



HAL
open science

PRIORS COMPARISON IN BAYESIAN MEDIATION FRAMEWORK WITH BINARY OUTCOME

Jean-Michel Galharret, Anne Philippe

► **To cite this version:**

Jean-Michel Galharret, Anne Philippe. PRIORS COMPARISON IN BAYESIAN MEDIATION
FRAMEWORK WITH BINARY OUTCOME. 2019. hal-02070053

HAL Id: hal-02070053

<https://hal.science/hal-02070053>

Preprint submitted on 16 Mar 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

PRIORS COMPARISON IN BAYESIAN MEDIATION FRAMEWORK WITH BINARY OUTCOME

J.M. GALHARRET AND A. PHILIPPE

ABSTRACT. In human sciences, mediation designates a particular causal phenomenon where the effect of a variable X on another variable Y passes (partially or entirely) through a third variable M . The parameters of interest in mediation models are the direct effect of X on Y and the indirect effect of X on Y through M . We use a Bayesian framework to estimate these parameters and we compare different construction of prior distribution. The results show that providing information improves the quality of the estimation. We also propose a Bayesian procedure to test the absence of direct and indirect effect. Results of simulations show that this procedure works as well as the usual frequentist test for the direct effect and better than the usual bootstrapping test for the indirect effect. Finally, we apply our approach to real data from a longitudinal study on the well-being of children in school.

Keywords : mediation analysis, Bayesian estimation, direct and indirect effect

1. INTRODUCTION

Studying a simple mediation model (X, M, Y) consist of analysing the effect of X on Y (see VanderWeele [17], Hayes [6] for general presentation). In the mediation analysis, this effect is subdivided into the direct influencing of X on Y (direct effect) and the influence of X on Y through the mediator variable M (indirect effect). In many situations the mediation model is applied with X as a binary variable, M a continuous variable, and where Y is a binary outcome. For instance, in nutrition, Fisher et al. [2] investigated the effect of duration of breast-feeding on caloric intake of the toddlers mediated by their maternal control of feeding. In management, Wieder and Ossimitz [18] studied the direct and indirect effects of Business Intelligence Management on the quality of managerial decision through Data Quality and Information Quality.

To understand the causal relationship between X and Y , Robins and Greenland [14], Pearl [12] defined the counterfactual approach. For each individual, ideally, we would like to observe the outcome Y under both exposure $X = 1$ and $X = 0$, but it is not possible. The two fictitious variables are named the counterfactual variables and denoted by Y_x when $X = x$ ($x = 0, 1$). From the counterfactual context, the causal effect of X on Y is defined from:

$$\psi = \mathbb{E}(Y_1 - Y_0)$$

This causal effect can be broken down into the natural direct effect (NDE) and the natural indirect effect (NIE). Under a set of hypothesis (see Imai et al. [7],

Date: March 16, 2019, Laboratoire de Mathématiques Jean Leray (LMJL- UMR 6629), F-44000 Nantes, France;

VanderWeele [17], for details) we have for $x \in \{0, 1\}$:

$$\begin{aligned} \text{NDE}(x) &= \int_{\mathcal{M}} [\mathbb{E}(Y|X=1, M=m) - \mathbb{E}(Y|X=0, M=m)] dF_{M|X=x}(m) \\ \text{NIE}(x) &= \int_{\mathcal{M}} \mathbb{E}(Y|X=x, M=m) [dF_{M|X=1}(m) - dF_{M|X=0}(m)] \end{aligned}$$

where $F_{M|X=x}$ denotes the conditional distribution of M given $X=x$. We obtain the following decomposition of the causal effect:

$$(1) \quad \psi = \text{NDE}(0) + \text{NIE}(1)$$

The causal effect can be also written as $\psi = \text{NDE}(1) + \text{NIE}(0)$

Remark: The first mediation model was introduced by Baron and Kenny [1] in a regression framework (with Y as a continuous variable). The causal effect ψ of X on Y was defined by the equation:

$$Y = a_0 + \psi X + \varepsilon_{0Y}$$

where $\varepsilon_{0Y} \sim \mathcal{N}(0, \sigma_0^2)$ For natural direct effect and the natural indirect effect, $\text{NDE}(0) = \text{NDE}(1) = \gamma$ and $\text{NIE}(0) = \text{NIE}(1) = \alpha\beta$ with (α, β, γ) with:

$$\begin{aligned} Y &= \beta_0 + \beta M + \gamma X + \varepsilon_Y \\ M &= \alpha_0 + \alpha X + \varepsilon_M \end{aligned}$$

where $\varepsilon_Y \sim \mathcal{N}(0, \sigma^2)$ and $\varepsilon_M \sim \mathcal{N}(0, \tau^2)$.

Bayesian mediation model. Let (X, M, Y) the model studied with X, Y binary variables and M continuous. The aim of the study is to estimate ψ . The estimation of ψ is based on (1). The definition of the natural direct and indirect effects requires $\mathbb{E}(Y|X=x, M=m)$ and $F_{M|X=x}$. The most common chosen for $\mathbb{E}(Y|X=x, M=m)$ is the logistic model and for $F_{M|X=x}$ the normal distribution. Thus we consider the model defined by the two linear equations and described in the Figure 1:

$$\begin{aligned} (2) \quad M &= \alpha_0 + \alpha X + \varepsilon_M \\ (3) \quad \text{logit}(\mathbb{E}(Y|X, M)) &= \beta_0 + \beta M + \gamma X \end{aligned}$$

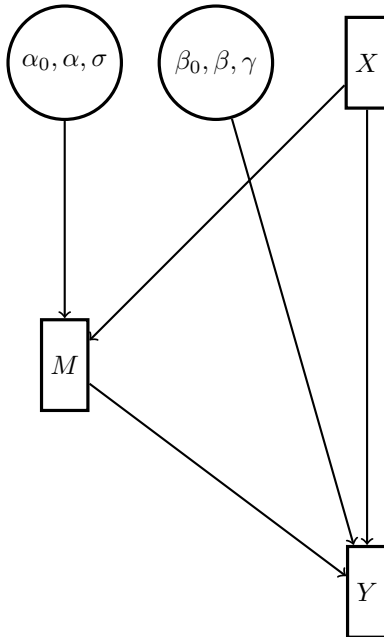
where $\varepsilon_M \sim \mathcal{N}(0, \sigma^2)$. Hereafter, we refer to the model defined by (2), (3) as \mathcal{BM} (Bayesian Mediation).

Let $\theta := (\alpha_0, \alpha, \sigma^2, \beta_0, \beta, \gamma)$ the vector of unknown parameters in the model \mathcal{BM} . Let NDE_θ denotes the parameter $\text{NDE}(0)$, and NIE_θ denotes the parameter $\text{NIE}(1)$.

$$(4a) \quad \text{NDE}_\theta = \frac{1}{\sigma\sqrt{2\pi}} \int \left(\frac{1}{1 + e^{-(\beta_0 + \beta m + \gamma)}} - \frac{1}{1 + e^{-(\beta_0 + \beta m)}} \right) e^{-\frac{1}{2} \left(\frac{m - \alpha_0}{\sigma} \right)^2} dm$$

$$(4b) \quad \text{NIE}_\theta = \frac{1}{\sigma\sqrt{2\pi}} \int \frac{1}{1 + e^{-(\beta_0 + \beta m + \gamma)}} \left(e^{-\frac{1}{2} \left(\frac{m - (\alpha_0 + \alpha)}{\sigma} \right)^2} - e^{-\frac{1}{2} \left(\frac{m - \alpha_0}{\sigma} \right)^2} \right) dm$$

From the two relations, we deduce the posterior distribution of our interest parameters $\text{NDE}_\theta, \text{NIE}_\theta$ from the one of parameter θ . To complete the definition of the model \mathcal{BM} , we have to choose the prior distribution on θ . For the applications in psychology, the information required to construct the prior distribution may come from previous measures on the same population (e.g. longitudinal study) or from

FIGURE 1. DAG (Directed Acyclic Graph) of mediation model \mathcal{BM}

an independent study conducted under the same conditions. One of the objectives of this paper is to discuss on the choice of the prior distribution and to show the interest to consider prior information. We illustrate through some simulations that the interest of including information in the construction of the prior distribution. We evaluate the impact of the prior using frequentist criteria.

The paper is organized as follow. In Section 2, we describe the Bayesian model, and the different prior distributions on the parameter θ : an informative prior introduced by Launay et al. [9], a weakly informative prior introduced by Gelman et al. [4] and a noninformative G-prior introduced by Zellner [19]. In Section 3 the simulations show that the informative approach based on historical informations allow reduction of the statistical bias and variability of the estimators. In Section 4 we propose a Bayesian test for the absence of the direct effect and the absence of the indirect effect. An application to schoolchildren's well-being is given Section 5.

2. DESCRIPTION OF BAYESIAN MODELS

Bayesian estimation. We propose Bayesian estimations for NDE_θ and NIE_θ . The likelihood of the model \mathcal{BM} defined in (2) is:

$$\mathcal{L}_\theta(\underline{Y}, \underline{M}, \underline{X}) = \prod_{i=1}^N \frac{e^{-(1-y_i)(\beta_0 + \beta m_i + \gamma x_i)}}{1 + e^{-(\beta_0 + \beta m_i + \gamma x_i)}} \times \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{m_i - \alpha_0 - \alpha x_i}{\sigma}\right)^2}$$

where for instance \underline{Y} denotes the vector $= (y_1, y_2, \dots, y_N)$ of the N independent observations for the variable Y . The posterior distribution of $\theta = (\alpha_0, \alpha, \sigma, \beta_0, \beta, \gamma)$

is

$$\pi(\theta|\underline{Y}, \underline{M}, \underline{X}) \propto \mathcal{L}_\theta(\underline{Y}, \underline{M}, \underline{X})\pi(\theta).$$

The parameters of interest NDE_θ and NIE_θ are non explicit function of θ defined in 1. The Bayes estimate of these two quantities are respectively:

$$\begin{aligned} \mathbb{E}(NDE_\theta|\underline{Y}, \underline{M}, \underline{X}) &= \int_{\Theta} NDE_\theta \pi(\theta|\underline{Y}, \underline{M}, \underline{X})d\theta \\ \mathbb{E}(NIE_\theta|\underline{Y}, \underline{M}, \underline{X}) &= \int_{\Theta} NIE_\theta \pi(\theta|\underline{Y}, \underline{M}, \underline{X})d\theta \end{aligned}$$

We consider now different strategies to construct the prior distributions according to the approaches proposed in Launay et al. [9], Gelman et al. [4], Zellner [19].

Prior distribution for variance parameter σ^2 . : For all models, the scale parameter σ^2 in equation (2) had an improper prior on non-negative real numbers. This prior is acceptable since it leads to proper posterior for the parameters of the model (see Gelman [3] for discussion on this choice).

Model 1: The first model comes from Launay et al. [9]. In this case, the prior distribution of θ is based on historical informations. On an historical dataset, we calculate the mean M and the covariance V of the posterior distribution $(\alpha_0, \alpha, \beta_0, \beta, \gamma)$

$$\begin{aligned} M &= \mathbb{E}((\alpha_0, \alpha, \beta_0, \beta, \gamma)|historical\ data) \\ V &= \text{cov}((\alpha_0, \alpha, \beta_0, \beta, \gamma)|historical\ data) \end{aligned}$$

For this historical sample we choose the prior distribution described in Model 3. Assuming that both samples come from similar studies, we take as prior distribution on $(\alpha_0, \alpha, \beta_0, \beta, \gamma)$ the distribution defined by:

$$(\alpha_0, \alpha, \beta_0, \beta, \gamma) \sim \mathcal{N}_5(\mu, V)$$

$$\text{where } \mu = \begin{pmatrix} k_1 & & & & \\ & \ddots & & & \\ & & & & \\ & & & & k_5 \end{pmatrix} M \text{ and } k_j \sim \mathcal{N}(1, 5) \text{ The hyper-parameters } k_1, \dots, k_5$$

allow that the parameter $(\alpha_0, \alpha, \beta_0, \beta, \gamma)$ varies slightly between the two studies. As for the structure of dependence between the variables, we impose a priori that it is unchanged. For example, if a first study shows that the Business Intelligence Management has a positive effect on the quality of managerial decision, we can expect that this effect will also be positive in a futur study. What may change from one study to another is the importance of the relationship between these two variables. In a longitudinal study, we are interested by the evolution of the characteristics of the participants. These characteristics should evolve over time, but not their correlations.

Remark: In Launay et al. [9], the hierarchical model is more complex, in particular, a scale parameter is added to consider the historical sample is very large compared to the one of interest. In this model, V is replaced by λV , where λ is a positive unknown parameter. This parameter λ plays an important role in terms of robustness if the historical information is not relevant.

Model 2: Without historical data, a noninformative prior is proposed. We assume that the prior distribution of (α_0, α) in (2) is a Zellner's G -prior Zellner [19]. Let ϕ

the $N \times 2$ matrix defined by $\phi = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_N \end{pmatrix}$, the G -prior distribution of (α_0, α) is:

$$(\alpha_0, \alpha) | \sigma \sim \mathcal{N}_2(0_2, g_0 \sigma^2 (\phi' \phi)^{-1})$$

We set $g_0 = N$ to ensure that the prior has the same weight as one observation of the sample (see Marin and Robert [11] for details). For the logistic equation (3), we use a weakly informative prior proposed by Gelman et al. [4]. As suggested, X is centered and M is centered and scaled to attribute a value of 0.5 to the standard deviation. The corresponding variables are denoted by $\tilde{\bullet}$. Let $(\tilde{\beta}_0, \tilde{\beta}, \tilde{\gamma})$ defined by: $\text{logit}(\mathbb{E}(Y | \tilde{X}, \tilde{M})) = \tilde{\beta}_0 + \tilde{\beta} \tilde{M} + \tilde{\gamma} \tilde{X}$. The relation between $(\tilde{\beta}_0, \tilde{\beta}, \tilde{\gamma})$ and (β_0, β, γ) is given by

$$(5a) \quad \gamma = \tilde{\gamma}$$

$$(5b) \quad \beta = \frac{\tilde{\beta}}{2 \sigma_M}$$

$$(5c) \quad \beta_0 = \tilde{\beta}_0 - \gamma \bar{X} - \beta \bar{M}$$

In Gelman et al. [4], the prior distributions of the parameters $(\tilde{\beta}_0, \tilde{\beta}, \tilde{\gamma})$ are

$$\tilde{\beta}_0 \sim \mathcal{C}(0, 10)$$

$$\tilde{\beta} \sim \mathcal{C}(0, 2.5)$$

$$\tilde{\gamma} \sim \mathcal{C}(0, 2.5)$$

where $\mathcal{C}(m, \tau)$ denotes the Cauchy distribution with m the location parameter and τ the scale parameter. The posterior distributions of the parameters (β_0, β, γ) can be deduced from the posterior distributions of the parameters $(\tilde{\beta}_0, \tilde{\beta}, \tilde{\gamma})$ using (5). Model 3. The third one is a non informative model. For the linear equation (2) we keep the same prior distribution as the one for Model 2. For the logistic equation (3), we choose a Zellner G -prior for (β_0, β, γ)

$$(\beta_0, \gamma, \beta) \sim \mathcal{N}_3(b_1, g_1 N^2 (\psi' \psi)^{-1})$$

where ψ is the $N \times 3$ matrix: $\psi = \begin{pmatrix} 1 & x_1 & m_1 \\ \vdots & \vdots & \vdots \\ 1 & x_N & m_N \end{pmatrix}$. The choice $b_1 = 0_3$ and

$g_1 = 4$ ensures that (β_0, β, γ) has a flat prior distribution (for details see Marin and Robert [11]).

Remark: Hanson et al. [5] introduced informative G -priors for logistic regression where the hyperparameter b_1, g_1 is estimated from historical dataset. The construction is restricted to Gaussian design matrix i.e.

$\begin{pmatrix} x_1 & m_1 \\ \vdots & \vdots \\ x_N & m_N \end{pmatrix}$ has a multivariate

normal distribution. When X is a binary variable, the same strategy can not be applied to bring prior information. In this case an explicit form of b_1, g_1 is not available.

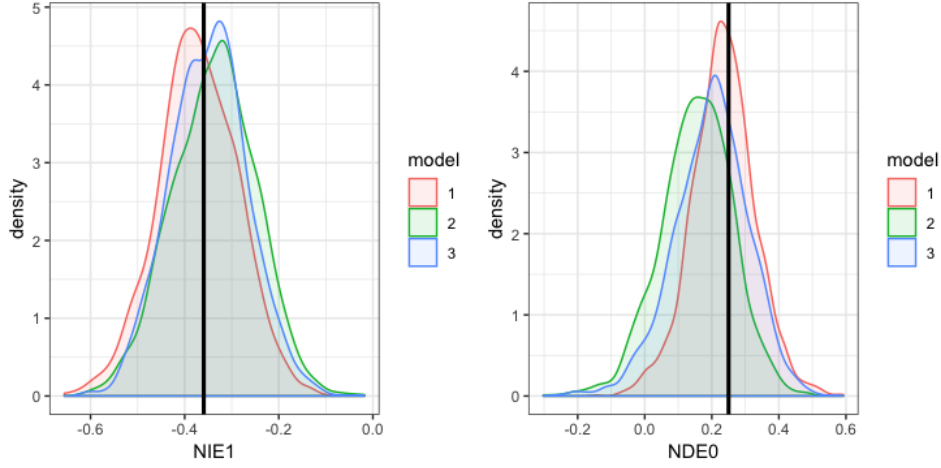


FIGURE 2. Estimated posterior distribution of NDE_θ, NIE_θ in the CASE 1 with $N = 50$.

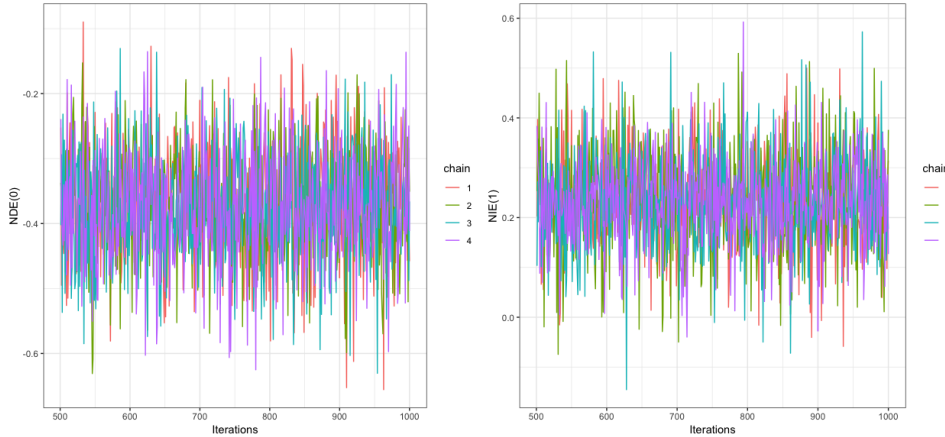


FIGURE 3. MCMC samples of NDE_θ, NIE_θ under Model 1 in the CASE 1 with $N = 50$

3. NUMERICAL RESULTS:

Method. To compare the three prior distributions presented previously, we evaluate the accuracy of Bayes estimates of $NDE(0)$ and $NIE(1)$ using frequentist criteria. From Monte Carlo experiments, we calculate the bias, the root of mean square error (RMSE) and the coverage probability of the credible interval. We also provide the length of credible interval, which evaluates the precision of the posterior distributions.

To approximate the posterior distributions, Markov Chain Monte Carlo (MCMC) method is required. We use Hamiltonian algorithm implemented in Stan application (R Core Team [13], Stan Development Team [15]). Let $(\theta^t)_{1 \leq t \leq T}$ be the resulting

sampling generated from the posterior distribution of θ . From $(\theta^t)_{1 \leq t \leq T}$, we deduce a MCMC sample from the posterior distribution of both parameters of interest NDE(0) and NIE(1). For each θ^t , we take NDE_{θ^t} , NIE_{θ^t} defined in (4a) and (4b). Note that this step requires a numerical approximation of integrals in (4a), (4b).

We also add to the comparison the estimate implemented in the R package `mediation` (see Tingley et al. [16]). In the function `mediate` a quasi-Bayesian Monte Carlo method based on normal approximation is implemented to estimate NDE(0) and NIE(1) (see Imai et al. [8] for details).

Simulated dataset. We generate historical and current data from three distinct situations. In all cases, the standard deviation of the linear model (2) is set to $\sigma = 0.75$ for both datasets. For the historical sample we choose the parameters $\alpha_0 = 1, \alpha = -2, \beta_0 = -0.5, \beta = 1, \gamma = 1.5$ for the three different scenarios. The current sample is simulated with the following values of the parameters $(\alpha_0, \alpha, \beta_0, \beta, \gamma)$:

- case 1 the same values as the historical data.
- case 2 the values increased by 10% .
- case 3 only the direction of the relationships between the variables (X, M, Y) does not change. We set $\alpha_0 = 1, \alpha = -3, \beta_0 = -1, \beta = 0.75, \gamma = 2$. It is the worst of the three cases for Model 1. The historical dataset bring misleading informations in the prior distribution.

These three situations make sense in real studies. For instance, in a longitudinal study, one might think that the relationships between the variables could be comparable over time. To assess the quality criteria, we generate $N = 1000$ independent copies. Figure 2 compares the posterior distributions of NDE(0), NIE(1) calculated on one of the samples simulated in the case 1.

From the posterior distributions of NDE(0), we clearly show the interest of including historical information in the prior distribution. Indeed, the accuracy and the precision of posterior distribution of NDE(0) are better with informative prior (Model 1). The same improvement is not observed as clearly for NIE(1). Figure 3 assesses graphically the convergence of the Markov chains for the mixing properties. We observed the same behavior for all the other models, independently of the values of the parameters.

Results. All the numerical results are summarized in Table 1. For all the situations, the informative model (Model 1) performed better than the others for all the statistical properties.

In the case 1 and 2, the historical informations reduced by half the bias of the estimation for the two parameters of interest NDE(0), NIE(1). Even in the worst case (case 3), the performances are lower, but the historical information still reduced the bias.

For the RMSE, the informative model's gain was about 20% relative to the others. The credible intervals of the estimations in Model 1,3 and the mediate function have correct confidence level, approximatively 95% for all sample size N . But, Model 1 provide a shorter credible than the others.

Model 2 based on weakly informative priors and introduced by Gelman et al. [4] had the worst statistical properties in all case. Note that we apply this prior distribution to the logistic part of a mediation model which is not the context studied in Gelman et al. [4]. Model 3 based on the G -priors and the mediate function have similar statistical properties.

TABLE 1. m denotes the mediate function, 1,2,3 the number of the precedent models. CL denotes the confidence level of the credible interval and L its length.

N	prior	$NIE(1)$						$NDE(0)$					
		bias	%bias	rmse	%rmse	CL	L	bias	%bias	rmse	%rmse	CL	L
CASE 1: Time-invariant parameters													
30	m	0.075	-20.81	0.162	45.19	0.946	0.63	-0.059	-23.79	0.195	77.90	0.968	0.76
	1	0.031	-8.49	0.133	37.08	0.941	0.51	-0.020	-7.83	0.152	60.63	0.943	0.58
	2	0.115	-32.09	0.173	48.11	0.893	0.57	-0.119	-47.46	0.203	81.09	0.926	0.71
	3	0.072	-20.03	0.160	44.64	0.940	0.57	-0.046	-18.39	0.199	79.40	0.934	0.72
50	m	0.057	-15.96	0.122	33.92	0.948	0.48	-0.043	-17.27	0.151	60.33	0.957	0.59
	1	0.026	-7.33	0.099	27.45	0.950	0.40	-0.011	-4.32	0.116	46.47	0.931	0.46
	2	0.089	-24.64	0.135	37.61	0.907	0.46	-0.088	-35.35	0.162	64.70	0.924	0.58
	3	0.054	-14.98	0.120	33.25	0.948	0.45	-0.033	-13.35	0.151	60.48	0.943	0.57
100	m	0.029	-7.96	0.080	22.37	0.944	0.33	-0.020	-8.01	0.101	40.29	0.959	0.41
	1	0.013	-3.52	0.070	19.44	0.956	0.29	-0.002	-0.70	0.080	32.20	0.950	0.34
	2	0.049	-13.50	0.089	24.82	0.920	0.32	-0.049	-19.60	0.109	43.56	0.947	0.41
	3	0.026	-7.30	0.079	22.01	0.944	0.31	-0.014	-5.77	0.101	40.22	0.944	0.40
CASE 2: Parameters grow of 10%													
30	m	0.082	-20.02	0.167	40.59	0.953	0.65	-0.064	-26.06	0.186	75.17	0.976	0.77
	1	0.042	-10.10	0.135	32.74	0.946	0.52	-0.035	-14.14	0.149	60.23	0.942	0.56
	2	0.128	-31.07	0.183	44.44	0.897	0.59	-0.132	-53.22	0.203	81.98	0.919	0.72
	3	0.080	-19.42	0.164	39.90	0.939	0.58	-0.051	-20.83	0.189	76.60	0.957	0.72
50	m	0.052	-12.65	0.124	30.13	0.949	0.50	-0.040	-16.16	0.142	57.27	0.968	0.58
	1	0.023	-5.67	0.103	25.10	0.945	0.41	-0.016	-6.32	0.110	44.70	0.944	0.44
	2	0.087	-21.21	0.139	33.88	0.900	0.47	-0.090	-36.58	0.158	64.03	0.932	0.57
	3	0.050	-12.04	0.121	29.49	0.945	0.46	-0.030	-12.14	0.142	57.48	0.949	0.56
100	m	0.030	-7.23	0.082	19.95	0.951	0.33	-0.024	-9.90	0.099	40.00	0.960	0.40
	1	0.014	-3.33	0.071	17.26	0.958	0.29	-0.009	-3.47	0.077	31.32	0.953	0.32
	2	0.051	-12.42	0.092	22.40	0.926	0.33	-0.056	-22.50	0.110	44.67	0.937	0.40
	3	0.028	-6.77	0.081	19.59	0.952	0.31	-0.019	-7.67	0.098	39.76	0.954	0.38
CASE 3: Invariant direction in the relationship													
30	m	0.124	-27.34	0.230	50.74	0.949	0.82	-0.125	-31.45	0.254	64.08	0.966	0.93
	1	0.064	-14.15	0.178	39.39	0.961	0.68	-0.089	-22.48	0.199	50.30	0.941	0.73
	2	0.213	-47.03	0.264	58.27	0.862	0.75	-0.228	-57.63	0.291	73.48	0.890	0.88
	3	0.123	-27.20	0.227	50.13	0.941	0.74	-0.110	-27.75	0.255	64.40	0.950	0.88
50	m	0.077	-16.90	0.169	37.27	0.966	0.66	-0.075	-18.96	0.193	48.65	0.965	0.75
	1	0.040	-8.78	0.132	29.08	0.955	0.54	-0.060	-15.25	0.157	39.65	0.923	0.59
	2	0.154	-34.01	0.206	45.38	0.898	0.64	-0.165	-41.61	0.231	58.31	0.921	0.74
	3	0.073	-16.17	0.166	36.54	0.958	0.62	-0.063	-15.85	0.191	48.11	0.960	0.72
100	m	0.058	-12.81	0.131	28.81	0.958	0.48	-0.056	-14.00	0.152	38.23	0.951	0.55
	1	0.040	-8.78	0.104	23.01	0.955	0.41	-0.049	-12.23	0.124	31.33	0.920	0.46
	2	0.109	-24.00	0.158	34.78	0.901	0.49	-0.114	-28.82	0.179	45.02	0.909	0.57
	3	0.055	-12.14	0.129	28.37	0.950	0.46	-0.048	-12.16	0.150	37.93	0.943	0.54

To summarize the numerical results, the performance of the estimate of the informative prior are similar to others calculated on twice as many observations.

4. TESTING THE ABSENCE OF EFFECT:

Investigators may be interested by the absence of direct effect or indirect effect. When X is a binary variable, it amounts to test $\text{NDE}(0) = 0$, $\text{NDE}(1) = 0$, $\text{NIE}(0) = 0$ or $\text{NIE}(1) = 0$.

Proposition 4.1. *Let (X, M, Y) a mediation model with a binary outcome Y . Let suppose $M = \alpha_0 + \alpha X + \varepsilon_M$ with $\mathcal{N}(0, \sigma^2)$ and:*

$$\logit(\mathbb{E}(Y|X, M)) = \beta_0 + \beta M + \gamma X \text{ (logistic model)}$$

or

$$\log(\mathbb{E}(Y|X, M)) = \beta_0 + \beta M + \gamma X \text{ (Poisson model)}$$

The following statements are equivalents:

$$\begin{aligned} \text{NDE}(0) = 0 &\iff \text{NDE}(1) = 0 \iff \gamma = 0 \\ \text{NIE}(0) = 0 &\iff \text{NIE}(1) = 0 \iff \alpha\beta = 0 \end{aligned}$$

Proof. We consider the logistic model. For the natural direct effect, we have:

$$\text{NDE}(x) = \frac{1}{\sigma\sqrt{2\pi}} \int \left(\frac{1}{1 + e^{-(\beta_0 + bm + \gamma)}} - \frac{1}{1 + e^{-(\beta_0 + bm)}} \right) e^{-\frac{1}{2} \left(\frac{m - (\alpha_0 + \alpha x)}{\sigma} \right)^2} dm$$

Let f_γ the function defined for all m by

$$f_\gamma(m) = \frac{1}{1 + e^{-(\beta_0 + bm + \gamma)}} - \frac{1}{1 + e^{-(\beta_0 + bm)}}$$

The sign of f_γ is the same as γ , thus NDE is the integral of a non-negative (or negative) function. $\text{NDE}(x) = 0$ is obviously equivalent to $f_\gamma(m) = 0$ for all m , which is in turn equivalent to $\gamma = 0$.

For the indirect effect, by the method of substitution, a straight forward calculus leads to:

$$\text{NIE}(x) = \frac{1}{\sigma\sqrt{2\pi}} \int \left(\frac{1}{1 + e^{-(\beta_0 + \beta(m + \alpha_0) + \gamma x)}} - \frac{1}{1 + e^{-(\beta_0 + \beta(m + \alpha_0) + \alpha + \gamma x)}} \right) e^{-\frac{1}{2\sigma^2} m^2} dm$$

The function $f_{(\alpha, \beta)}$ defined by $f_{(\alpha, \beta)}(m) = \frac{1}{1 + e^{-(\beta_0 + \beta(m + \alpha_0) + \gamma x)}} - \frac{1}{1 + e^{-(\beta_0 + \beta(m + \alpha_0) + \alpha + \gamma x)}}$ had the same sign as $\alpha\beta$.

The previous arguments are valid for the Poisson model and the proof is similar. \square

Testing the absence of the direct or the indirect effect is one of the major issues of the mediation analysis. Let α_c be the significance level of the test. We can perform standard tests based on the parameters α, β, γ :

- For the absence of direct effect ($\gamma = 0$), in logistic model, we can use a likelihood ratio test (LR test) or a Wald test.
- For the absence of indirect effect ($\alpha\beta = 0$), MacKinnon [10] proposed a bootstrapping procedure. The $(1 - \alpha_c)$ BCa (bias-corrected and accelerated) bootstrapping confidence interval for the parameter $\alpha\beta$ is computed. The decision rule is to reject the absence of indirect effect when 0 is not in the confidence interval.

Procedure testing for the absence of effects: Thank to the results of Proposition 4.1, the hypothesis can be expressed using the parameters $\text{NDE}(x)$ and $\text{NIE}(x)$ ($x \in \{0, 1\}$). Let us consider the hypothesis of absence of direct effect defined by $\mathcal{H}_0 : \text{NDE}(x) = 0$. We propose decision rules based on our Bayesian analysis. The main idea is to construct the critical region of the test from the $(1 - \alpha_c)$ credible interval \mathcal{I}_{α_c} of the parameter $\text{NDE}(x)$. The decision rule is to reject the null hypothesis when the credible interval \mathcal{I}_{α_c} does not contain 0. The asymptotic properties of the posterior distribution ensure that the significance level of this procedure is α_c .

The same procedure can be implementing for testing:

- the absence of indirect effect defined by $\mathcal{H}_0 : \text{NIE}(x) = 0$
- the absence of causal effect defined by $\mathcal{H}_0 : \psi = 0$ where $\psi = \text{NDE}(x) + \text{NIE}(1 - x)$.

Numerical comparison: We compare our testing procedure to the LR test (respectively bootstrapping procedure) for the absence of direct (respectively indirect) effect. For this comparison, we generate $B = 1000$ sample from the following model: for all $i \in \{1, \dots, N\}$, the triple (X_i, M_i, Y_i) is i.i.d. from

$$\begin{aligned}
 X_i &\sim \mathcal{B}(0.4), \\
 M_i &\sim \mathcal{N}(1 + \alpha X_i, 0.75^2), \\
 Y_i &\sim \mathcal{B}(p_i), \\
 p_i &= \frac{1}{1 + e^{-(2 + \beta M_i + \gamma X_i)}}.
 \end{aligned}
 \tag{6}$$

We set the significant level $\alpha_c = .05$ and we compare the empirical probability to reject the null hypothesis.

Table 2 gives the results for testing the absence of direct effect. The value $\gamma = 0$ corresponds to the empirical size of tests, and $\gamma > 0$ to the empirical power. The two approaches have the same performances in terms of empirical size. As expected, the empirical power of the different tests increases with γ and N , with the same rate of convergence.

Table 3 gives the results for testing the absence of indirect effect. The performances of the test depend on the product $\alpha\beta$, but also the values of each parameters α, β . That's why α is fixed and we illustrate the convergence rate of the power as a function of $\alpha\beta$. Let's remember that the value $\alpha\beta = 0$ corresponds to the empirical size of tests, and $\alpha\beta > 0$ to the empirical power. Both Bayesian procedures have an empirical size closer to the nominal level $\alpha_c = 0.05$ than the bootstrapping procedure. As expected, the empirical power of the different tests increases with $\alpha\beta$ and N , but the Bayesian procedure are uniformly more powerful. In addition, Bayesian procedure is less expensive in term of computation time.

5. APPLICATION TO SCHOOLCHILDREN'S WELL-BEING:

The data come from a longitudinal study of schoolchildren aged 6 to 9 years. Their academic performance was assessed three times at 6-month intervals. The researchers would like to prove that the schoolchildren's academic performance at time $t + 1$ depends directly on their academic performance at time t and indirectly through their Self Efficacy Feeling (see Figure 4). The higher the academic performances, the greater the Self Efficacy Feeling, which in turn generates higher

TABLE 2. The empirical probability to reject the absence of direct effect using our Bayesian decision rule and using the LR test. We set the significant level and $\alpha_c = .05$ $\alpha = \beta = 2$ in the model defined in (6)

		NDE(0)	NDE(1)	LR test
$\gamma = 0$	$N = 30$	0.04	0.04	0.05
	$N = 50$	0.03	0.03	0.05
	$N = 100$	0.05	0.05	0.07
$\gamma = 1$	$N = 30$	0.09	0.09	0.10
	$N = 50$	0.11	0.11	0.12
	$N = 100$	0.21	0.21	0.21
$\gamma = 2$	$N = 30$	0.23	0.23	0.24
	$N = 50$	0.36	0.36	0.37
	$N = 100$	0.64	0.64	0.63
$\gamma = 3$	$N = 30$	0.41	0.41	0.39
	$N = 50$	0.78	0.78	0.77
	$N = 100$	0.97	0.97	0.96

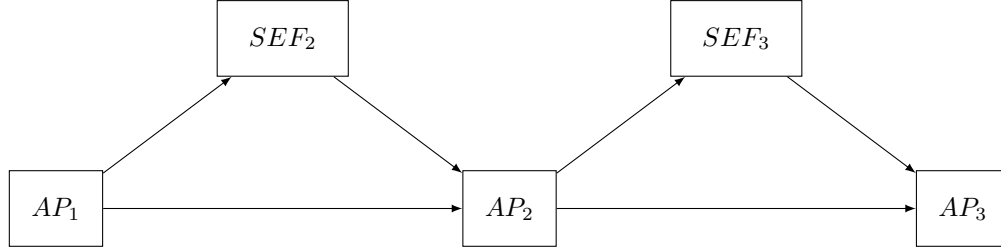


FIGURE 4. Structural model. AP : Academic Performance, SEF : Self Efficacy Feeling.

academic performances. Teachers evaluate academic performance as above follows: $AP = 1$ if the schoolchildren's results reach (or exceed) the expected level and $AP = 0$ otherwise. The Self-Efficacy Feeling is a 11-item questionnaire scale. Schoolchildren responded to the item on a scale ranging from 1 (not agree at all) to 6 (totally agree).

The structural model was tested controlling the gender (W). To include the categorical covariate W on a mediation model (X, M, Y) , VanderWeele [17] define the natural direct effect by:

$$\text{NDE}(x) = \sum_w \text{NDE}_w(x) \mathbb{P}(W = w)$$

where

$$\text{NDE}_w(x) := \int_{\mathcal{M}} [\mathbb{E}(Y|X = 1, M = m, W = w) - \mathbb{E}(Y|X = 0, M = m, W = w)] dF_{M|X=x, W=w}(m).$$

TABLE 3. The empirical probability to reject the absence of indirect effect using our Bayesian decision rule and the bootstrapping procedure. We set $\alpha_c = .05$, $\gamma = 2$, $\alpha = 2$ in the model defined in (6). The indicated values of the product $\alpha\beta$ correspond to $\beta \in \{0, 0.05, \dots\}$

$\alpha\beta$	N	NIE(0)	NIE(1)	boot
$\alpha\beta = 0$	30	0.07	0.07	0.01
	50	0.06	0.06	0.04
	100	0.06	0.06	0.04
$\alpha\beta = 0.1$	30	0.06	0.06	0.01
	50	0.08	0.08	0.04
	100	0.04	0.04	0.03
$\alpha\beta = 0.5$	30	0.07	0.07	0.02
	50	0.10	0.10	0.06
	100	0.10	0.10	0.07
$\alpha\beta = 1.00$	30	0.13	0.13	0.03
	50	0.20	0.20	0.12
	100	0.28	0.28	0.21
$\alpha\beta = 1.50$	30	0.20	0.20	0.05
	50	0.30	0.30	0.22
	100	0.57	0.57	0.46
$\alpha\beta = 2.00$	30	0.35	0.35	0.24
	50	0.30	0.30	0.23
	100	0.78	0.78	0.73
$\alpha\beta = 3.00$	30	0.60	0.60	0.42
	50	0.59	0.59	0.47
	100	0.98	0.98	0.97
$\alpha\beta = 4.00$	30	0.81	0.81	0.51
	50	0.82	0.82	0.73
	100	1.00	1.00	1.00

Similarly, we can define the natural indirect effect NIE with a categorical covariate W . The Bayesian model with a covariate W is defined by:

$$\begin{aligned} M &= \alpha_0 + \alpha X + cW + \varepsilon_M \\ \text{logit}(\mathbb{E}(Y|X, M)) &= \beta_0 + \beta M + \gamma X + dW \end{aligned}$$

We do not propose longitudinal modeling, but we apply the mediation model to the two measurement times independently. For the first time measurement, the dataset is $(Y, X, M) = (AP_2, AP_1, SEF_2)$. We apply only Model 3 since no historical information is available to construct Model 1. For the second time measurement, $(Y, X, M) = (AP_3, AP_2, SEF_3)$, both models can be implemented. For Model 1, the prior distribution is constructed taking $(Y, X, M) = (AP_2, AP_1, SEF_2)$ as historical sample. Note that AP_2 can be used to construct the informative prior because the mediation model is defined conditionally to (X, W) .

In Figure 5, we see that Model 1 and Model 3 give similar posterior distribution for $\text{NDE}(x)$. The use of historical data in Model 1 does not improve the estimation compared to Model 3. This result is not amazing when we compare the posterior

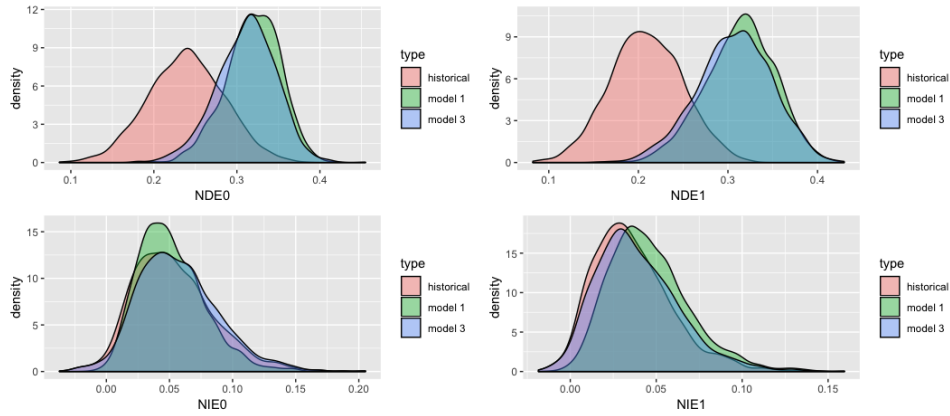


FIGURE 5. Posterior distribution of the parameters of interest $NDE(x)$, $NIE(x)$ ($x \in \{0, 1\}$) calculated on the data of the second measurement time. The density in pink (denoted by historical) represents the prior information (used in Model 1) and corresponds to the posterior distribution obtained at the first measurement time.

TABLE 4. 95 %-Credible Interval for $NDE(x)$, $NIE(x)$ ($x \in \{0, 1\}$) parameters at the second time measurement

	Direct Effect NDE		Indirect Effect NIE	
	Model 1	Model 3	Model 1	Model 3
$X = 0$	[0.236 ; 0.376]	[0.251 ; 0.378]	[0.009 ; 0.106]	[0.014 ; 0.115]
$X = 1$	[0.220 ; 0.384]	[0.232 ; 0.383]	[0.008 ; 0.096]	[0.014 ; 0.103]

distribution obtained by Model 3 at the two time measurements. Indeed, both density are very far away, so the information coming from historical sample is not relevant

For the parameter $NDE(x)$, the situation is very different since we notice that the estimations at the measurement times are similar. This context is favorable to Model 1, and this is confirmed by the more precise estimates.

Table 4 provides 95% credible intervals for the same parameter as Figure 5. These numerical values confirmed our previous conclusions. In addition, the application of our decision rule based on the credible interval (see section 4) allows us to test the significance of these parameters. The conclusion of testing procedure are the following:

- the past academic performances have a significant direct effect for the two groups ($AP = 1$ and $AP=0$)
- the direct effect varies over time while the indirect effect does not.
- the direct and indirect effects for the two groups ($AP = 0$ and $AP = 1$) are significant at the second time measurement.

6. DISCUSSION AND CONCLUSION

The Bayesian approach introduced by Launay et al. [9] provides informative prior based on the knowledge of historical sample. The key idea is to learn the correlation between the parameters from the historical sample. As shown in the numerical results, the improvement brought upon non or weak informative prior is more particularly important when there are weak variation on the parameters. This situation seems to be realistic in a longitudinal study. Its application to schoolchildren's Well-Being confirms this fact. We obtain estimation of direct and indirect effects more precise than the standard method implemented in R package mediation (see Tingley et al. [16]). We propose also a Bayesian decision rule for testing the absence of a direct effect or an indirect effect. For the direct effect we can see numerically that the Bayesian rule performs similarly than the classical Student's t-test. For the indirect effect, there is no exact test but a bootstrapping test is typically used. The comparison with this test is in favor of our Bayesian rule.

ACKNOWLEDGEMENTS

This research was funded by several grants from the French Ministère de l'Éducation Nationale et de l'Enseignement Supérieur, the Défenseur des Droits, and l'Agence pour la Cohésion Sociale et l'Égalité des Chances.

REFERENCES

- [1] Baron, R. and Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51(6):1173.
- [2] Fisher, J. O., Birch, L. L., and Smiciklas-Wright (2000). Breast-feeding through the first year predicts maternal control in feeding and subsequent toddler energy intakes. *Journal of the American Dietetic Association*, 100(6):641–646.
- [3] Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by browne and draper). *Bayesian Anal.*, 1(3):515–534.
- [4] Gelman, A., Jakulin, A., Pittau, M. G., and Su, Y.-S. (2007). A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*, 2.
- [5] Hanson, T. E., Branscum, A. J., and Johnson, W. O. (2014). Informative g -priors for logistic regression. *Bayesian Anal.*, 9(3):597–612.
- [6] Hayes, A. (2013). *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach*. Methodology in the Social Sciences Series. Guilford Publications.
- [7] Imai, K., Keele, L., and Tingley, D. (2010a). A general approach to causal mediation analysis. *Psychological Methods*, 15(4):309–334.
- [8] Imai, K., Keele, L., Tingley, D., and Yamamoto, T. (2010b). Causal mediation analysis using r. In Vinod, H. D., editor, *Advances in Social Science Research Using R*. Springer-Verlag, New York.
- [9] Launay, T., Philippe, A., and Lamarche, S. (2015). Construction of an informative hierarchical prior for a small sample with the help of historical data and application to electricity load forecasting. *TEST*, 24(2):361–385.

- [10] MacKinnon, D. P. (2008). *Introduction to Statistical Mediation Analysis*. Taylor & Francis Group, Routledge.
- [11] Marin, J.-M. and Robert, C. (2014). *Bayesian essentials with R*. Springer Textbooks in Statistics. Springer Verlag, New York.
- [12] Pearl, J. (2001). Direct and indirect effects. In *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence, UAI'01*, pages 411–420, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.
- [13] R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- [14] Robins, J. and Greenland, S. (1992). Identifiability and exchangeability for direct and indirect effects. *Epidemiology*, 3:143–155.
- [15] Stan Development Team (2018). RStan: the R interface to Stan. R package version 2.18.1.
- [16] Tingley, D., Yamamoto, T., Hirose, K., Keele, L., and Imai, K. (2014). mediation: R package for causal mediation analysis. *Journal of Statistical Software*, 59((5)):1–38.
- [17] VanderWeele, T. (2015). *Explanation in Causal Inference: Methods for Mediation and Interaction*. Oxford University Press.
- [18] Wieder, B. and Ossimitz, M.-L. (2015). The impact of business intelligence on the quality of decision making – a mediation model. *Procedia Computer Science*, 64:1163 – 1171.
- [19] Zellner, A. (1971). *An Introduction to Bayesian Inference in Econometrics*. A Wiley-Interscience publication. Wiley.