \cdot P(X_1 = 1)} \\
&= \frac{P(X_1 = 1 \mid D) \cdot P(X_1 = 0)}{[1 - P(X_1 = 1 \mid D)] \cdot P(X_1 = 1)}
\end{aligned} \quad , \quad \text{(o)}
$$

$$
\begin{aligned}
& RR_{X1} \cdot P(X_1 = 1) \cdot [1 - P(X_1 = 1 \mid D)] = P(X_1 = 0) \cdot P(X_1 = 1 \mid D) \\
& RR_{X1} \cdot P(X_1 = 1) = RR_{X1} \cdot P(X_1 = 1) \cdot P(X_1 = 1 \mid D) + P(X_1 = 0) \cdot P(X_1 = 1 \mid D) \\
& RR_{X1} \cdot P(X_1 = 1) = [RR_{X1} \cdot P(X_1 = 1) + P(X_1 = 0)] \cdot P(X_1 = 1 \mid D)
\end{aligned}
$$

$$
P(X_1 = 1 \mid D) = \frac{RR_{X1} \cdot P(X_1 = 1)}{RR_{X1} \cdot P(X_1 = 1) + P(X_1 = 0)} \quad , \quad \text{(p)}
$$

$$
P(X_2 = 1 \mid D) = \frac{RR_{X2} \cdot P(X_2 = 1)}{RR_{X2} \cdot P(X_2 = 1) + P(X_2 = 0)} \, . \quad \text{(q)}
$$

According to Equation n, Equation p and Equation q, $P(X_1 = 1, X_2 = 1 \mid D)$, $P(X_1 = 1, X_2 = 0 \mid D)$, $P(X_1 = 0, X_2 = 1 \mid D)$ and $P(X_1 = 0, X_2 = 0 \mid D)$ can be calculated using known parameters as follows

$$
\begin{aligned}
& P(X_1 = 1, X_2 = 1 \mid D) = (1+K) \cdot P(X_1 = 1 \mid D) \cdot P(X_2 = 1 \mid D) \\
& P(X_1 = 1, X_2 = 0 \mid D) = P(X_1 = 1 \mid D) - (1+K) \cdot P(X_1 = 1 \mid D) \cdot P(X_2 = 1 \mid D) \\
& P(X_1 = 0, X_2 = 1 \mid D) = P(X_2 = 1 \mid D) - (1+K) \cdot P(X_1 = 1 \mid D) \cdot P(X_2 = 1 \mid D) \\
& P(X_1 = 0, X_2 = 0 \mid D) = 1 - P(X_1 = 1 \mid D) - P(X_2 = 1 \mid D) + (1+K) \cdot P(X_1 = 1 \mid D) \cdot P(X_2 = 1 \mid D)
\end{aligned} \quad . \text{(r)}
$$

Thus, f_{11D} can be calculated from Equation k, Equation m and Equation r. $f_{12D}, f_{22D}, f_{21D}, f_{11d}, f_{12d}, f_{22d}$ and f_{21d} can be calculated in the same way.