	5%	95%
$\text{DDE}_{\text{exposure}}$	-0.05	0.01
$\text{PCB}_{\text{exposure}}$	-2.90	-0.81
$\text{DDE}_{\text{exposure}}$*white	-0.12	0.02
$\text{PCB}_{\text{exposure}}$*white	0.07	3.38

	5%	95%
$\text{DDE}_{\text{exposure}}$	-0.05	0.01
$\text{PCB}_{\text{exposure}}$	-2.96	-0.87
$\text{DDE}_{\text{exposure}}$*white	-0.13	0.01
$\text{PCB}_{\text{exposure}}$*white	0.00	3.48

	5%	95%
$\text{DDE}_{\text{exposure}}$	-0.05	0.01
$\text{PCB}_{\text{exposure}}$	-3.01	-0.83
$\text{DDE}_{\text{exposure}}$	-0.13	0.01
$\text{PCB}_{\text{exposure}}$	0.05	3.55

Table 1: 90% credible intervals for all coefficients, under prior on R^2 with location 0.3 (Left), under prior on R^2 with location 0.5 (Middle) under prior on R^2 with location 0.8 (Right)

1. Sensitivity Analysis

We vary different priors (prior on R^2 with location 0.3, 0.5 and 0.8 respectively) to check if the model is sensitive to priors (See Table 1). We attempted to check with other priors, like Cauchy, but it's impossible to specify such prior in the *stan_polr* function of this package. To conclude, our method is not sensitive to the choice of priors.

2. Model Checking

Since the ordinal data is used, the common residual model checking plot is no longer applicable. Instead, the surrogate residual method suggested by () is used.

Latent variables Z can be used to parameterize the Bayesian logistics model. Specifically, $Z = -X\beta + \epsilon$ and $Y = j$ if $Z \in [\alpha_{j-1}, \alpha_j]$, where ϵ is a random variable with cumulative distribution $G(\cdot)$ and α_j is some threshold value. $G^{-1}(\cdot)$ is the link function of the model. Surrogate residual is defined as $R_S = S - E(S|X)$, where S is some continuous variable generated from the conditional distribution of latent variables Z given observation Y. If the model assumptions are satisfied, the surrogate residual R_S should display three characteristics:

1. $E(R_S|X) = 0$
2. $Var(R_S|X) = c$, the conditional variance of R_S is constant.
3. The empirical distribution of R_S resembles an explicit distribution that is related to the link function $G^{-1}(\cdot)$. Specifically, $R_S \sim G(c + \int u dG(u))$ and R_S is independent of X, where c is a constant.

To explain more straightforwardly, if the model assumptions are satisfied, R_S should distribute evenly around 0, independent of X. Besides, the empirical quantiles of R_S should match those of the theoretical distribution.

3. Full Model Output

The comprehensive output of our model is also included (See Table 3 for credible intervals and Figure 1 for the histogram). Although the effects of variables other than $\text{DDE}_{\text{exposure}}$

	5%	95%
$DDE_{exposure}$	-0.05	0.01
$PCB_{exposure}$	-2.73	-0.75
white	0.10	0.85
center10	-0.13	0.77
center15	-1.11	-0.28
center31	-0.25	0.80
center37	-1.01	-0.32
center45	-0.42	0.35
center50	-0.54	0.23
center55	-0.78	0.08
center60	-0.66	0.16
center66	-0.48	0.19
center71	-0.46	0.28
center82	-1.09	-0.29
smoking_status1	-0.31	0.00
$DDE_{exposure}$:white	-0.12	0.02
$PCB_{exposure}$:white	0.01	3.21
Dangerous—Pre term	-3.24	-2.54
Pre term—At term	-1.88	-1.22

Table 2: 90% credible intervals for all coefficients, under the uniform prior

and $PCB_{exposure}$ an are not the focus on this report, we can still interpret the coefficients of variables like intercepts and center.

- Intercept: when a subject is non-white, measured at center 5, doesn't smoke, and exposed to 0 level of DDE and PCB, her 90% credible interval for the risk of dangerous preterm is $\frac{1}{1+e^{[-3.24,-2.54]}} * 100\% = [3.77\%, 7.31\%]$. 90% credible interval for the dangerous preterm or preterm is $\frac{1}{1+e^{[-1.88,-1.22]}} * 100\% = [13.24\%, 22.79\%]$.
- Center: There are clear heterogeneity across centers. Center5 is chosen to be the baseline here. Center 15, 37, 82 are significantly different from the baseline because their 90% credible intervals do not cover 0.

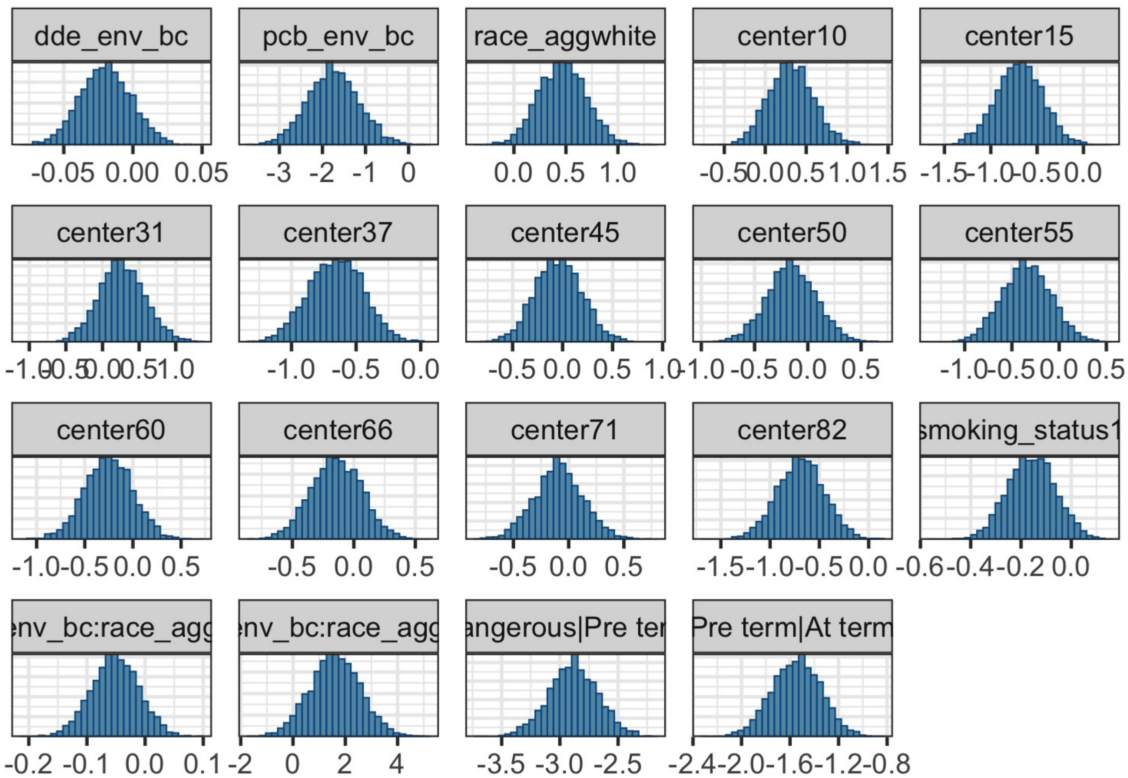

Figure 1: Histogram of 90% credible intervals