HHS Public Access

Author manuscript

Stat Med. Author manuscript; available in PMC 2018 November 20.

Published in final edited form as:

Stat Med. 2017 November 20; 36(26): 4182–4195. doi:10.1002/sim.7423.

Quantile causal mediation analysis allowing longitudinal data

M.-A. Bind[*,1], **T. J. VanderWeele**[2], **J. D. Schwartz**[3], and **B. A. Coull**[4]

[1]Department of Statistics, University of Harvard, One Oxford Street, Science Center, 7th floor, Cambridge, Massachusetts, United States

[2]Department of Epidemiology, Harvard School of Public Health, 677 Huntington Avenue, Boston, Massachusetts, United States

[3]Department of Environmental Health, Harvard School of Public Health, Landmark Center, Suite 415, Boston, Massachusetts, United States

[4]Department of Biostatistics, Harvard School of Public Health, 655 Huntington Avenue, Boston, Massachusetts, United States

Abstract

Mediation analysis has mostly been conducted with mean regression models. With this approach modeling means, formulae for direct and indirect effects are based on changes in means, which may not capture effects that occur in units at the tails of mediator and outcome distributions. Individuals with extreme values of medical endpoints are often more susceptible to disease and can be missed if one investigates mean changes only. We derive the controlled direct and indirect effects of an exposure along percentiles of the mediator and outcome using quantile regression models and a causal framework. The quantile regression models can accommodate an exposure-mediator interaction and random intercepts to allow for longitudinal mediator and outcome. Because DNA methylation acts as a complex "switch" to control gene expression and fibrinogen is a cardiovascular factor, individuals with extreme levels of these markers may be more susceptible to air pollution. We therefore apply this methodology to environmental data to estimate the effect of air pollution, as measured by particle number, on fibrinogen levels through a change in interferon-gamma (*IFN-γ*) methylation. We estimate the controlled direct effect of air pollution on the q^{th} percentile of fibrinogen and its indirect effect through a change in the p^{th} percentile of *IFN-γ* methylation. We found evidence of a direct effect of particle number on the upper tail of the fibrinogen distribution. We observed a suggestive indirect effect of particle number on the upper tail of the fibrinogen distribution through a change in the lower percentiles of the *IFN-γ* methylation distribution.

Keywords

Mediation analysis; Quantile regression; Causal inference; Longitudinal data

[*]Correspondence to: Harvard University, Department of Statistics, Science Center 7th floor, Cambridge, MA, 02138.

1. Introduction

Mediation analysis aims to identify a direct effect of an exposure on an outcome and an indirect effect between the same exposure and outcome via a change in a mediator [1]. Scientists have recently developed and used mediation analysis tools in many disciplines, such as causal inference [2–6], social sciences [7,8], and epidemiological research [9–12]. The total effect decomposition (i.e., into a indirect effect via a mediator and a direct effect not through this mediator) allows one to investigate biological pathways and thus, disease mechanisms [13–15].

Mediation analysis has largely focused on mean regression models to define controlled direct, natural direct, and natural indirect effects. These estimands may not reflect effects that occur primarily in the tails of the distributions of the mediator and the outcome. We have illustrated that standard regression models for the outcome mean could miss effects in susceptible subgroups, that is, individuals with already extreme levels of mediator and/or outcome that are risk factors for disease [12, 16, 17]. Quantile regression is an estimation method that addresses this issue. If one chooses regression coefficients that minimize the sum of the absolute values of the residuals instead of the sum of squared residuals, the result is an estimate of covariate effects on the median, instead of the mean, of the outcome distribution. Quantile regression generalizes this approach by weighting the positive and negative residuals differently, which forces the regression line to other percentiles of the distribution, which is useful to examine the exposure-outcome relationship at different locations of the outcome distribution, especially when this relationship is not homogenous across quantiles of the outcome. This distribution-free approach, modeling quantiles of a distribution instead of the mean, has already been used successfully to describe: 1) how air pollution effects on cardiovascular markers are different across the distribution of these same cardiovascular markers [17], and 2) how air pollution and temperature exposures change the shape of DNA methylation distributions, an epigenetic outcome [12, 16]. In these recent papers, the authors identified susceptible subgroups and therefore allows one to provide more precise results for risk assessment in biomedicine.

Recent research has also identified epigenetics as an important molecular response to environmental pollutants [18–20]. Epigenetics refers to chromosome changes that influence gene expression without modifying the genetic code. The most frequently studied epigenetic mechanism is DNA methylation which involves methylation of cytosine in CpG pairs. Because DNA methylation often acts as a complex “switch” to control gene expression, individuals with extreme methylation level may be more susceptible to environmental exposures. This result has been suggested recently with air pollution [12] and temperature [16]. DNA methylation has been related to cardiovascular and pulmonary outcomes [21,22], that have, in turn, been associated with exposure to air pollution [23]. These findings raised interest for examining DNA methylation as an intermediate player in air pollution adverse responses [24,25].

The motivation of this paper is to study the mediated effects of air pollution on an cardiovascular marker via DNA methylation, but across the distributions of: 1) the mediator and 2) the outcome. To estimate these effects, we need to expand standard mediation

methods and consider quantiles of mediators and outcomes as dependent variables. Mediation formulae based on quantiles of the outcome distribution have been derived [14, 26]. However, these formulae focuses mostly on mean changes in the mediator [14] or do not explicitly define the causal estimands as functions of potential outcomes [26]. In this manuscript, we define and derive the controlled direct and the indirect effects along the distribution of the outcome with two quantile regression models, one for the mediator and one for the outcome. We formalize our estimands using the potential outcome framework and allow for an exposure-mediator interaction in the model for the outcome. Because many observational studies collect longitudinal data, we also allow the mediator and outcome variables to be repeated measurements. This work generalizes previous work on mediation models now allowing for quantile estimation and longitudinal data [14,26–28].

Using a causal framework, we apply the quantile mediation model to an environmental health example in which we examine the effects of air pollution and methylation on the entire distribution of fibrinogen, a known coagulation marker. We hypothesize that individuals with higher fibrinogen levels may be more susceptible to air pollution and changes in methylation. In this manuscript, we estimate the indirect effect of particle number exposure on the fibrinogen distribution via a distributional change in interferon-gamma (*IFN-γ*) methylation and its direct effect not through *IFN-γ* methylation. While particle number is a marker for traffic-related air pollution, fibrinogen and IFN-γ are known to be cardiovascular markers.

2. Methods

2.1. Notations

This paper aims to define and derive the direct exposure effect on the distribution of an outcome and its indirect effect via a change in a given percentile of a mediator. We allow exposure-mediator interaction and repeated mediator and outcome measurements from a longitudinal cohort. Let A_{ij}, M_{ij}, and Y_{ij} represent the observed exposure, mediator, and outcome of interest for an individual i at a visit j, respectively. Let $\psi_p(M_{ij})$ be the p^{th} percentile of the observed M_{ij} distribution and let $\zeta_q(Y_{ij})$ be the q^{th} percentile of the observed Y_{ij} distribution. Let M_{ij}^a denote the potential mediator that would have been observed if A were set to a and let $Y_{i,j}^{a,m}$ represent the potential outcome that would have been observed if A were set to a and M were set to the value m. Let $\Psi_p(M_{ij}^a)$ be the p^{th} percentile of the distribution of the potential outcome M_{ij} distribution when A_{ij} is set to a and let $\xi_q(Y_{ij}^{a,m})$ denote the q^{th} percentile of the distribution of the potential outcome Y_{ij} distribution when A_{ij} is set to a and M_{ij} is set to m. Thus, $\xi_q\left(Y_{ij}^{a,\Psi_p\left(M_{ij}^{a^*}\right)}\right)$ corresponds to the q^{th} percentile of the potential outcome distribution that would have been observed if A_{ij} were set to a and if M_{ij} were set to the value of the p^{th} percentile of the distribution of the potential outcome M_{ij} if A_{ij} were set to a^*. Contrasts of these percentiles will be used to define direct and indirect effects of exposure along the outcome distribution. Because the exposure and mediator are generally not randomized in observational studies, we define the

distributions in terms of quantiles conditional on covariates, and conditional on random intercepts to allow longitudinal mediator and outcome.

2.2. Quantile regression models with random intercepts

We consider two quantile regression models that include covariates, random intercepts b_{0i} and g_{0i} (not necessarily independent of each other), and an exposure-mediator interaction for the outcome model. C_{ij} corresponds to the sets of covariates included in the mediator and outcome models:

$$\psi_p(M_{ij}|Cij{=}c, A_{ij}{=}a, b_{0i}){=}(\beta_0{+}b_{0i}){+}\beta_1 a{+}\beta_c^T c$$

$$\zeta_q(Y_{ij}|C_{ij}{=}c, A_{ij}{=}a, M_{ij}{=}m, g_{0i}){=}(\gamma_0{+}g_{0i}){+}\gamma_1 a{+}\gamma_2 m{+}\gamma_3 am{+}\gamma_c^T c.$$

We note that the parameters depend on the chosen quantiles p and q. For simplicity in the proof (i.e., derivation of the estimands), the coefficients β's and γ's introduced above will not carry the quantile-specific indices p and q, respectively. Later on, when presenting the motivating example, the coefficients of interest β_1, γ_1, γ_2, will be denoted with their associated quantile-specific indices, that is, $\beta_{1,p}$, $\gamma_{1,q}$, $\gamma_{2,q}$.

2.3. Identification assumptions

Four assumptions will suffice to identify the direct and indirect effects on the outcome distribution that we will present later:

1. $Y_{ij}^{a,m} \coprod A_{ij}|C_{ij}{=}c, b_{0i}, g_{0i}$
2. $Y_{ij}^{a,m} \coprod M_{ij}|A_{ij}{=}a, C_{ij}{=}c, b_{0i}, g_{0i}$
3. $M_{ij}^{a} \coprod A_{ij}|C_{ij}{=}c, b_{0i}, g_{0i}$
4. $Y_{ij}^{a,m} \coprod M_{ij}^{a*}|C_{ij}{=}c, b_{0i}, g_{0i},$

where the symbol $\coprod$ conveys the independence between two random variables.

Conditional on the measured covariates C_{ij} and the random intercepts, the first assumption is met if there is no unmeasured exposure-outcome confounding. The second is met if there is no unmeasured mediator-outcome confounding. The third holds if there is no unmeasured exposure-mediator confounding, and the fourth holds if there is no mediator-outcome confounder affected by the exposure. Similar assumptions have been presented in a recent paper presenting identification assumptions that hold conditionally on random effects, but not unconditionally [28].

Although assumption 2 requires that we control for all mediator-outcome confounders, assumption 4 requires that the exposure does not affect any of these confounders that we adjusted for in assumption 2. In principle, we could have a set of mediator-outcome confounders that included variables affected by the exposure in which case assumption 4

would not hold and thus, it is essential to check whether the exposure affects the confounders of the mediator-outcome relationship.

Because unmeasured confounding is difficult to rule out, complementary analyses can be performed. Researchers can: 1) consider subsets of potential confounding variables and assess whether the resulting conclusions change when controlling for them, 2) perform some sensitivity analyses that estimate "how much" confounding is required to reverse the results, or 3) add a design stage involving matching strategies in order to create balance in covariates in the exposed vs. unexposed groups.

2.4. Assumptions of no time-varying confounding and no interference

We assume exogeneity with respect to the exposure and the mediator. Time-varying confounding in this setting has already been described [28]. Briefly, time-varying confounding would occur if there were an effect of M_{ij} or Y_{ij} on a subsequent measurement of $A_{ij'}$ or an effect of A_{ij} or Y_{ij} on a subsequent measurement of $M_{ij'}$ (j' > j). This would likely be plausible in environmental health context with exogenous exposure (e.g., air pollution), as shown in a recent mediation analysis investigating the effects of air pollution on inflammation via changes in DNA methylation [28]. In addition, we assume no interference between units *i*.

2.5. Estimands

In the Appendix, we show that, under the mixed-effects models we have considered, $\xi_q\left(Y_{ij}^{a,\Psi_p\left(M_{ij}^{a^*}|C_{ij}=c,b_{0i},g_{0i}\right)}|C_{ij}=c,b_{0i},g_{0i}\right)$, the q^{th} percentile of the potential outcome Y_{ij} distribution setting A_{ij} to a and M^{ij} to $\Psi_p(M_{ij}^{a^*})$ conditioning on the covariates and random intercepts, is equal to $\gamma_0+g_{0i}+\gamma_1 a+(\gamma_2+\gamma_3 a)(\beta_0+b_{0i}+\beta_1 a^*+\beta_c^T c)+\gamma_c^T c$.

2.5.1. Controlled direct effect—Recall that the controlled direct effect (CDE) of exposure A on outcome Y comparing $A = a$ with $A = a^*$ and setting M to m can be obtained by comparing $Y^{a,m}$ to $Y^{a^*,m}$. We define and derive the controlled direct effect of the exposure on the q^{th} percentile of the outcome Y_{ij} distribution by contrasting two percentiles of the potential outcome distributions $\xi_q(Y_{ij}^{a,m})$, one setting A_{ij} to a and the other one setting A_{ij} to a^*; while holding M_{ij} to the same value $\Psi_p(M_{ij}^{a^*})$:

$$\xi_q\left(Y_{ij}^{a,\Psi_p\left(M_{ij}^{a^*}|C_{ij}=c,b_{0i},g_{0i}\right)}|C_{ij}=c,b_{0i},g_{0i}\right)-\xi_q\left(Y_{ij}^{a^*,\Psi_p\left(M_{ij}^{a^*}|C_{ij}=c,b_{0i},g_{0i}\right)}|C_{ij}=c,b_{0i},g_{0i}\right).$$

This direct effect captures the effect on the q^{th} quantile of the distribution of Y when changing the exposure from a^* to a, but in both cases fixing the mediator to the p^{th} quantile of the distribution of the mediator when exposure is set to a^*. Under the quantile regression models we have considered and in the absence of exposure-mediator interaction, this is: $\gamma_1(a - a^*)$. In the absence of exposure-mediator interaction, the direct effect does not depend on the quantile p, that is, this is the same for all mediator quantiles. In the presence of exposure-mediator interaction, the contrast between

$$\xi_q\left(Y_{ij}^{a,\Psi_p\left(M_{ij}^{a^*}|C_{ij}=c,b_{0i},g_{0i}\right)}|C_{ij}=c,b_{0i},g_{0i}\right)-\xi_q\left(Y_{ij}^{a^*,\Psi_p\left(M_{ij}^{a^*}|C_{ij}=c,b_{0i},g_{0i}\right)}|C_{ij}=c,b_{0i},g_{0i}\right)$$

depends on the individual random effects: $\gamma_1(a-a^*)+\gamma_3(a-a^*)(\beta_0+b_{0i}+\beta_1 a+\beta_c^T c)$. To obtain an estimate of the population controlled direct effects, the random effects "b_{0i}" can be integrated out. If researchers are interested in individual-level controlled direct effects, quantile mixed-effects models can be considered to estimate the random effects b_{0i} (code not shown). The natural direct effect (NDE) of exposure A on outcome Y comparing $A = a$ with $A = a^*$ intervening to set M to what it would have been if exposure had been $A = a^*$ is generally obtained by comparing $Y^{a,M^{a^*}} - Y^{a^*,M^{a^*}}$. Because of the focus on specific mediator and outcome quantiles, the natural direct effect (in its usual definition) cannot be defined.

2.5.2. Indirect effect—We now define the indirect effect on the q^{th} percentile of the potential outcome Y_{ij} distribution through the p^{th} percentile of the M_{ij} distribution as the contrast between the percentiles of the potential outcome distribution $\xi_q(Y_{ij}^{a,m})$, one setting M_{ij} to $\Psi_p\left(M_{ij}^{a}\right)$ and the other setting M_{ij} to $\Psi_p\left(M_{ij}^{a^*}\right)$; while holding A_{ij} to a:

$$\xi_q\left(Y_{ij}^{a,\Psi_p\left(M_{ij}^{a}|C_{ij}=c,b_{0i},g_{0i}\right)}|C_{ij}=c,b_{0i},g_{0i}\right)-\xi_q\left(Y_{ij}^{a,\Psi_p\left(M_{ij}^{a^*}|C_{ij}=c,b_{0i},g_{0i}\right)}|C_{ij}=c,b_{0i},g_{0i}\right).$$

This indirect effect captures the effect on the q^{th} quantile of the potential outcome Y_{ij} distribution, when the exposure is fixed to a, but the mediator is changed from the p^{th} quantile of mediator when the exposure is a^* to the p^{th} quantile of the mediator distribution when the exposure is a. Under the mixed-effects effects model we have considered, this is: $(\gamma_2 + \gamma_3 a)\,\beta_1\,(a - a^*)$. If there is no exposure-mediator interaction, this is equal to the standard "mediation formula", $\gamma_2\,\beta_1\,(a - a^*)$. Note that these indirect effects depend on both quantiles p and q. Similarly as before, the natural indirect effect (NIE) comparing $A = a$ with $A = a^*$ and intervening to set exposure A to a is generally defined by comparing Y^{a,M^a} to $Y^{a,M^{a^*}}$. Note that, because of our focus on specific quantiles, the natural indirect effect (in its usual definition) cannot be defined here.

2.5.3. Additional remarks—We note that the controlled direct and indirect effects we have defined above do not add up to the (usual) total effect (of the exposure on the q^{th} percentile of the outcome distribution) resulting from the difference between $\xi_q(Y_{ij}^{a})$ and $\xi_q(Y_{ij}^{a^*})$. Instead, the sum of the two effects is equal to the contrast:

$$\xi_q\left(Y_{ij}^{a,\Psi_p\left(M_{ij}^{a}|C_{ij}=c,b_{0i},g_{0i}\right)}|C_{ij}=c,b_{0i},g_{0i}\right)-\xi_q\left(Y_{ij}^{a^*,\Psi_p\left(M_{ij}^{a^*}|C_{ij}=c,b_{0i},g_{0i}\right)}|C_{ij}=c,b_{0i},g_{0i}\right).$$

This contrast can be decomposed into sum of the controlled direct and indirect effects we have defined, which is not a quantity that has been of interest in causal inference, but could be defined as a new type of "total effect".

The framework described in this manuscript is not restricted to models with random intercepts only. It can be extended to models including random intercepts and slopes. For example, one can add two random slopes b_{1i} and g_{1i} that allow for heterogeneous exposure and mediator effects across subjects, respectively. In this case, the direct and indirect effects

(conditional on covariates and random effects) can be obtained by replacing β_1 and γ_1 by $\beta_1 + b_{1i}$ and $\gamma_1 + g_{1i}$, respectively.

The method developed here applies to cross-sectional data as well by setting the variance of the random intercept b_{0i} and g_{0i} to zero. In such a setting, the controlled direct effect and and the indirect effect still equals $\gamma_1 (a - a^*)$ and $\gamma_2 \beta_1 (a - a^*)$, respectively. In situations with missing covariates, we recommend to multiply-impute them before conducting the quantile mediation analysis. If variables are measured with differential error, we recommend a recent developed method relying on sensitivity parameters [29].

2.6. Variance of the control direct and indirect effects

If there is no exposure-mediator interaction in the outcome model, the variance of the direct effect (γ_1) can be easily obtained from the software output. The indirect effect formula in the absence of exposure-mediator interaction consists of a product term $\gamma_2\beta_1$. The bivariate delta method can be used to derive $Var(\gamma_2\beta_1)$ which is approximately equal to $\sigma^2_{\beta_1}\gamma_2^2+2\sigma_{\beta_1,\gamma_2}\gamma_2\beta_1+\sigma^2_{\gamma_2}\beta_1^2$. The Bootstrap procedure can also be used to estimate the variances of the indirect effects in complex settings. The Bootstrap technique is useful to estimate variances non-parametrically, especially with quantile regression models allowing for longitudinal data [30], with mediation models with exposure-mediator interactions [28], or in more complex situations in which asymptotic theory does not apply. In longitudinal settings, the observations should be sampled with replacement by subject in order to preserve the correlation structure of the dataset [31].

2.7. Estimation

We used the R package *rqpd* to estimate the coefficients of the quantile regressions for longitudinal data [30]. We used the Bootstrap procedure to estimate the variance of the direct and indirect effects on the outcome distribution. We provide software in the form of R code in the Appendix.

Similarly as in mean mediation analysis for longitudinal data [28], the random effects could be dependent with each other. For example, in a situation in which random slopes for the mediator and outcome models are correlated, these two models (i.e., mediator and outcome models) should be estimated jointly. Sophisticated software should be considered.

3. Motivating example: Air pollution, DNA methylation, and fibrinogen

3.1. Scientific question

The role of DNA methylation, although very complex, has been summarized as a “switch” that can control gene expression. Recent results have suggested that air pollution distorts the distribution of epigenetic outcomes (e.g., *IFN-γ* methylation), and that larger impacts are observed among individuals whose epigenetic outcomes’ levels are low [12]. These results suggest that participants who already have higher risk of coagulation may also be the ones primarily affected by air pollution. With our developed method, we conduct a quantile mediation analysis to examine this hypothesis, but in a mediation analysis context with a cardiovascular marker as the outcome. Our approach modeling percentiles of mediator and

outcomes can detect whether larger effects are observed at the adverse end(s) of health outcomes' distributions.

Using the quantile causal mediation model described in the previous sections, we investigate whether there is a direct air pollution (i.e., particle number) effect on different quantiles of the outcome distribution of a coagulation marker (i.e., fibrinogen) and whether there is an indirect air pollution effect on this outcome distribution through different quantiles of the epigenetic mediator (i.e., *IFN-γ* methylation).

3.2. Data description

We analyze data from a longitudinal cohort study including participants from the Normative Aging Study. This investigation, described in a previous paper [32], was established in Boston in 1963 by the U.S. Veterans Administration. Between 1999 and 2009, a total of 777 participants had their levels of DNA methylation and fibrinogen measured one to five times with intervals of three to five years. Participants blood was collected at every medical visit after an overnight fast and smoking abstinence.

The exposure of interest is particle number, which is commonly used as a surrogate of ultrafine particles from traffic. Hourly particle number concentrations were measured with a Condensation Particle Counter (TSI Inc, Model 3022A, Shoreview, MN) 1 km away from the medical center. The intermediary mechanism of interest is DNA methylation on the *IFN-γ* gene. IFN-γ is an important cytokine that plays a role in innate and adaptive immunity against foreign compounds. *IFN-γ* methylation was assessed with highly quantitative methods based on bisulfite polymerase chain reaction (PCR) pyrosequencing. The outcome of interest is fibrinogen, which is a precursor of fibrin involved in blood clotting. Plasma fibrinogen concentration was measured using MDA Fibriquick (Trinity Biotech, Bray, Ireland), a bovine thrombin reagent.

The relevant exposure windows for the air pollution effects on DNA methylation and fibrinogen are not clearly determined. We consider the intermediate-term exposure window of 28-day moving average preceding each individual medical examination. This time window could serve as a median choice between short- and long-term exposures to air pollution. The medians of the mean distances of the participant homes from the monitor were about 20 km.

3.3. Methods

We standardize the exposure (i.e., particle number), the mediator (i.e., *IFN-γ* methylation), and the outcome (i.e., fibrinogen) of interest and present the controlled direct and the indirect effects as the result of a one-unit increase of the standardized exposure. We fit two separate quantile regressions, one to model the exposure-mediator relationship and another one to model the effects of the exposure and the mediator on the outcome:

$$\psi_p(M_{ij}|C_{ij}=c, A_{ij}=a, b_{0i})=(\beta_0+b_{0i})+\beta_1 a+\beta_c^T c$$

$$\zeta_q(Y_{ij}|C_{ij}=c, A_{ij}=a, M_{ij}=m, g_{0i})=(\gamma_0+g_{0i})+\gamma_1 a+\gamma_2 m+\gamma_c^T c.$$

Using the same method described in a previous paper [12], we construct an alternative way of presenting the estimated coefficients from the quantile regressions by illustrating the distributional shift associated with a two-unit increase in standardized exposure or mediator. The predicted curve, constructed using the quantile regression coefficients, assumes a constant trend within decile intervals.

We compare our quantile mediation analysis results to what we would have found using a standard mediation analysis modeling the means of the mediator and outcome:

$$E(M_{ij}|C_{ij}=c, A_{ij}=a, b_{0i})=(\beta_0+b_{0i})+\beta_1 a+\beta_c^T c$$

$$E(Y_{ij}|C_{ij}=c, A_{ij}=a, M_{ij}=m, g_{0i})=(\gamma_0+g_{0i})+\gamma_2 m+\gamma_c^T c.$$

Note that in the regressions modeling the outcome mean, the exposure-mediator interaction was not significant (p-value$_{interaction}$=0.88) and its addition did not change the effect estimates. We therefore fit all mediation models with no exposure-mediator interaction.

Our four assumptions about confounding in Section 2.3 need to hold for the controlled direct and indirect effects to be identified from the data. Therefore, we included “a priori” the following covariates C_{ij} in the regression models: temperature, relative humidity, seasonal sine and cosine (to allow the regression analysis to estimate both the amplitude and the phase of the seasonal cycle), batch of DNA methylation measurement, age, body mass index (BMI), smoking, diabetes, statin use, percentage of neutrophils in blood count, percentage of lymphocytes in blood count, percentage of monocytes in blood count, and percentage of basophils in blood count.

As in a recent mean mediation analysis [28], we assume no endogeneity. For instance, we assume that the mediator (i.e., *IFN-*γ methylation) and outcome (i.e., fibrinogen) measured at previous visits do not affect the concentrations of particle number at any subsequent visit; and that particle number and fibrinogen measured at previous visits do not affect *IFN-*γ methylation at any subsequent visit. This is plausible because air pollution is an exogenous variable and medical visits occur far apart, i.e., three to five years.

3.4. Results

3.4.1. Direct effect—We found evidence of heterogeneity in the direct effect of particle number on fibrinogen across the outcome distribution (Figures 1.1 and 1.2). Figure 1.1 represents the effect of (standardized) particle number on each decile of the fibrinogen distribution. Figure 1.2 reports the predicted distributional change on fibrinogen per two-unit increase in the standardized exposure. Only participants with higher fibrinogen levels were significantly affected by particle number exposure, which tends to increase their fibrinogen

levels even further. For instance, for a one-unit increase in standardized particle number concentrations, there is a 0.29 [95%$CI_{Bootstrap}$: 0.08; 0.44] increase in fibrinogen levels among participants belonging to the 90^{th} percentile, although we observed null associations among participants belonging to the lower half of the distribution (see coefficients $\gamma_{1,q}$ in Table 1).

3.4.2. Indirect effect—We followed our investigation by decomposing the mediated effect in two parts: we first examined the effect of the exposure on the mediator distribution (see coefficients $\beta_{1,p}$ in Table 1) and then investigated the effect of the mediator on the outcome distribution (see coefficients $\gamma_{2,q}$ in Table 1).

Exposure-mediator relationship: We observed heterogenous effects of particle number exposure on the *IFN-γ* methylation mediator distribution across deciles (Figures 2.1 and 2.2). Only participants with low methylation levels seem to be affected by ultrafine particles. For instance, a one-unit increase in standardized particle number was associated with a 0.19 decrease in the 10^{th} decile of the standardized *IFN-γ* methylation and was not significantly associated with the upper deciles (i.e., 70^{th} to 90^{th}) of standardized *IFN-γ* methylation.

Mediator-outcome relationship: The effect of the *IFN-γ* methylation mediator on the fibrinogen outcome was less clear (Figures 3.1 and 3.2), but Figure 3.1 suggests a negative impact of *IFN-γ* hypomethylation among participants belonging to the 90^{th} percentile of the fibrinogen distribution. A decrease in *IFN-γ* methylation among these participants was marginally associated with an increase in fibrinogen. Therefore, for those with lower levels of *IFN-γ* methylation and higher levels of fibrinogen, an increase in particle number will generally further decrease their *IFN-γ* methylation levels, resulting in an increase in fibrinogen.

This sequence of effects among participants with lower *IFN-γ* methylation and higher fibrinogen is consistent with the positive association between exposure to particle number and fibrinogen at the higher quantiles of the fibrinogen distribution (as seen in Figure 1.1).

Quantile mediated effects: We then calculated the indirect effects of particle number on the fibrinogen distribution through three percentiles (20^{th}, 50^{th}, and 80^{th}) of the *IFN-γ* DNA methylation distribution (Figures 4.1 to 4.3). We calculated 95% confidence intervals via the Bootstrap procedure we described in section 2.6. Although the quantile-specific indirect effects were not strictly significant, Figures 4.1 and 4.2 suggest some indirect effect among participants in the 90^{th} percentile of fibrinogen having also low *IFN-γ* methylation. We observed no indirect effect of particle number on fibrinogen via *IFN-γ* methylation among participants in the upper percentile of *IFN-γ* methylation (Figure 4.3).

We also present the indirect effects of particle number on the q^{th} percentile of the fibrinogen distribution through the p^{th} percentile of the *IFN-γ* DNA methylation distribution on a heat map (Figure 5.1) and a 3D-plot (Figure 5.2). We observe that participants with low levels of *IFN-γ* methylation and high levels of fibrinogen have the largest indirect effects of particle number on fibrinogen via a decrease in *IFN-γ* methylation.

Mean mediated effects vs. quantile mediated effects: We contrasted our findings using quantile regression to what one would obtain if a standard mean regression analysis was conducted (see coefficient estimates from the mean models in Table 1). We used a mean mediation model for longitudinal data [28]. We fit two mixed-effects, one modeling the mediator mean, the second modeling the outcome mean. We control for the same covariates we adjusted for in the quantile mediation analysis. The direct effect of particle number (standardized) on the mean of fibrinogen (standardized) was equal to 0.16 [95%CI: 0.05 to 0.26]. The indirect effect of a one unit increase in standardized particle number on the mean of standardized fibrinogen via a change in the mean standardized *IFN-γ* DNA methylation was equal to 0.00 [95%CI: −0.01 to 0.02], while the estimated indirect effect is equal to 0.03 [95%CI: −0.01 to 0.06] at the 20^{th} quantile of *IFN-γ* methylation and the 90^{th} quantile of fibrinogen.

4. Simulation study

We present a simulation study contrasting: 1) two scenarios for which the effects are only present at the extreme quantiles of the simulated study population, vs. 2) a scenario for which the effects are present for a larger portion of the population (See supplementary materials).

5. Discussion

This approach extends previous work that defined direct and indirect effects using quantile regressions [4, 26]. Using a causal framework, we considered two quantile regression models and formally derived the controlled direct exposure effect along the outcome distribution and its indirect effect via a given percentile of the mediator using potential outcomes. Our approach allows for exposure-mediator interaction in the outcome model and allows for longitudinal setting with repeated measurements of mediator and outcome. With this methodology, we present an alternative method (vs. the standard mean mediation analysis) that investigates shifts in the distributions of the mediator and outcome, rather than just shifts in the means of these distributions. We are able to detect interesting patterns of particle number health effects that would be missed using ordinary mean mediation analysis. This research tool is valuable for many disciplines, especially in epidemiological studies that aim to investigate pathways and determine susceptibility.

The direct and indirect effects we derived based on specific quantiles of the mediator and outcome distributions may be more difficult to interpret as compared to those estimated in a mean mediation analysis are conducted. The controlled direct effect on the q^{th} percentile of the outcome Y_{ij} distribution, defined by contrasting $\xi_q(Y_{ij}^{a,m})$, one setting A_{ij} to a and the other one setting A_{ij} to a^*; while holding M_{ij} to the same value $\Psi_p(M_{ij}^{a^*})$ may not generally be of interest if the p represents a quantile at the tail of the distribution, unless the direct effect occurs for participants belonging to the tail of the mediator. Other quantities of interest could be considered. For example, one could consider a mean regression model for the mediator and a quantile regression model for the outcome. If the median and mean of the mediator are close to each other, this analysis should provide similar estimates to what one

would obtain by fitting a median regression for the mediator and the same quantile regression model for the outcome.

Many disciplines have shown interest in quantile regression [12, 16, 17] and mediation analysis to investigate existant or new biological mechanisms, as well as their relative importance [4–6, 8–11]. These biological findings are relevant to the understanding of air pollution adverse effects. They provide evidence of an impact of ultrafine particles on *IFN-γ* methylation followed by a potential increase in coagulation. Such findings are relevant to several fields, such as molecular biology, environmental epidemiology, and risk assessment. These results also add evidence towards the intermediate role that DNA methylation may play in air pollution effects on cardiovascular disease and provide evidence towards susceptibility to air pollution (i.e., ultrafine particles) based on levels of mediator (i.e., DNA methylation) and outcome (i.e., fibrinogen).

We applied the method to an interesting environmental health example and identified cardiovascular-related responses with different individual susceptibility. This method can be applied in many environmental studies with exogenous exposure to examine whether the exposure-outcome, exposure-mediator, and mediator-outcome effects are heterogeneous within the study population, and whether certain tails of the distributions of the mediator and outcome are more affected by the exposure and/or the mediator. We highlight a way to examine heterogeneity in responses without making assumptions about the mediator and outcome distributions or their residuals. It was interesting that, under high particle number exposure, the right-tail of the fibrinogen distribution and the left-tail of the *IFN-γ* DNA methylation distribution become longer than would be expected under low exposure conditions. Participants with elevated fibrinogen, often due to clotting dysfunction, were more likely to be affected by particle number exposure and a change in *IFN-γ* methylation. Participants with low *IFN-γ* methylation were also more susceptible to particle number exposure. The key result is that stronger effects were found among participants in the lower tail of the *IFN-γ* methylation distribution and in the higher tail of the fibrinogen distribution. These larger effects were therefore seen in people at the adverse end of the mediator and outcome distributions, which have important public health implications and need to be considered by air pollution health impact assessments.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank Andrea Baccarelli, Tania Kotlov, Petros Koutrakis, Pantel Vokonas, and the participants from the Normative Aging Study. Research reported in this publication was supported by the Ziff fund at the Harvard University Center for the Environment, the Office of the Director, National Institutes of Health under Award Number DP5OD021412, the NIH grants ES000002, ES015172, and CA134294, as well as the EPA grant RD-83479801. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

1. Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. J Pers Soc Psychol. Dec; 1986 51(6): 1173–1182. [PubMed: 3806354]
2. Rubin DB. Practical implications of modes of statistical inference for causal effects and the critical role of the assignment mechanism. Biometrics. Dec; 1991 47(4):1213–1234. [PubMed: 1786315]
3. Rubin DB. Direct and indirect causal effects via potential outcomes. Scandinavian Journal of Statistics. Jun; 2004 31(2):161–170.
4. Imai K, Keele L, Tingley D. A general approach to causal mediation analysis. Psychol Methods. Dec; 2010 15(4):309–334. [PubMed: 20954780]
5. Suzuki E, Yamamoto E, Tsuda T. Identification of operating mediation and mechanism in the sufficient-component cause framework. Eur J Epidemiol. May; 2011 26(5):347–357. [PubMed: 21448741]
6. VanderWeele TJ. A three-way decomposition of a total effect into direct, indirect, and interactive effects. Epidemiology. Mar; 2013 24(2):224–232. [PubMed: 23354283]
7. Cole DA, Maxwell SE. Testing mediational models with longitudinal data: questions and tips in the use of structural equation modeling. J Abnorm Psychol. Nov; 2003 112(4):558–577. [PubMed: 14674869]
8. Hoven H, Siegrist J. Work characteristics, socioeconomic position and health: a systematic review of mediation and moderation effects in prospective studies. Occup Environ Med. Sep; 2013 70(9):663–669. [PubMed: 23739492]
9. Valeri L, Vanderweele TJ. Mediation analysis allowing for exposure-mediator interactions and causal interpretation: theoretical assumptions and implementation with SAS and SPSS macros. Psychol Methods. Jun; 2013 18(2):137–150. [PubMed: 23379553]
10. Vanderweele TJ, Adami HO, Tamimi RM. Mammographic density as a mediator for breast cancer risk: analytic approaches. Breast Cancer Res. Jul.2012 14(4):317. [PubMed: 22838961]
11. Kosma M, Ellis R, Cardinal BJ, Bauer JJ, McCubbin JA. The mediating role of intention and stages of change in physical activity among adults with physical disabilities: an integrative framework. J Sport Exerc Psychol. Feb; 2007 29(1):21–38. [PubMed: 17556774]
12. Bind MA, Coull BA, Peters A, Baccarelli AA, Tarantini L, Cantone L, Vokonas PS, Koutrakis P, Schwartz JD. Beyond the Mean: Quantile Regression to Explore the Association of Air Pollution with Gene-Specific Methylation in the Normative Aging Study. Environ Health Perspect. Mar. 2015
13. VanderWeele TJ. Mediation and mechanism. Eur J Epidemiol. 2009; 24(5):217–224. [PubMed: 19330454]
14. Imai K, Keele L, Tingley D, Yamamoto T. Unpacking the Black Box of Causality: Learning about Causal Mechanisms from Experimental and Observational Studies. American Political Science Review. Nov; 2011 105(4):765–789.
15. Lange T, Vansteelandt S, Bekaert M. A simple unified approach for estimating natural direct and indirect effects. Am J Epidemiol. Aug; 2012 176(3):190–195. [PubMed: 22781427]
16. Bind MA, Coull BA, Baccarelli A, Tarantini L, Cantone L, Vokonas P, Schwartz J. Distributional changes in gene-specific methylation associated with temperature. Environ Res. Oct.2016 150:38–46. [PubMed: 27236570]
17. Bind MA, Peters A, Koutrakis P, Coull B, Vokonas P, Schwartz J. Quantile Regression Analysis of the Distributional Effects of Air Pollution on Blood Pressure, Heart Rate Variability, Blood Lipids, and Biomarkers of Inflammation in Elderly American Men: The Normative Aging Study. Environ Health Perspect. Aug; 2016 124(8):1189–1198. [PubMed: 26967543]
18. Bollati V, Baccarelli A. Environmental epigenetics. Heredity (Edinb). Jul; 2010 105(1):105–112. [PubMed: 20179736]
19. Peluso M, Bollati V, Munnia A, Srivatanakul P, Jedpiyawongse A, Sangrajrang S, Piro S, Ceppi M, Bertazzi PA, Boffetta P, et al. DNA methylation differences in exposed workers and nearby residents of the Ma Ta Phut industrial estate, Rayong, Thailand. Int J Epidemiol. Dec; 2012 41(6): 1753–1760. [PubMed: 23064502]

20. Hou L, Zhang X, Wang D, Baccarelli A. Environmental chemical exposures and human epigenetics. Int J Epidemiol. Feb; 2012 41(1):79–105. [PubMed: 22253299]
21. Castro R, Rivera I, Struys EA, Jansen EE, Ravasco P, Camilo ME, Blom HJ, Jakobs C, Tavares de Almeida I. Increased homocysteine and S-adenosylhomocysteine concentrations and DNA hypomethylation in vascular disease. Clin Chem. Aug; 2003 49(8):1292–1296. [PubMed: 12881445]
22. Kim M, Long TI, Arakawa K, Wang R, Yu MC, Laird PW. DNA methylation as a biomarker for cardiovascular disease risk. PLoS ONE. 2010; 5(3):e9692. [PubMed: 20300621]
23. Brook RD, Rajagopalan S, Pope CA, Brook JR, Bhatnagar A, Diez-Roux AV, Holguin F, Hong Y, Luepker RV, Mittleman MA, et al. Particulate matter air pollution and cardiovascular disease: An update to the scientific statement from the American Heart Association. Circulation. Jun; 2010 121(21):2331–2378. [PubMed: 20458016]
24. Bellavia A, Urch B, Speck M, Brook RD, Scott JA, Albetti B, Behbod B, North M, Valeri L, Bertazzi PA, et al. DNA hypomethylation, ambient particulate matter, and increased blood pressure: findings from controlled human exposure experiments. J Am Heart Assoc. Jun.2013 2(3):e000 212.
25. Bind MA, Lepeule J, Zanobetti A, Gasparrini A, Baccarelli A, Coull BA, Tarantini L, Vokonas PS, Koutrakis P, Schwartz J. Air pollution and gene-specific methylation in the Normative Aging Study: association, effect modification, and mediation analysis. Epigenetics. Mar; 2014 9(3):448–458. [PubMed: 24385016]
26. Shen E, Chou CP, Pentz MA, Berhane K. Quantile Mediation Models: A Comparison of Methods for Assessing Mediation Across the Outcome Distribution. Multivariate Behav Res. 2014; 49(5): 471–485. [PubMed: 26732360]
27. Bauer DJ, Preacher KJ, Gil KM. Conceptualizing and testing random indirect effects and moderated mediation in multilevel models: new procedures and recommendations. Psychol Methods. Jun; 2006 11(2):142–163. [PubMed: 16784335]
28. Bind MA, Vanderweele TJ, Coull BA, Schwartz JD. Causal mediation analysis for longitudinal data with exogenous exposure. Biostatistics. Jan; 2016 17(1):122–134. [PubMed: 26272993]
29. Valeri L, Vanderweele TJ. The estimation of direct and indirect causal effects in the presence of misclassified binary mediator. Biostatistics. Jul; 2014 15(3):498–512. [PubMed: 24671909]
30. Koenker R. Quantile regression for longitudinal data. Journal of Multivariate Analysis. 2004; 91(1):74–89.
31. Moulton L, Zeger S. Analysing Repeated Measures on Generalized Linear Models via the Bootstrap. Biometrics. Jun; 1989 45(2):381–394.
32. Bind MA, Baccarelli A, Zanobetti A, Tarantini L, Suh H, Vokonas P, Schwartz J. Air pollution and markers of coagulation, inflammation, and endothelial function: associations and epigene-environment interactions in an elderly cohort. Epidemiology. Mar; 2012 23(2):332–340. [PubMed: 22237295]

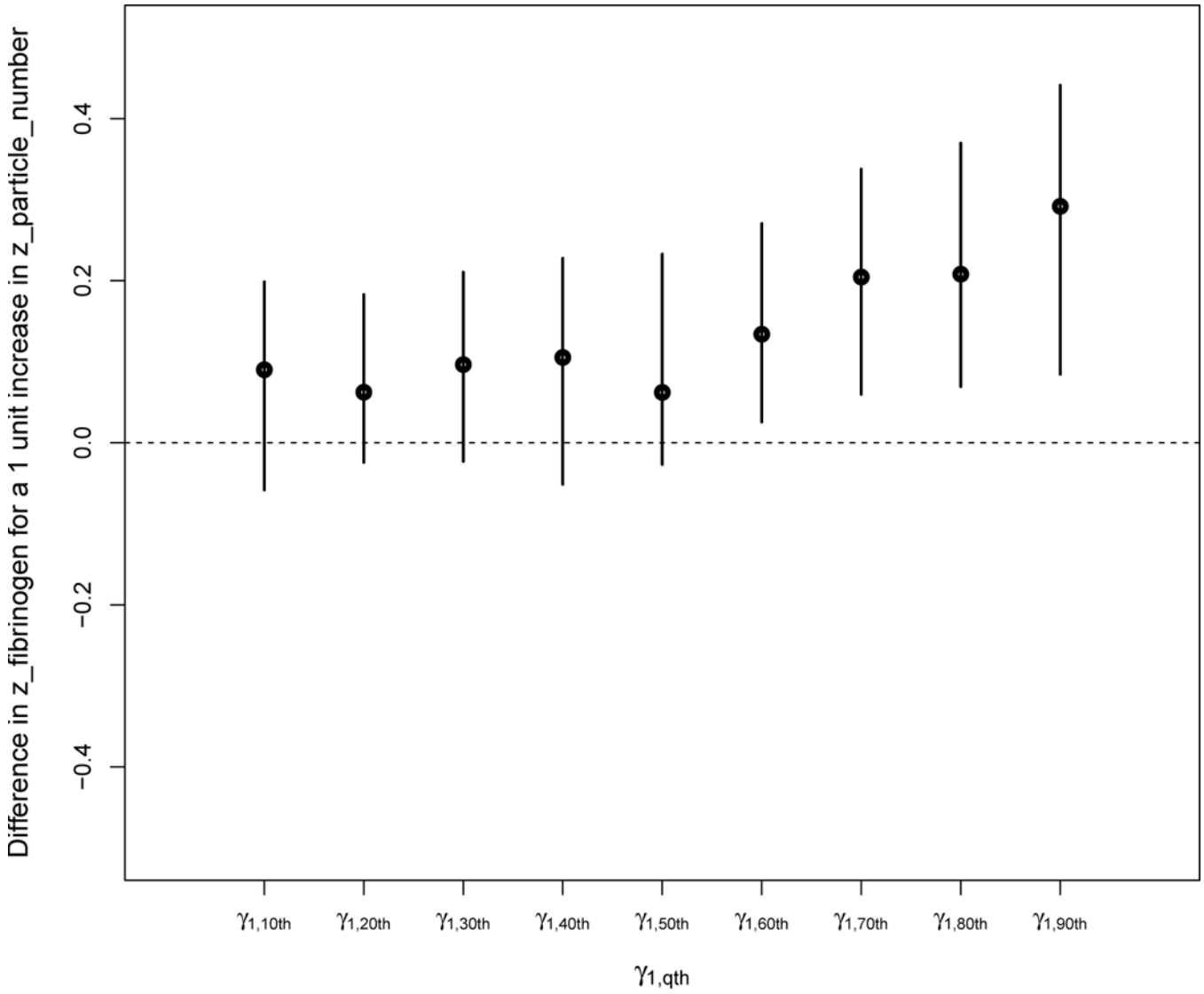

FIGURE 1.1.
Quantile regression coefficients of the associations between particle number exposure and the deciles of the fibrinogen distribution.

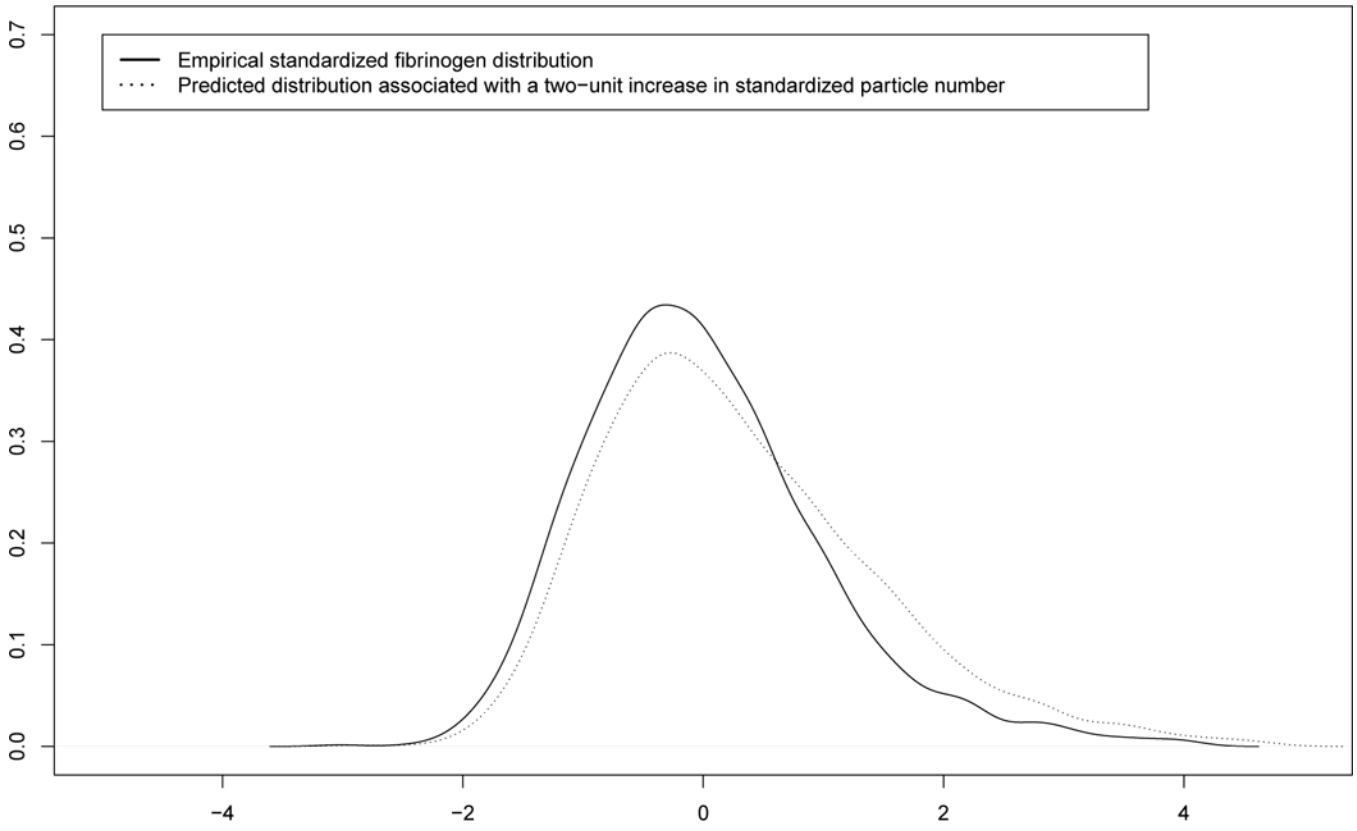

FIGURE 1.2.
Distributional change of the standardized *IFN-γ* methylation distribution due an increase in particle number exposure

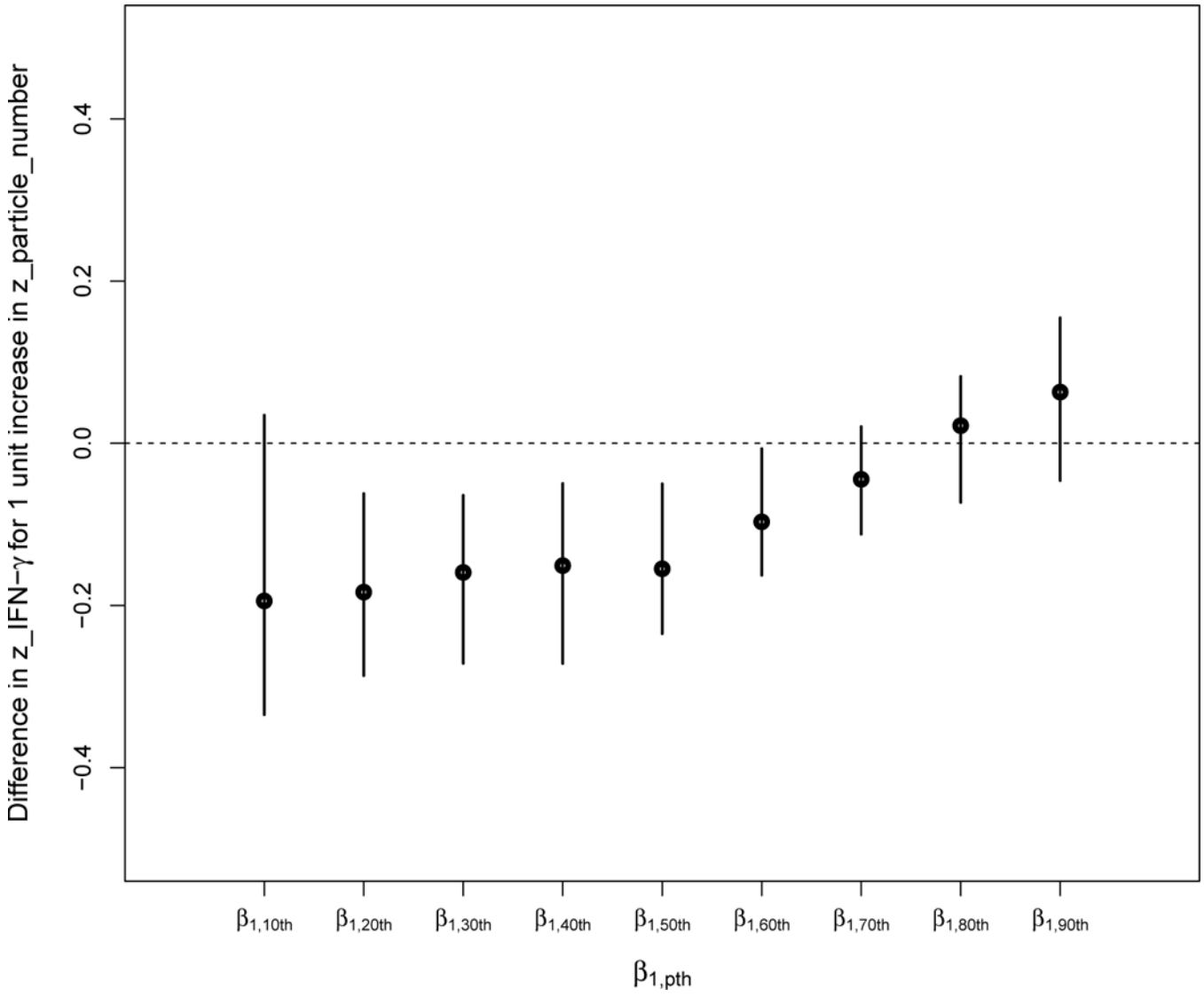

FIGURE 2.1.
Quantile regression coefficients of the associations between particle number exposure and the deciles of the *IFN-γ* methylation distribution.

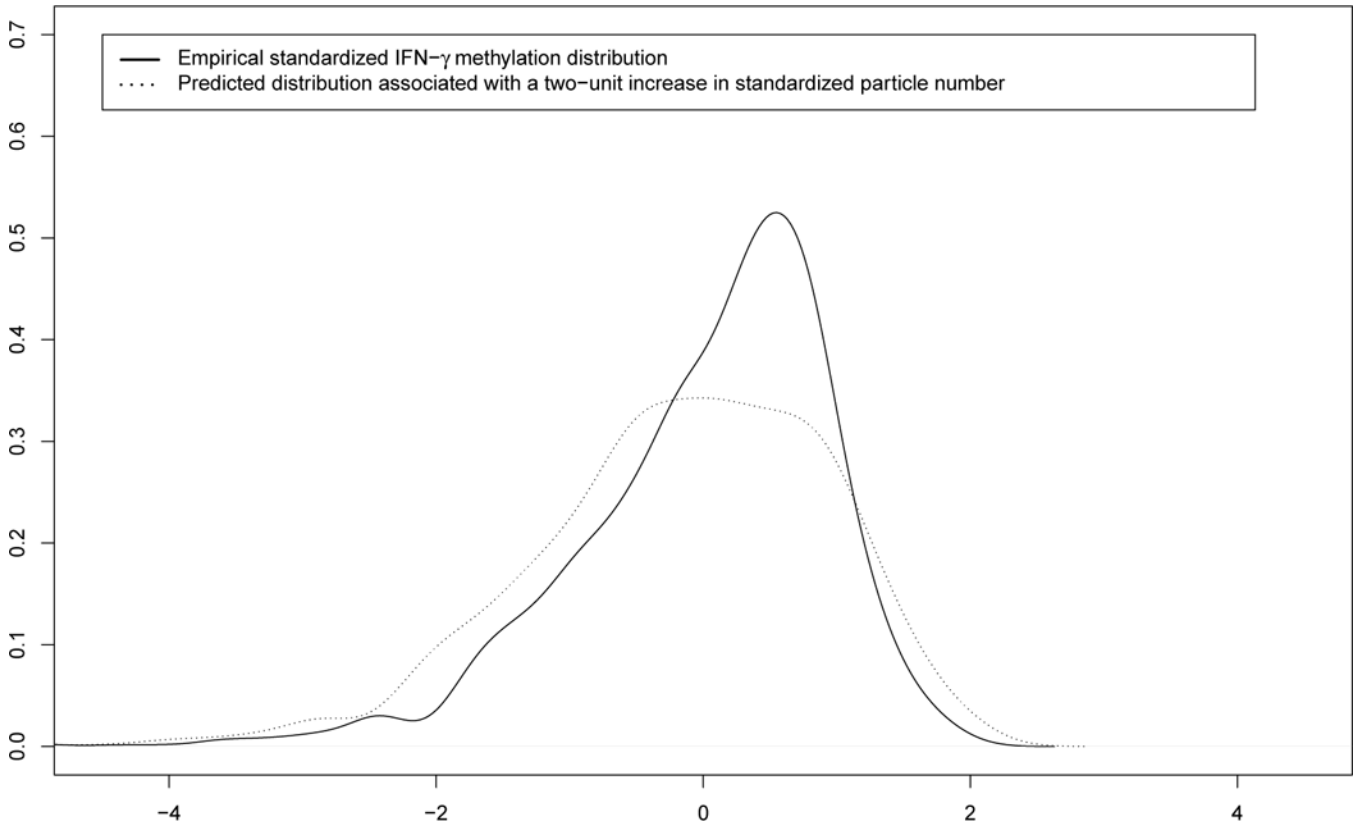

FIGURE 2.2.
Distributional change of the standardized *IFN-γ* methylation distribution due to standardized particle number exposure.

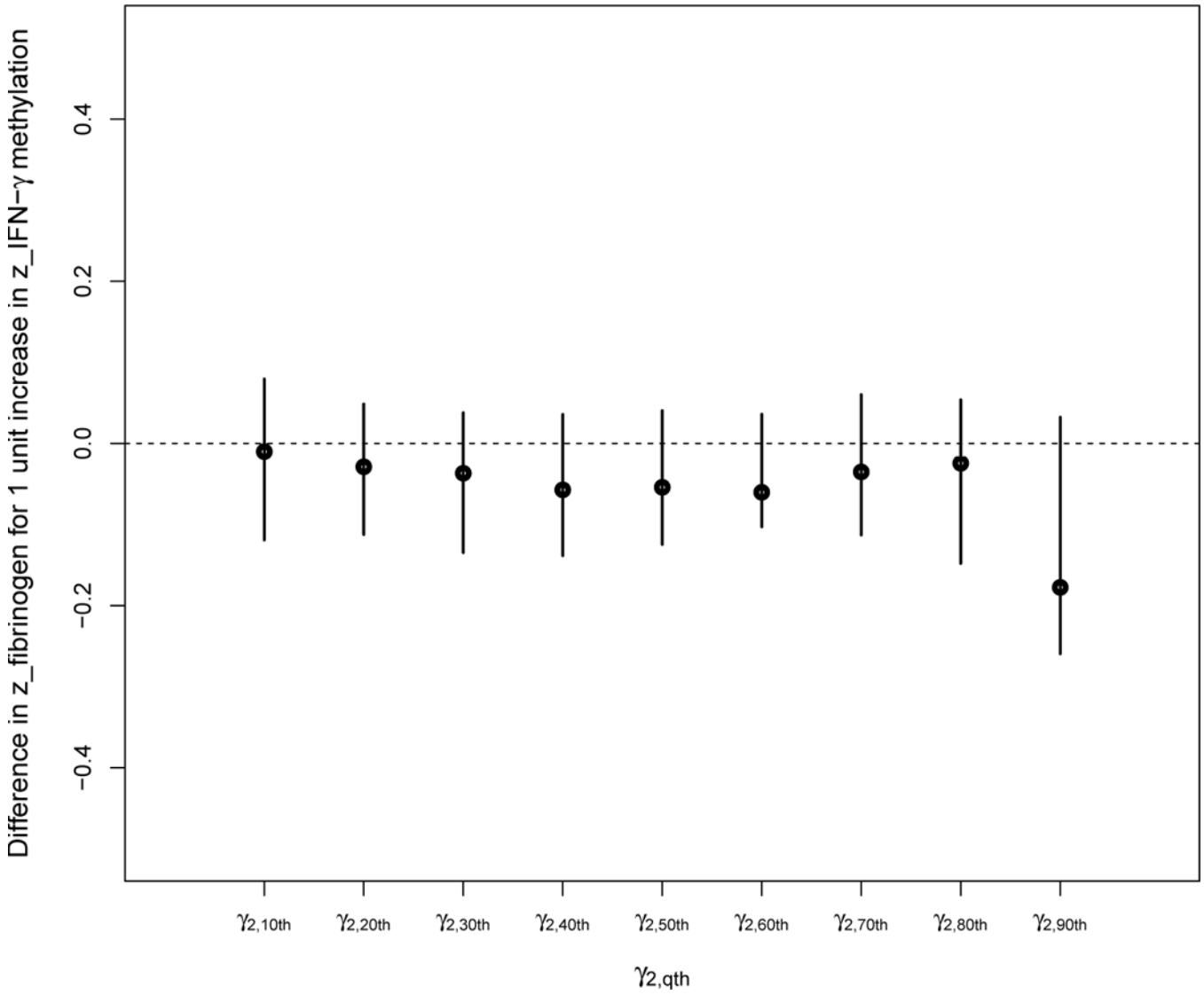

FIGURE 3.1.
Quantile regression coefficients of the associations between *IFN-γ* methylation and the deciles of the fibrinogen distribution.

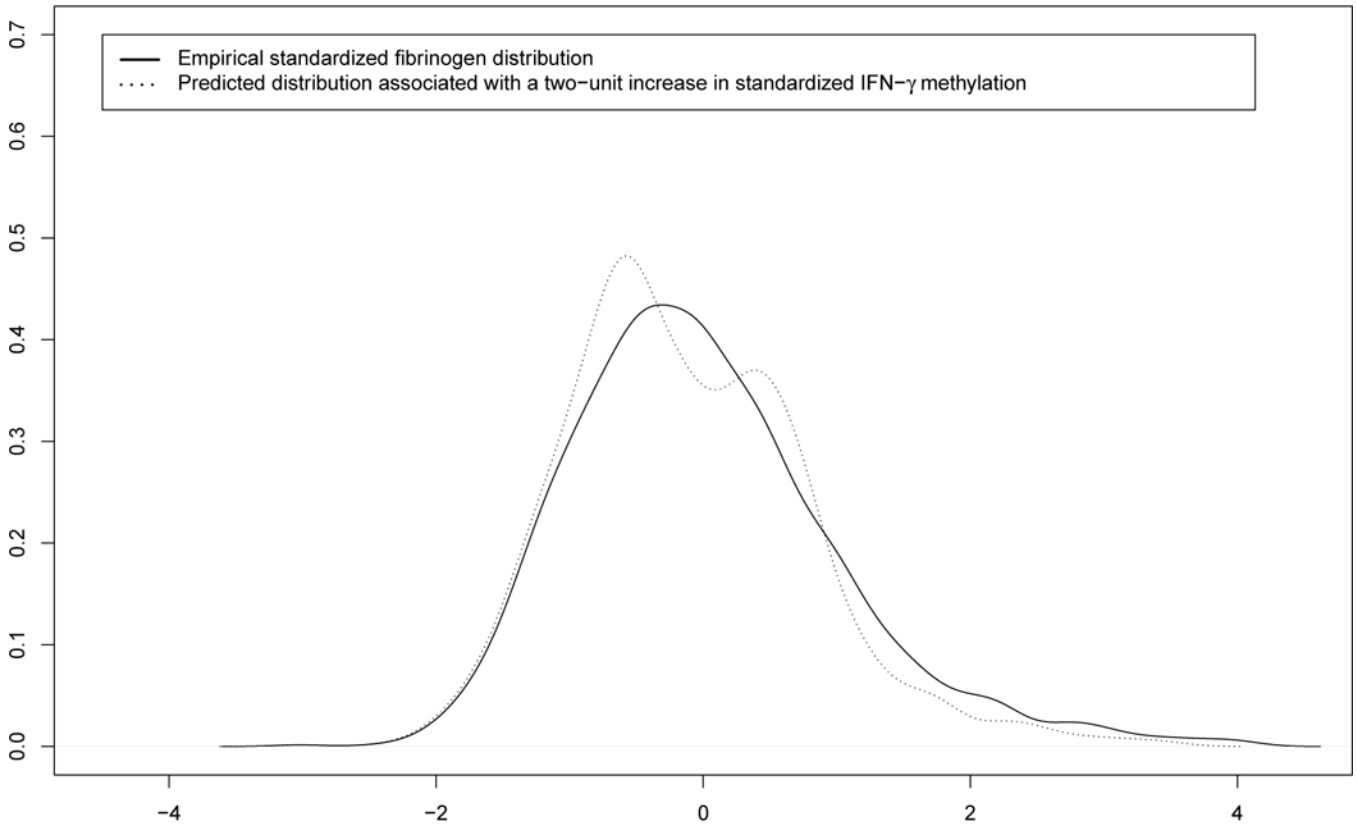

FIGURE 3.2.
Distributional change of the standardized fibrinogen distribution due to a change in standardized *IFN-γ* methylation.

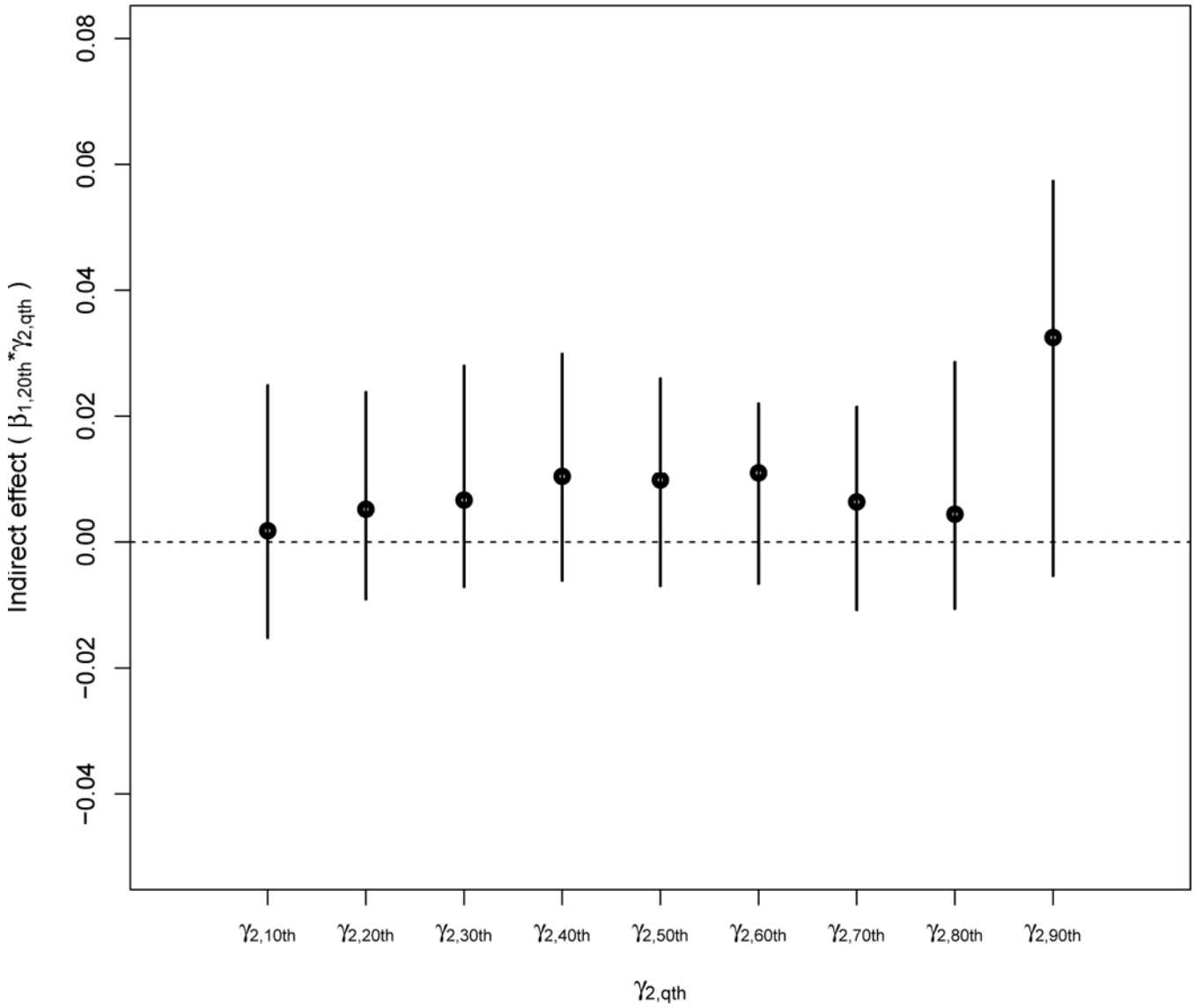

FIGURE 4.1.

Indirect effects of standardized particle number on standardized fibrinogen through a change in the 20th percentile of the standardized *IFN-γ* methylation distribution.

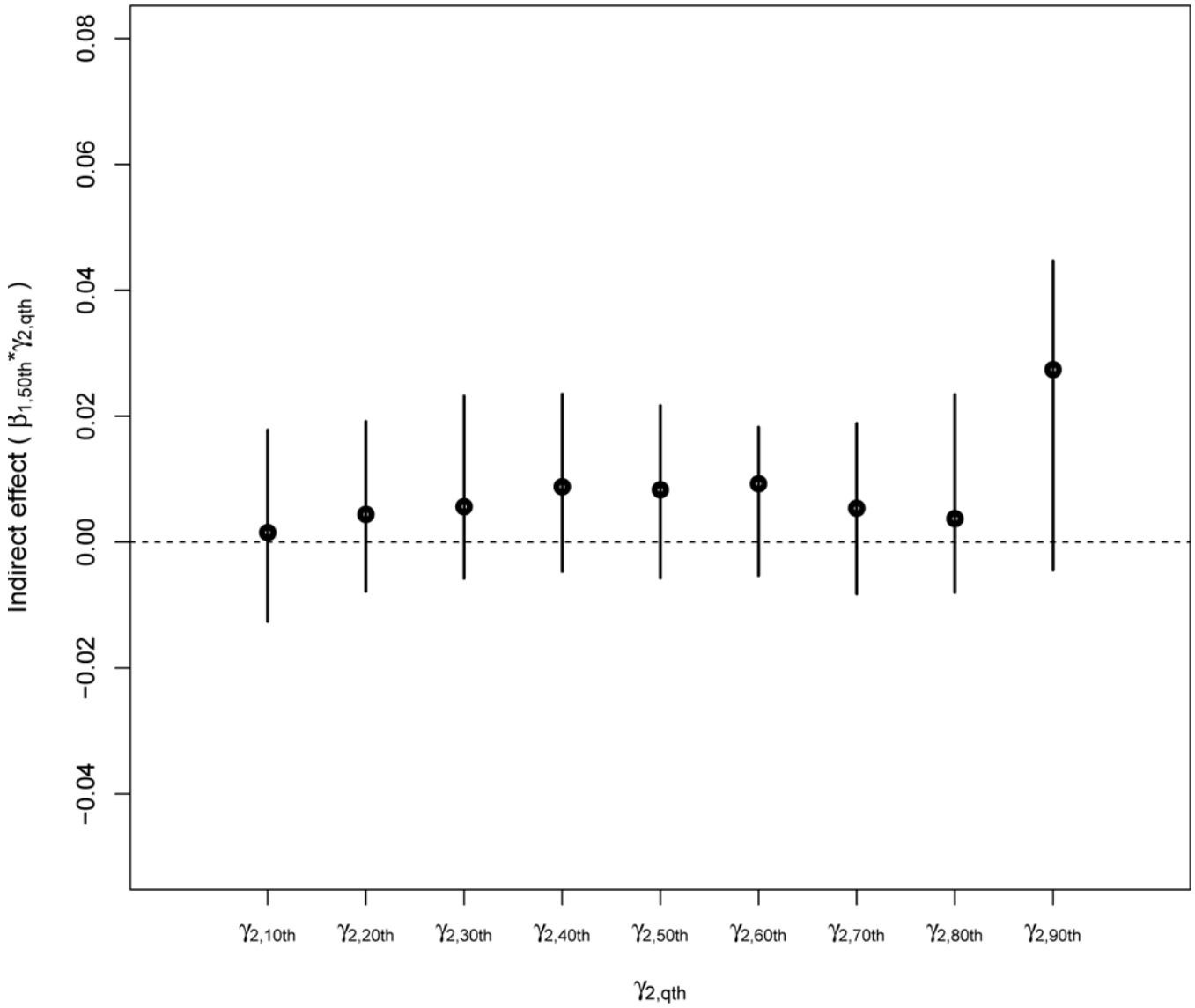

FIGURE 4.2.

Indirect effects of standardized particle number on standardized fibrinogen through a change in the 50*th* percentile of the standardized *IFN-γ* methylation distribution

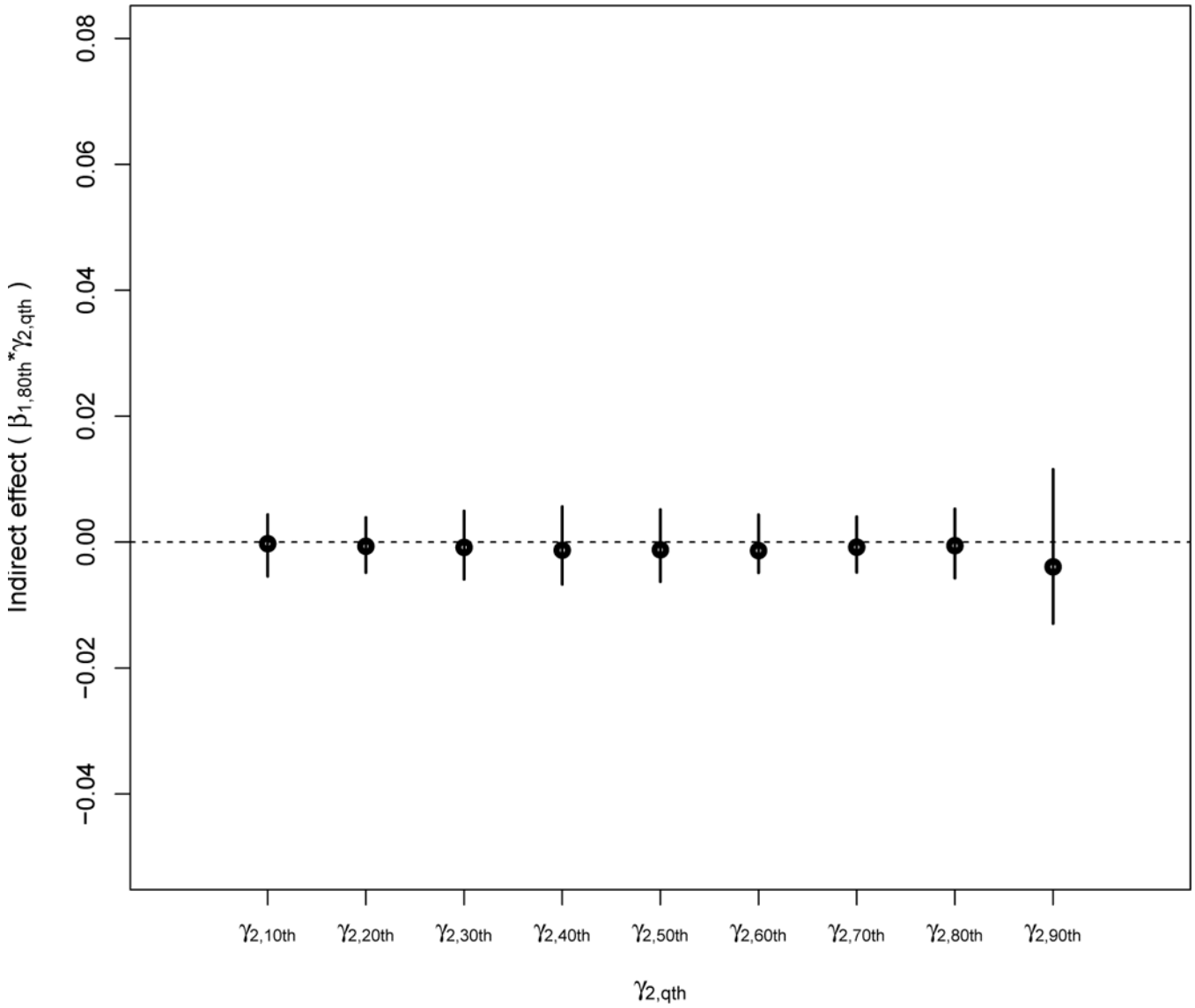

FIGURE 4.3.

Indirect effects of standardized particle number on standardized fibrinogen through a change in the 80*th* percentile of the standardized *IFN-γ* methylation distribution

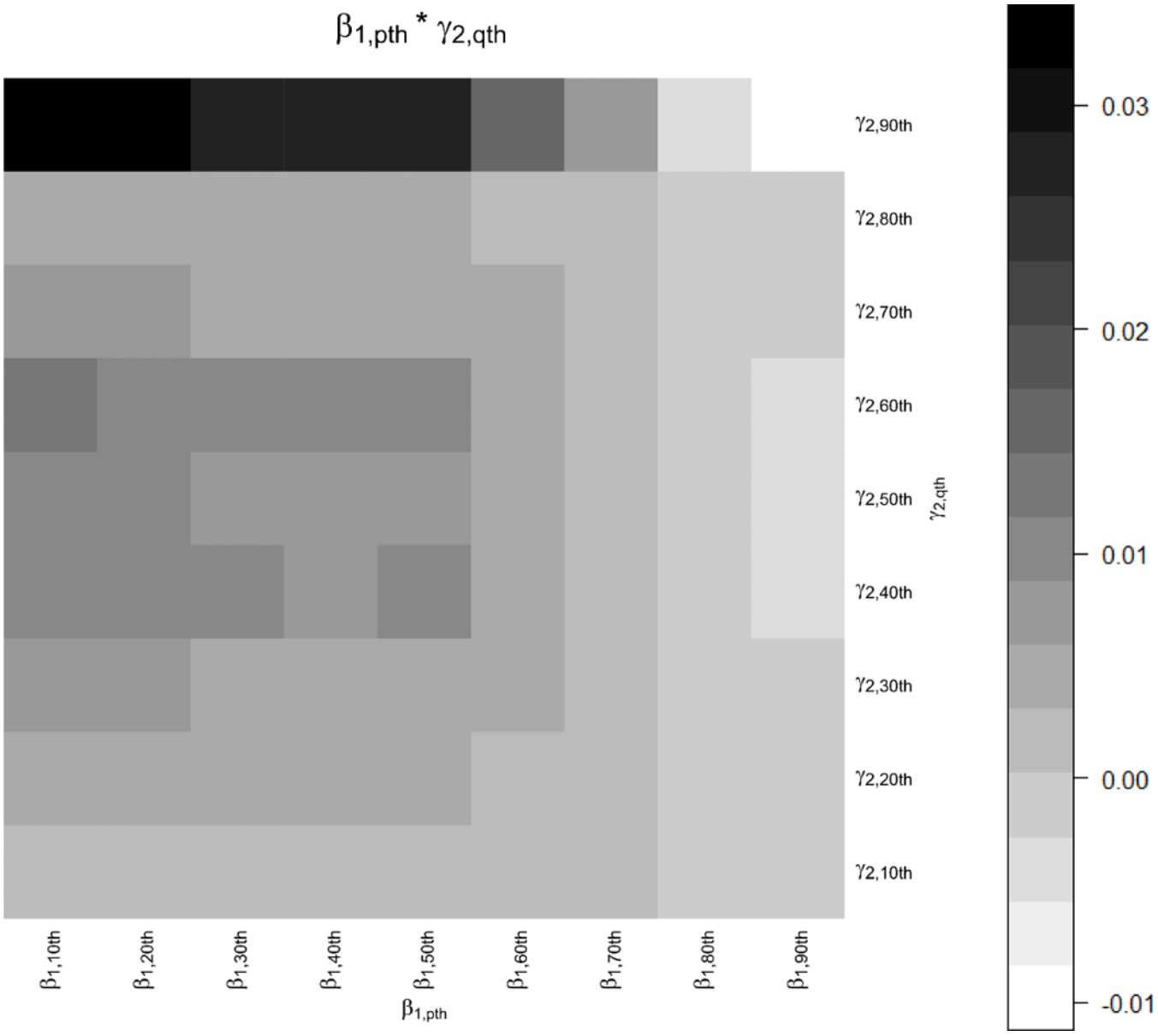

FIGURE 5.1.

Heatmap representing the indirect effects of standardized particle number on the *q*th percentiles of the standardized fibrinogen distribution through a change in the *pth* percentiles of the standardized *IFN-γ* methylation distribution

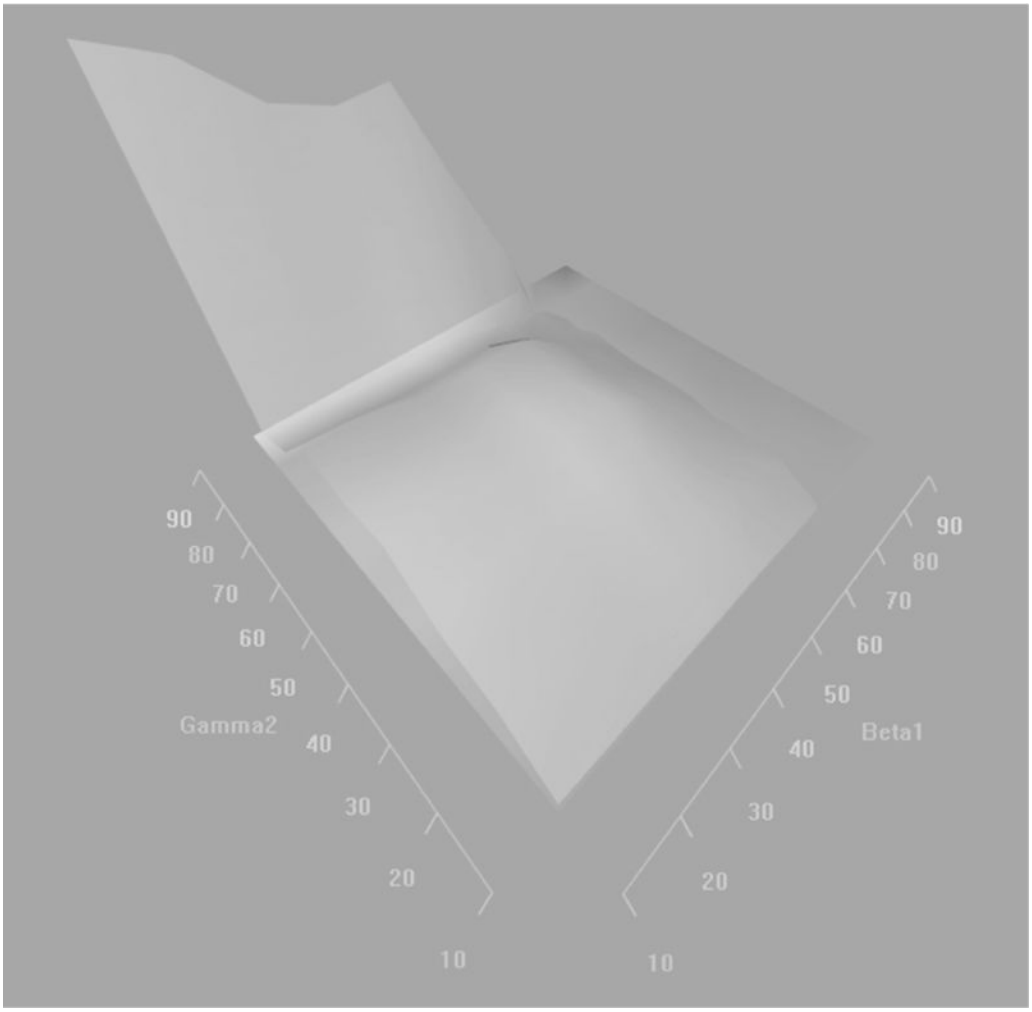

FIGURE 5.2.
3-D plot representing the indirect effects of standardized particle number on the *qth* percentiles of the standardized fibrinogen distribution through a change in the *pth* percentiles of the standardized IFN-*y* methylation distribution

Table 1

Quantile and mean regressions results

Coefficient $\gamma_{1,q}$	**Quantile estimate**	**95% $CI_{Bootstrap}$**
$\gamma_{1,10}$	0.0903	[−0.0583 to 0.1986]
$\gamma_{1,20}$	0.0625	[−0.0242 to 0.1830]
$\gamma_{1,30}$	0.0966	[−0.0230 to 0.2106]
$\gamma_{1,40}$	0.1055	[−0.0514 to 0.2277]
$\gamma_{1,50}$	0.0623	[−0.0268 to 0.2329]
$\gamma_{1,60}$	0.1342	[0.0256 to 0.2705]
$\gamma_{1,70}$	0.2047	[0.0597 to 0.3378]
$\gamma_{1,80}$	0.2083	[0.0693 to 0.3698]
$\gamma_{1,90}$	0.2920	[0.0844 to 0.4413]
Coefficient $\beta_{1,p}$	Quantile estimate	95% $CI_{Bootstrap}$
$\beta_{1,10}$	−0.1942	[−0.3346 to 0.0347]
$\beta_{1,20}$	−0.1833	[−0.2865 to −0.0621]
$\beta_{1,30}$	−0.1591	[−0.2714 to −0.0641]
$\beta_{1,40}$	−0.1507	[−0.2716 to 0.0495]
$\beta_{1,50}$	−0.1546	[−0.2348 to −0.0500]
$\beta_{1,60}$	−0.0967	[−0.1628 to −0.0066]
$\beta_{1,70}$	−0.0442	[−0.1121 to 0.0206]
$\beta_{1,80}$	−0.0220	[−0.0730 to 0.0824]
$\beta_{1,90}$	−0.0633	[−0.0460 to 0.1548]
Coefficient $\gamma_{2,q}$	Quantile estimate	95% $CI_{Bootstrap}$
$\gamma_{2,10}$	−0.0100	[−0.1191 to 0.0794]
$\gamma_{2,20}$	−0.0287	[−0.1124 to 0.0485]
$\gamma_{2,30}$	−0.0366	[−0.1347 to 0.0378]
$\gamma_{2,40}$	−0.0571	[−0.1385 to 0.0357]
$\gamma_{2,50}$	−0.0540	[−0.1246 to 0.0404]
$\gamma_{2,60}$	−0.0601	[−0.1029 to 0.0360]
$\gamma_{2,70}$	−0.0350	[−0.1131 to 0.0602]
$\gamma_{2,80}$	−0.0243	[−0.1481 to 0.0537]
$\gamma_{2,90}$	−0.1774	[−0.2595 to 0.0323]
Mean coefficient	Quantile estimate	95% $CI_{Asymptotic}$
$\beta_{1,mean}$	−0.2453	[−0.3572 to −0.1334]
$\gamma_{1,mean}$	0.1586	[0.0508 to 0.2665]
$\gamma_{2,mean}$	−0.0027	[−0.0629 to 0.0574]