# Causal Mediation Analysis for Difference-in-Difference Design and Panel Data

**Pei-Hsuan Hsia[1], An-Shun Tai[1], Chu-Lan Micheal Kao[1], Yu-Hsuan Lin[2], Sheng-Hsuan Lin[1]***

[1] Institute of Statistics, National Yang Ming Chiao Tung University, Hsin-Chu, Taiwan. 1001 University Road, Hsinchu, Taiwan 300

[2] Institute of Population Health Sciences, National Health Research Institutes, Miaoli, Taiwan 35 Keyan Road, Zhunan, Miaoli County 35053, Taiwan.

***Corresponding author**
Sheng-Hsuan Lin, MD, ScD
Institute of Statistics, National Chiao Tung University, Hsin-Chu, Taiwan
1001 University Road,
Hsinchu, Taiwan 300
Cell: +886 (3) 5712121 ext.56822
E-mail: shenglin@stat.nctu.edu.tw

## Abstract

Advantages of panel data, i.e., difference in difference (DID) design data, are a large sample size and easy availability. Therefore, panel data are widely used in epidemiology and in all social science fields. The literatures on causal inferences of panel data setting or DID design are growing, but no theory or mediation analysis method has been proposed for such settings. In this study, we propose a methodology for conducting causal mediation analysis in DID design and panel data setting. We provide formal counterfactual definitions for controlled direct effect and natural direct and indirect effect in panel data setting and DID design, including the identification and required assumptions. We also demonstrate that, under the assumptions of linearity and additivity, controlled direct effects can be estimated by contrasting marginal and conditional DID estimators whereas natural indirect effects can be estimated by calculating the product of the exposure-mediator DID estimator and the mediator-outcome DID estimator. A panel regression-based approach is also proposed. The proposed method is then used to investigate mechanisms of the effects of the Covid 19 pandemic on the mental health

status of the population. The results revealed that mobility restrictions mediated approximately 45 % of the causal effect of Covid 19 on mental health status.

# Introduction

Panel data is a form of dataset widely used in all social science fields, including sociology [1], economics [2, 3], and political science [4-7]. In epidemiology [8-11], panel data setting is usually treated as a quasi-experimental study to evaluate the role of a policy implementation in a specific health-related outcome. If only considering two time points (pretreatment and posttreatment time-points), panel data setting is also known as "difference-in-difference (DID) design" [10]. The panel data format (or DID design format) is the most common format for open access data, which usually includes nationwide or even worldwide information along with repeated observations between groups. For example, the scientific question of interest in this study was the causal effect of daily COVID-19 deaths on mental health status in a national population and the mediating effect (also termed as indirect effect) mediated through mobility restriction. The dataset is in the form of panel data containing information about daily COVID-19 deaths in 40 countries in the past 1 year. The literature on causal inferences of panel data settings or DID design have grown in the past decade. However, all methods proposed so far have only been used to investigate the average causal effect of a treatment among the treated (ATT) instead of the average causal effect in a national population, which is of interest and importance in epidemiology. In addition, methods for investigating causal mechanism under this setting are also required. Causal mediation analysis is a popular technique for investigating causal pathways through mediators of interest [12, 13]. Various studies have proposed methodologies for conducting mediation analysis under different settings [14-18]. However, no method of performing causal mediation analysis of panel data has been reported.

Therefore, this study developed a method of causal mediation analysis in which the causal effect of the overall population is decomposed into two parts: a mediating part and a non-mediating part. We provide formal definitions for causal effects of interest based on counterfactual models. We then provide the required assumptions for identification and estimation. We show that all causal effects can be expressed in terms of traditional DID estimators. Additionally, we illustrate that, when linearity is assumed, the two widely used methods, i.e., difference method and product method, are valid and correspond to different causal interpretations [19].

Our method is motivated by research on the effect of the nationwide COVID-19 outbreak on mental health and its mechanism. Previous studies have investigated how the COVID-19 outbreak affects mental health in the general population [20, 21]. We were interested in the extent to which this effect is mediated by mobility restriction in the general population. Pandemic-induced restrictions on mobility might result from fear of infection or from

government policy. Some researchers have hypothesized that restrictions on various activities and constant inconvenience in daily life likely result in feelings of loneliness and social isolation, which can then have negative mental health impacts, e.g., insomnia [22]. As a result, we hypothesized that restricted mobility is an important mediator of the causal effect of the national COVID-19 outbreak on mental health status. We obtained panel data for the extent of the national COVID-19 outbreak, the extent of mobility restriction, and the mental health status of the population. The Illustration section then demonstrates an application of causal mediation analysis.

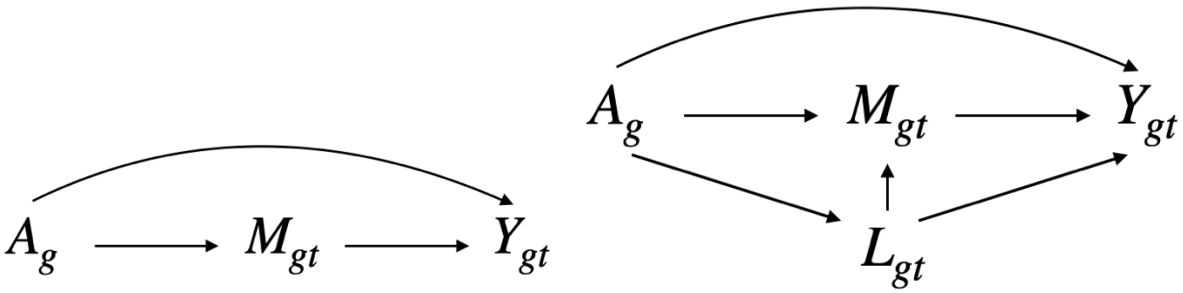
# Methods

*Notations and causal structures*

The variables from the panel data structure are labeled with time and group. For a unit $g$ (g =1, 2, …, G), let $A_g$ denote the exposure of interest. At time $t$, let $M_{gt}$ denote a potential mediator, and let $Y_{gt}$ denote the outcome. In time $t$, $M_{gt}$ is assumed to occur earlier than $Y_{gt}$. Moreover, $Y_{g0}$ and $M_{g0}$ denote baseline values (i.e., "pretreatment values" in previous works) of $Y_{gt}$ and $M_{gt}$, respectively. Here we only consider two time points: time t ("post-treatment period" in previous works) and the pre-treatment period. In our study, $A_g$ is the extent of the national COVID-19 outbreak, $M_{gt}$ is the extent of mobility restriction, and $Y_{gt}$ is the mental health status of the general population. In this example, the question is how to quantify the causal effect of the extent of the national COVID-19 outbreak $A_g$ on mental health status $Y_{gt}$ in the general population and the extent to which this effect is mediated by mobility restriction. For simplicity, we dichotomize $A_g$ and $M_{gt}$ as follows: $A_g = 0$ denotes that the number of cases of COVID-19 in country g is relatively low compared to the median number of deaths in country g over time; $A_g = 1$ denotes that the extent of the national COVID-19 outbreak is relatively high; $M_{gt} = 0$ denotes that mobility restriction is low compared to the median value for mobility restriction in country g over time; $M_{gt} = 1$ denotes that the mobility restrictions are relatively high.

We then use the counterfactual model to define all causal effects of interest [23, 24]. Let $Y_{gt}(a)$ be the hypothetical value of the outcome that would have occurred if $A_g$ had been set to $a$. Let $Y_{gt}(a, m)$ be the hypothetical value of the outcome that would have occurred if $A_g$ had been set to $a$ and $M_{gt}$ had been set to $m$. Similarly, let $M_{gt}(a)$ be the hypothetical value of the mediator if $A_g$ had been set to $a$. Consistency and composition are assumed[25-27] as follows: the value of $Y_{gt}(a)$ is the exact value of $Y_{gt}$ when $A_g = a$; the value of $Y_{gt}(a, m)$ is the observed value of $Y_{gt}$ when $A_g = a$ and $M_{gt} = m$; the value of $M_{gt}(a)$ is

the observed value of $M_{gt}$ when $A_g = a$. Figure 1(a) depicts the causal relations among variables $A_g$, $M_{gt}$, and $Y_{gt}$. A further assumption is that both exposure and mediator had no effect on the variables in pre-treatment period (NEPT) [28], which can be mathematically expressed as $Y_{g0} = Y_{g0}(a)$, $Y_{g0} = Y_{g0}(a, m)$ and $M_{g0} = M_{g0}(A_g = a)$ for each $a$ and $m$.




(a) (b)

**Fig. 1.** (a) Causal Diagram of exposure ($A_g$), mediator ($M_{gt}$), and outcome ($Y_{gt}$). (b) Causal Diagram in the presence of "recanting witness" (i.e., exposure-inducing-mediator-outcome confounder) ($L_{gt}$)

*General common trend assumption, unconfoundedness and exchangeability in DID design and panel data*

Before introducing the causal effects of interest, we discuss the common trend assumption, which is required for identification [9, 10, 29]. The common trend assumption is that, in a counterfactual world in which no one has been exposed, changes in outcome over time are same for all exposure levels. For example, the common trend assumption of the effect from $A_g$ to $Y_{gt}$ implies that $E[Y_{gt}(0) - Y_{g0}(0)|A_g = 0] = E[Y_{gt}(0) - Y_{g0}(0)|A_g = 1]$. Previous literature reveal that, under the common trend assumption, the causal effect of the treatment group (ATT) can be identified [28]. In this study, we are interested in all populations. Therefore, we must apply the common trend assumption in another counterfactual world in which everyone has been exposed, i.e., $E[Y_{gt}(1) - Y_{g0}(1)|A_{gt} = 0] = E[Y_{gt}(1) - Y_{g0}(1)|A_g = 1]$.

A more general assumption that satisfies the above two common trend assumptions is $Y_{gt}(a) - Y_{g0}(a) \perp A_g \; \forall a$ (CT-0 assumption). That is, the counterfactual trend of $Y_{gt}$ is exchangeable among groups with different values for $A_g$. This general common trend assumption mathematically equals to $Y_{gt}(a) \perp A_g|Y_{g0} \; \forall a$ (see Appendix A for detailed proof). This equation also reveals how the common trend assumption is related to the unconfoundedness assumption: the general common trend assumption can be interpreted as the pretreatment outcome (i.e., $Y_{g0}$) that captures all potential confounders between $A_g$ and $Y_{gt}$.

Based on the general common trend assumption and its interpretation, we introduce the

assumptions for mediation analysis. We first determine the general common trend assumption of the effect from $A_g$ and $M_{gt}$ to $Y_{gt}$: (CT-1 assumption) $Y_{gt}(a,m) - Y_{g0}(a,m) \perp A_g \; \forall a$ and (CT-2 assumption) $Y_{gt}(a,m) - Y_{g0}(a,m) \perp M_{gt}|A_g \; \forall a,m$. We also determine the general common trend assumption of the effect from $A_g$ to $M_{gt}$: (CT-3 assumption) $M_{gt}(a) - M_{g0}(a) \perp A_g \; \forall a$. Similar interpretation of CT-0 can be implied for the above three assumptions (i.e. CT-1, CT-2, and CT-3). The CT-1 assumption and the CT-0 assumption have identical interpretations; the CT-2 assumption can be interpreted as the assumption that both pretreatment outcome and exposure capture all potential confounders between $M_{gt}$ and $Y_{gt}$; the CT-3 assumption can be interpreted as the assumption that the pretreatment mediator captures all potential confounders between $M_{gt}$ and $A_g$.

We also require one additional"cross-world" common trend assumption: (CT-4 assumption) $Y_{gt}(a,m) - Y_{g0}(a,m) \perp M_{gt}(a^*)|A_g \; \forall a, a^*, m$. The sufficient conditions for this assumption are CT-2 assumption and no mediator-outcome confounder affected by the exposure. This assumption is required for identifying the NDE and NIE even in the non-DID design setting. The exposure-inducing-mediator-outcome confounder is also termed the "recanting witness" in some literature. Figure 1(b) is an example of a CT-4 violation in the presence of a recanting witness ($L_{gt}$) [30].

The following sections discuss the conventional definitions of total effect, direct effect, and indirect effect in the DID setting and panel data [31]. We also show that, with binary $A_g$ and $M_{gt}$, all effects can be expressed in terms of traditional DID estimates. It merits noting that the assumptions of consistency, composition, and NEPT are always presumed to hold true. However, the rest of this article only discusses when CT assumptions should be applied.

*Definition, identification and estimation of total effect*

The total effect (TE) from $A_g$ to $Y_{gt}$ is defined as $E\left[Y_{gt}(1) - Y_{gt}(0)\right]$, where the average outcome obtained if $A_g$ had been set to 1 is compared with the average outcome obtained if $A_g$ had been 0. In this study, TE is the effect of the extent of the national COVID-19 outbreak on the mental health status of the general popualtion. Under CT-0 assumption, TE can be identified as $\boldsymbol{DID}_{Y_{gt},A_g}$, where $\boldsymbol{DID}_{Y_{gt},A_g} := \left\{E\left[Y_{gt}\middle|A_g = 1\right] - E\left[Y_{g0}\middle|A_g = 1\right]\right\} - \left\{E\left[Y_{gt}\middle|A_g = 0\right] - E\left[Y_{g0}\middle|A_g = 0\right]\right\}$.

Details of the identification process appear in Appendix B. The $\boldsymbol{DID}_{Y_{gt},A_g}$ is a DID estimator [11, 32-34]. Notably, previous literature only apply the conventional common trend

assumption and interpret $\boldsymbol{DID}_{Y_{gt},A_g}$ as the ATT. In contrast, our study applies a general common trend assumption (which is stronger than the traditional one): the identical $\boldsymbol{DID}_{Y_{gt},A_g}$ can be interpreted as the causal effect for the total population. The $\boldsymbol{DID}_{Y_{gt},A_g}$ can be used to estimate TE as the difference in the conditional means of $Y_{gt}$ and $Y_{g0}$ where $A_g = 1$ minus that of where $A_g = 0$.

For random variables W and X (where X is binary), a marginal DID estimator can be denoted in simplified notation as $\boldsymbol{DID}_{W,X} := \{E[W_{gt}|X = 1] - E[W_{g0}|X = 1]\} - \{E[W_{gt}|X = 0] - E[W_{g0}|X = 0]\}$. Additionally, for random variables W and X and an additional random variable Z, a conditional DID estimator can be denoted as $\boldsymbol{DID}_{W,X|Z} := \{E[W_{gt}|X = 1, Z] - E[W_{g0}|X = 1, Z]\} - \{E[W_{gt}|X = 0, Z] - E[W_{g0}|X = 0, Z]\}$.

*Definition, Identification and Estimation based on DID estimate of Controlled Direct Effect (CDE)*

According to different research questions, two decomposition strategies for mediation analysis are available [35]. To evaluate the extent of the effect of an exposure, which is eliminated by the intervening the mediator, TE is decomposed into CDE and portion eliminated (PE) [31, 36], i.e., $\text{TE} = \text{CDE} + \text{PE}$. the extent of effect of the exposure, which can be explained by the mediation mechanisms via the mediator [36, 37], TE is decomposed into natural direct effect (NDE) and natural indirect effect (NIE), i.e. $\text{TE} = \text{NDE} + \text{NIE}$. The causal mediation analysis method proposed in this method is applicable for DID design and panel data setting. This section discusses CDE and PE, and the next section discusses NDE and NIE.

The CDE, which is defined as $CDE(m) = E[Y_{gt}(1, m)] - E[Y_{gt}(0, m)]$, compares exposure between level $1$ and level $0$ and sets the mediator level to $m$. The CDE captures the effect of exposure $A_g$ on outcome $Y_{gt}$ by setting $M_{gt}$ to $m$. Note that CDE may vary with $m$ if effects of $A_g$ interact with effects of $M_{gt}$. In our motivating example, CDE (1) is interpreted as the change in mental health status caused by the increased severity of the national COVID-19 outbreak when the mobility is always highly restricted by government policy (such as lockdown policy).

Under assumptions CT-1 and CT-2, CDE can be identified as a conditional DID estimator of $Y_{gt}$ and $A_g$ conditioning on $M_{gt}$ as $m$, $\boldsymbol{DID}_{Y_{gt},A_g|M_{gt}=m}$, where

$$\boldsymbol{DID}_{Y_{gt},A_g|M_{gt}=m} := \left[E[Y_{gt}|A_g = 1, M_{gt} = m] - E[Y_{g0}|A_g = 1, M_{gt} = m]\right] -$$

$$\left[E[Y_{gt}|A_g = 0, M_{gt} = m] - E[Y_{g0}|A_g = 0, M_{gt} = m]\right].$$

Appendix C provides the details of the identification process. The CDE is a function of m, which is set according to the research question. The $\boldsymbol{DID}_{Y_{gt},A_g|M_{gt}=m}$ can be used to estimate CDE by calculating the difference in the conditional means of $Y_{gt}$ and $Y_{g0}$ where $A_g = 1$ and $M_{gt} = m$ minus the difference in the conditional means of $Y_{gt}$ and $Y_{g0}$ where $A_g = 0$ and $M_{gt} = m$.

The PE can then be calculated as the difference between TE and CDE as

$$\mathrm{PE(m)} = \mathrm{TE} - \mathrm{CDE(m)} = \boldsymbol{DID}_{Y_{gt},A_g} - \boldsymbol{DID}_{Y_{gt},A_g|M_{gt}=m},$$

which is the contrast of marginal DID mediator of $Y_{gt}$ and $A_g$ and conditional DID mediator conditional on $M_{gt}$ as $m$. After TE and CDE are estimated, PE can be estimated.

*Definition, Identification and Estimation based on DID estimate of NDE and NIE*

Based on the traditional definition of NDE [31, 38], this study defined NDE for DID design and panel data settings as

$$E\left[Y_{gt}\left(1, M_{gt}(0)\right) - Y_{gt}\left(0, M_{gt}(0)\right)\right],$$

the contrast of the expectation of hypothetical values of $Y_{gt}$ had the exposure been set to 1 (versus 0) and had the mediator been always set to the hypothetical value had the exposure been set to 0. In our motivating example, this would capture the change in mental health status between two levels of severity of the national COVID-19 outbreak if the maximum restriction on mobility for each country had always been set to the mobility level in the absence of a national COVID-19 outbreak. The NIE can be defined as $E\left[Y_{gt}\left(1, M_{gt}(1)\right) - Y_{gt}\left(1, M_{gt}(0)\right)\right]$, which is the difference in the hypothetical values of $Y_{gt}$ had the exposure always been set to 1, but had the mediator been set to two hypothetical values had the exposure been set to 1 versus 0. In our example, NIE captures mental health status by comparing the mobility restriction policy under two levels of severity of the national COVID-19 outbreak if the severity of the national COVID-19 outbreak had in fact been 1.

Four common trend assumptions are needed to identify NDE and NIE. Under assumptions CT-1 to 4, NIE can be identified as $NIE = \boldsymbol{DID}_{M_{gt}|A_g} \times \boldsymbol{DID}_{Y_{gt},M_{gt}|A_g=1}$, where

$$\boldsymbol{DID}_{M_{gt},A_g} := \left\{E\left[M_{gt}\middle|A_g = 1\right] - E\left[M_{g0}\middle|A_g = 1\right]\right\} - \left\{E\left[M_{gt}\middle|A_g = 0\right] - E\left[M_{g0}\middle|A_g = 0\right]\right\},$$

and

$$\boldsymbol{DID}_{Y_{gt},M_{gt}|A_g=1} := \left\{\left[E\left[Y_{gt}\middle|A_g = 1, M_{gt} = 1\right] - E\left[Y_{g0}\middle|A_g = 1, M_{gt} = 1\right]\right] -$$

$\left[E\left[Y_{gt}\middle|A_g=1,M_{gt}=0\right]-E\left[Y_{g0}\middle|A_g=1,M_{gt}=0\right]\right]\}$.

The details of the procedures appear in Appendix D. The first term ($\boldsymbol{DID}_{M_{gt},A_g}$) is the marginal DID estimator of $M_{gt}$ and $A_g$, which can be interpreted as the effect of $A_g$ on $M_{gt}$. The second term ($\boldsymbol{DID}_{Y_{gt},M_{gt}|A_g=1}$) is the conditional DID estimator of $Y_{gt}$ and $M_{gt}$ when $A_g = 1$, which can be interpreted as the effect of $M_{gt}$ on $Y_{gt}$. Therefore, the NIE can be expressed as the product of the effect of $A_g$ on $M_{gt}$ and the effect of $M_{gt}$ on $Y_{gt}$, expressed in the form of DID estimators. The $\boldsymbol{DID}_{M_{gt},A_g}$ can be used to estimate the effect of $A_g$ on $M_{gt}$ based on the difference between the conditional means of $M_{gt}$ and $M_{g0}$ where $A_g = 1$ and the conditional means of $M_{gt}$ and $M_{g0}$, where $A_g = 0$. Meanwhile, $\boldsymbol{DID}_{Y_{gt},M_{gt}|A_g=1}$ can be used to estimate the effect of $M_{gt}$ on $Y_{gt}$ by calculating the difference in the conditional means of $Y_{gt}$ and $Y_{g0}$ where $A_g = 1$ and $M_{gt} = 1$ minus the conditional means of $Y_{gt}$ and $Y_{g0}$ where $A_g = 1$ and $M_{gt} = 0$. The product of estimators $\boldsymbol{DID}_{M_{gt},A_g}$ and $\boldsymbol{DID}_{Y_{gt},M_{gt}|A_g=1}$ obtains the estimation of NIE. Thus, NDE is the difference between TE and NIE, i.e., $NDE = \text{TE} - \text{NIE}$. Estimation of TE and NIE enables estimation of NDE.

*Empirical 2FE regression-based method*

In the above discussion, the mediator was binary, and effects were only estimated at a certain time point. This section proposes an empirical regression-based method that can be used for a continuous mediator. The aim is to determine the mean effect over time. The $A_g$ as $A_{gt}$ are used to estimate the effects at each time $t$ under the assumption of time sequence $A_{gt} \rightarrow M_{gt} \rightarrow Y_{gt}$. For analysis of continuous variables, this study used two-way fixed effects (2FE) regression models, to enable simultaneous adjustment for unobserved unit-specific and time-specific confounders [39]. Here we assume that there is no long-term effects, e.g., insomnia status in a given month was assumed to be affected only by the national COVID-19 outbreak and by mobility restriction in that month; similarly, insomnia status in a given month was assumed to be affected only by the severity of the national COVID-19 outbreak and mobility restriction in that month. Under assumptions of linearity, additivity (no interaction), and absence of long-term effects, the estimation of panel regression is a generalized form of DID estimator with multiple treatment level and time-points. According to the literature, the two methods perform equally when both the exposure and mediator are binary with two time-points [39-41].

Three models of $Y_{gt}$ and $M_{gt}$ were built for estimation. First, the role of exposure $A_{gt}$ on outcome $Y_{gt}$ was modeled. Assuming that the treatment effect is additive and that the error

terms are additively separable, the resulting 2FE regression model (Model 1) can be written as

$E[Y_{gt}|A_{gt} = a_{gt}] = \gamma_A a_{gt} + \gamma_t + \gamma_g$ (Model 1)

In Model 1, $\gamma_g$ captures the impact of any unobserved but temporally stable characteristic of group $g$ on outcome $Y_{gt}$; $\gamma_t$ captures the impact of time t, which is stable over groups, on outcome $Y_{gt}$. Additionally, since time has no effect at baseline, $Y_{g0}$ can be viewed as $\gamma_g$. Using $Y_{g0}$ to represent the effect of temporally stable characteristics of group $g$ on outcomes $Y_{gt}$ is intuitive. Coefficient, $\gamma_A$, which represents the causal effect of $A_{gt}$ on $Y_{gt}$ (i.e., TE), is assumed to be invariant over time.

A similar approach can be used to build a model of the effect of $Y_{gt}$ on both $A_{gt}$ and $M_{gt}$ and to build a model of the effect of $M_{gt}$ on $A_{gt}$ as follows:

$E[Y_{gt}|A_{gt} = a_{gt}, M_{gt} = m_{gt}] = \theta_A a_{gt} + \theta_M m_{gt} + \theta_t + \theta_g$ (Model 2)

$E\left[M_{gt}\middle|A_{gt} = a_{gt}\right] = \beta_A a_{gt} + \beta_t + \beta_g$ (Model 3)

Model 1 can be used to quantify TE based on $\gamma_A$. According to models 1 and 2, the estimated CDE is $\theta_A$, and the estimated PE is $\gamma_A - \theta_A$, which coincides with the difference method. According to models 2 and 3, the estimated NDE is $\theta_A$, and the estimated NIE is $\theta_M \beta_A$. Moreover, when models 1-3 are all correct, $\gamma_A - \theta_A = \theta_M \beta_A$ (see Appendix E for detailed proof), which are the same results obtained by difference method and product method obtains the same estimation results in linear models [42]. The above estimations are applicable when $A_{gt}$ can be binary or continuous. Notably, the results obtained when $M_{gt}$ is binary are identical to those obtained by the DID estimator version mentioned above.

# Illustration

*Data Description*

The method proposed in this study was used to investigate the causal effect of the national COVID-19 outbreak on mental health status and the mediating role of mobility restriction in this causal relationship. The dataset used in this study was for 40 countries throughout the world for the period from March, 2020, to February, 2021. Exposure $A_{gt}$ was the number of deaths per day (monthly average) due to COVID-19. Exposure $A_{gt}$ was used as an indicator of the severity of the national COVID-19 outbreak within a country during this period. Mediator $M_{gt}$ was cellular phone usage in residential areas, which was obtained from Google. Google obtains aggregated and anonymized cellular phone location data from users who have a Google account on their cellular phone and who opted to provide their location data to Google Location History. Cellular phone usage was recorded for six categories of places: workplace, retailers, transit stations, grocery stores, parks, and residences. In this study, residential cellular

phone usage was used as a surrogate of mobility restriction for each country during the period from March, 2020, to February, 2021.

Outcome variable $Y_{gt}$ was the volume of Google searches for "insomnia" during the period March, 2020, to February, 2021, according to Google Trends data. Google search volume was used as a surrogate for mental health status in 40 countries. Accurate translation of 'insomnia' into the local languages of the 40 countries was confirmed by using Google Translate for forward translations from Chinese and English and back translation to Chinese and English. For the 40 countries, average monthly values for each variable were collected over a 12-month period (Appendix F). Table S1 in Appendix G provides the descriptive statistics for these variables. The pretreatment time was the month of January, 2020, in which the impacts of the national COVID-19 outbreak had not yet occurred.

Data analysis was demonstrated in two cases (DID design and empirical 2FE regression-based estimation) as mentioned in the estimation section. In the first case, DID estimators were used to quantify the effects of the June, 2020, national COVID-19 outbreak on mobility restriction and insomnia status in July, 2021. The threshold for dichotomizing variable $A_g$ was the monthly median number of deaths in each country. If $A_g = 1$, monthly deaths have increased and the severity of the national COVID-19 outbreak is classified as high. In contrast, if $A_g = 0$, the severity of the pandemic is relatively low. The $M_{gt}$ is also dichotomized by the median value, where $M_{gt} = 1$ suggests that the time people spend in their homes is longer than usual. Standard errors were derived by bootstrap method. In the second case of the data analysis, we estimate the effects of the COVID-19 outbreak on mobility restriction and insomnia status in the period from March, 2020, to February, 2021, and all variables were considered continuous. Appendix H presents the models of the volumes of 'insomnia' searches and the model of the extent of mobility restriction. Since fixed time effects were controlled and eliminated during the estimation process, all 12 months of data were used in the analysis. In this estimation, standard deviation was used to standardize the number of deaths. Delta method was used to estimate standard error. All analyses were performed with R software, version 3.6.3 (R Foundation for Statistical Computing).

*DID estimation*

Table 1 presents the TE, CDE, PE, NDE and NIE. The TE of the severity of the national COVID-19 outbreak had a significant positive effect on the tendency of insomnia (10.16, 95% CI 2.81 to 17.51). This indicated that the tendency of insomnia increased during the COVID-19 pandemic outbreak, which is consistent with the literature. Additionally, the effect of A on M ($\boldsymbol{DID}_{M_{gt},A_g}$) was 0.59 (95% CI 0.35 to 0.83) while the effect of M on Y ($\boldsymbol{DID}_{Y_{gt},M_{gt}|A_g=1}$)

was 7.73 (95% CI 2.46 to 13.00), which respectively indicated that (1) mobility restrictions increased as the severity of the national COVID-19 outbreak increased and that (2) the tendency of insomnia increased as mobility restrictions increased. Mediation analysis revealed a significant NIE (4.62, 95% CI 1.27 to 7.97, with Proportion mediated as 0.45), i.e., that mobility restriction mediated the causal effects of COVID-19-induced insomnia and explained approximately 45 % of the causal mechanism. Since 55% of the effect was not explained by the mediation pathway, other important mechanisms may have had roles and need further study.

*Empirical Regression-based Estimation*

This study used 2FE models to estimate the effect over the year. Table S2 (see Appendix I) shows the maximum likelihood estimates for the coefficients in Models 4-6. Table 2 presents the estimated effects (TE, CDE, PE, NDE and NIE). All effects were positive. However, only NIE was statistically significant (0.41, 95% CI = 0.18 (0.06, 0.76)) with the proportion mediated equal to 0.29, which provided evidence that the mediating effect of mobility restriction on the association between the severity of the national COVID-19 outbreak and the tendency of insomnia continued throughout the entire 1-year period of the analysis.

# Discussion

To our best knowledge, this study is the first to extend causal mediation analysis to DID design and panel data settings. This study provided formal counterfactual definitions for direct and indirect effects in panel data settings and DID design, including both the identification and required assumptions. Researchers can use general statistical software to implement the proposed DID estimators and regression-based estimators.

Several limitations of the proposed methods are worth noting. First, the empirical regression-based estimator and model require strong causal assumptions. For example, the model assumes the absence of interactions between variables, including interactions between time and group, between time and variables, and between group and variables. However, interactions may exist in many real-world scenarios. Additionally, non-linear outcome models are widely used in epidemiology research. Additional studies are needed to develop a general methodology that is applicable under widely varying conditions. Moreover, our analysis assumed time sequences between variables, which must be confirmed by substantial background knowledge when conducting data analysis. Sensitivity analysis techniques must be developed to assess bias when assumptions are violated. Finally, all analyses in this study are

based on ecological data and the causal inference is only applicable in ecological level. Inferring causality to individual level will be subject to ecological fallacy.

# Reference


1. White, L. and J.N. Edwards, *Emptying the nest and parental well-being: An analysis of national panel data.* American sociological review, 1990: p. 235-242.
2. Blundell, R., et al., *Investment and Tobin's Q: Evidence from company panel data.* Journal of econometrics, 1992. **51**(1-2): p. 233-257.
3. Arellano, M., *Panel data econometrics*. 2003: Oxford university press.
4. Plümper, T., V.E. Troeger, and P. Manow, *Panel data analysis in comparative politics: Linking method to theory.* European Journal of Political Research, 2005. **44**(2): p. 327-354.
5. Eveland Jr, W.P., et al., *Understanding the relationship between communication and political knowledge: A model comparison approach using panel data.* Political Communication, 2005. **22**(4): p. 423-446.
6. Finkel, S.E. and E.N. Muller, *Rational choice and the dynamics of collective political action: Evaluating alternative models with panel data.* American Political Science Review, 1998. **92**(1): p. 37-49.
7. Boulianne, S., *Stimulating or reinforcing political interest: Using panel data to examine reciprocal effects between news media and political interest.* Political Communication, 2011. **28**(2): p. 147-162.
8. Macinko, J., F.C. Guanais, and M.D.F.M. De Souza, *Evaluation of the impact of the Family Health Program on infant mortality in Brazil, 1990–2002.* Journal of Epidemiology & Community Health, 2006. **60**(1): p. 13-19.
9. Dimick, J.B. and A.M. Ryan, *Methods for evaluating changes in health care policy: the difference-in-differences approach.* Jama, 2014. **312**(22): p. 2401-2402.
10. Wing, C., K. Simon, and R.A. Bello-Gomez, *Designing difference in difference studies: best practices for public health policy research.* Annual review of public health, 2018. **39**.
11. Chai, Y., A. Nandi, and J. Heymann, *Does extending the duration of legislated paid maternity leave improve breastfeeding practices? Evidence from 38 low-income and middle-income countries.* BMJ global health, 2018. **3**(5): p. e001032.
12. MacKinnon, D.P., A.J. Fairchild, and M.S. Fritz, *Mediation analysis.* Annu. Rev. Psychol., 2007. **58**: p. 593-614.
13. VanderWeele, T., *Explanation in causal inference: methods for mediation and interaction*. 2015: Oxford University Press.
14. Baron, R.M. and D.A. Kenny, *The moderator–mediator variable distinction in social*

*psychological research: Conceptual, strategic, and statistical considerations.* Journal of personality and social psychology, 1986. **51**(6): p. 1173.

15. Judd, C.M. and D.A. Kenny, *Process analysis: Estimating mediation in treatment evaluations.* Evaluation review, 1981. **5**(5): p. 602-619.
16. Valeri, L. and T.J. VanderWeele, *Mediation analysis allowing for exposure–mediator interactions and causal interpretation: theoretical assumptions and implementation with SAS and SPSS macros.* Psychological methods, 2013. **18**(2): p. 137.
17. VanderWeele, T.J. and S. Vansteelandt, *Odds ratios for mediation analysis for a dichotomous outcome.* American journal of epidemiology, 2010. **172**(12): p. 1339-1348.
18. VanderWeele, T.J., *Causal mediation analysis with survival data.* Epidemiology (Cambridge, Mass.), 2011. **22**(4): p. 582.
19. Pearl, J., *The causal mediation formula—a guide to the assessment of pathways and mechanisms.* Prevention science, 2012. **13**(4): p. 426-436.
20. Lin, Y.-H., T.-W. Chiang, and Y.-L. Lin, *Increased Internet Searches for Insomnia as an Indicator of Global Mental Health During the COVID-19 Pandemic: Multinational Longitudinal Study.* J Med Internet Res, 2020. **22**(9): p. e22181.
21. Wu, T., et al., *Prevalence of mental health problems during the COVID-19 pandemic: A systematic review and meta-analysis.* Journal of affective disorders, 2020.
22. Hawkley, L.C. and J.T. Cacioppo, *Loneliness matters: A theoretical and empirical review of consequences and mechanisms.* Annals of behavioral medicine, 2010. **40**(2): p. 218-227.
23. Rubin, D., *Formal modes of statistical inference for causal effects.* Quality Engineering, 1990. **36**: p. 185-188.
24. Hernán, M.A., *A definition of causal effect for epidemiological research.* Journal of Epidemiology & Community Health, 2004. **58**(4): p. 265-271.
25. Pearl, J., *Models, reasoning and inference.* Cambridge, UK: CambridgeUniversityPress, 2000. **19**.
26. VanderWeele, T.J., *Concerning the consistency assumption in causal inference.* Epidemiology, 2009. **20**(6): p. 880-883.
27. VanderWeele, T.J. and S. Vansteelandt, *Conceptual issues concerning mediation, interventions and composition.* Statistics and its Interface, 2009. **2**(4): p. 457-468.
28. Lechner, M., *The estimation of causal effects by difference-in-difference methods*. 2011: Now Hanover, MA.
29. Ryan, A.M., J.F. Burgess Jr, and J.B. Dimick, *Why we should not be indifferent to*

*specification choices for difference-in-differences.* Health services research, 2015. **50**(4): p. 1211-1235.

30. Vansteelandt, S. and T.J. VanderWeele, *Natural direct and indirect effects on the exposed: effect decomposition under weaker assumptions.* Biometrics, 2012. **68**(4): p. 1019-1027.
31. Pearl, J., *Direct and indirect effects*, in *Proceedings of the Seventeenth conference on Uncertainty in artificial intelligence*. 2001, Morgan Kaufmann Publishers Inc.: Seattle, Washington. p. 411–420.
32. Gangl, M., *Causal Inference in Sociological Research.* Annual Review of Sociology, 2010. **36**(1): p. 21-47.
33. Han, B., H. Yu, and M.W. Friedberg, *Evaluating the impact of parent-reported medical home status on children's health care utilization, expenditures, and quality: a difference-in-differences analysis with causal inference methods.* Health services research, 2017. **52**(2): p. 786-806.
34. Abadie, A., *Semiparametric difference-in-differences estimators.* The Review of Economic Studies, 2005. **72**(1): p. 1-19.
35. VanderWeele, T.J., *Explanation in Causal Inference: Methods for Mediation and Interaction*. 2015: Oxford University Press.
36. Robins, J.M., *Semantics of causal DAG models and the identification of direct and indirect effects.*
37. Joffe, M.M., D. Small, and C.-Y. Hsu, *Defining and estimating intervention effects for groups that will develop an auxiliary outcome.* Statistical Science, 2007. **22**(1): p. 74-97.
38. Robins, J.M. and S. Greenland, *Identifiability and exchangeability for direct and indirect effects.* Epidemiology, 1992: p. 143-155.
39. Imai, K. and I.S. Kim, *On the use of two-way fixed effects regression models for causal inference with panel data.* Political Analysis, 2021. **29**(3): p. 405-415.
40. Bertrand, M., E. Duflo, and S. Mullainathan, *How much should we trust differences-in-differences estimates?* The Quarterly journal of economics, 2004. **119**(1): p. 249-275.
41. Angrist, J.D. and J.-S. Pischke, *Mostly harmless econometrics*. 2008: Princeton university press.
42. Jiang, Z. and T.J. VanderWeele, *When Is the Difference Method Conservative for Assessing Mediation?* American Journal of Epidemiology, 2015. **182**(2): p. 105-108.

**Table 1**. Estimations for all effects of the covid-19 severities on search volumes of "insomnia".

| | **Estimation** | **SD** | **95%CI** | **P value** |
|---|---|---|---|---|
| Total effect | 10.16 | 3.75 | (2.81, 17.51) | 0.006 |
| Controlled direct effect | 6.67 | 4.49 | (-2.13, 15.47) | 0.13 |
| Portion Eliminated | 3.49 | 3.75 | (-3.86, 10.84) | 0.34 |
| Effect of A on M ($\boldsymbol{DID}_{M_{gt},A_g}$) | 0.59 | 0.12 | (0.35, 0.83) | <0.001 |
| Effect of M on Y ($\boldsymbol{DID}_{Y_{gt},M_{gt}\mid A_g=1}$) | 7.73 | 2.69 | (2.46, 13.00) | 0.004 |
| Natural direct effect | 5.54 | 2.89 | (-0.12, 11.20) | 0.053 |
| Natural indirect effect | 4.62 | 1.71 | (1.27, 7.97) | 0.013 |
| Proportion eliminated | 0.34 | | | |
| Proportion mediated | 0.45 | | | |

**Table 2**. Estimation for all effects of the covid-19 severities on search volumes of "insomnia".

| | **Estimation** | **SD** | **95%CI** | **P value** |
|---|---|---|---|---|
| Total effect | 1.41 | 0.72 | (-0.004, 2.82) | 0.051 |
| Controlled direct effect =Natural direct effect | 1.00 | 0.74 | (-0.46, 2.45) | 0.17 |
| Portion Eliminated =Natural indirect effect | 0.41 | 0.18 | (0.06, 0.76) | 0.021 |
| Proportion Mediated = Proportion Eliminated | 0.29 | | | |