# Continuous Time Causal Mediation Analysis

**Jeffrey M. Albert**[1], **Youjun Li**[1], **Jiayang Sun**[1], **Wojbor A. Woyczynski**[2], **Suchitra Nelson**[3]

[1]Department of Population and Quantitative Health Sciences, School of Medicine, Case Western Reserve University, 10900 Euclid Ave., Cleveland, Ohio, U.S.A.

[2]Department of Mathematics, Applied Mathematics, and Statistics, Case Western Reserve University, 10900 Euclid Ave., Cleveland, Ohio, U.S.A.

[3]Department of Community Dentistry, School of Dental Medicine, Case Western Reserve University, 10900 Euclid Ave., Cleveland, Ohio, U.S.A.

## Abstract

While causal mediation analysis has seen considerable recent development for a single measured mediator ($M$) and final outcome ($Y$), less attention has been given to repeatedly measured $M$ and $Y$. Previous methods have typically involved discrete-time models that limit inference to the particular measurement times used, and do not recognize the continuous nature of the mediation process over time. To overcome such limitations, we present a new continuous time approach to causal mediation analysis that uses a differential equations model in a potential outcomes framework to describe the causal relationships among model variables over time. A connection between the differential equations models and standard repeated measures models is made to provide convenient model formulation and fitting. A continuous time extension of the sequential ignorability assumption allows for identifiable natural direct and indirect effects as functions of time, with estimation based on a two-step approach to model fitting in conjunction with a continuous time mediation formula. Novel features include a measure of an overall mediation effect based on the 'area between the curves', and an approach for predicting the effects of new interventions. Simulation studies show good properties of estimators and the new methodology is applied to data from a cohort study to investigate sugary drink consumption as a mediator of the effect of socioeconomic status on dental caries in children.

### Keywords

dental caries; differential equations; longitudinal data; mediation formula; potential outcomes

---

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available in the Web Supporting Information.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

## 1. INTRODUCTION

Mediation analysis seeks to determine the extent to which the effect of an exposure or intervention on a health outcome is due to its effect on one or more causally intermediate variables (or mediators). A goal of such an analysis is to illuminate the mechanisms through which the exposure or treatment affects the health outcome. An important related application is in the development of more effective, acceptable, and cost-effective interventions.

Recently, mediation analysis has been formulated using a potential outcome (causal model) framework. The resulting methodology, referred to as causal mediation analysis, provides, and elucidates the basis for, causally interpretable mediation (direct and indirect) effects. Causal mediation analysis for a single mediator has been addressed, for example, by Robins and Greenland (1992)[1], Albert (2008)[2], and Imai et al. (2010)[3]. Recent developments have involved more complex situations, including multiple (non-causally-ordered) mediators[4,5], causally ordered mediators[6–8], and a repeatedly measured mediator[9]. Most of these methods involve versions of the mediation formula[10] or the G formula[11]; an alternative approach uses a natural effect model in conjunction with inverse weighting[12]. In addition, though not based on a potential outcomes framework, structural equation model[13,14] and related (e.g., cross-lagged model[15]; and state space model[16]) methods for mediation analysis, possibly utilizing repeated measures, have been offered. An approach using linear mixed effects model has also been proposed[17].

Causal mediation analysis, and causal modelling in general, has predominantly relied on discrete time models. The prototypical mediation model is represented by the causal diagram (technically, a directed acyclic graph, or DAG[18]) given in Figure 1 (left). Here, we suppose that the exposure ($X$), mediator ($M$) and final outcome ($Y$) are measured at times $t_1$, $t_2$, and $t_3$, respectively, with $t_1 < t_2 < t_3$. For example, this model might describe the direct effect of a behavioral intervention on body mass index (BMI) in children and its indirect effect via diet change[19]. A discrete causal mediation model for longitudinal data (repeated measures of $X$, $M$, and $Y$, with the subscript indicating the time point) is represented in Figure 1 (right).

There has been increasing recognition that conventional (discrete time) mediation models are inadequate for explaining, or providing predictions related to, many social/behavioral and biological processes in health, which may often be seen as evolving in continuous time[20]. Aalen et al. (2014)[21], for example, showed in simulation studies that the use of a discrete mediation model, when the true model is continuous, can seriously distort estimates of mediation effects, thus, imply a null or small mediation effect, when in fact a large effect is present, or vice versa. In contrast to discrete time models, continuous time models (as represented in Figure 2) represent the underlying processes which are considered as existing prior to the selection of measurement times.

Aside from the possible distortion induced, discrete time models for mediation/path analysis involving continuous time processes have a number of other shortcomings. One is that inferences are restricted to the particular measurement times used, a consequence being that different measurement times imply different questions and are apt to produce different conclusions. In addition, discrete time models are generally ill–equipped to handle

longitudinal data with unbalanced measurement times, that is, where subjects have different measurements times. A further limitation is that discrete time models become increasingly cumbersome as the number of repeated measures increases. Some work to address these issues within the SEM framework, for example, using latent trend variables, has been done[22].

Unfortunately, there has been little development of causal models for the continuous time context. Most of this work has been directed at determining the causal effect of a treatment varying over time rather than mediation analysis per se[23–26]. Recently proposed methods for mediation analysis in continuous time have been based on ordinary differential equations models[21,27]. However, these methods are not based on a causal (in particular, potential outcomes) model and thus do not provide clear implications for causal inference. Also, they deal with contexts in which a differential equations model can be directly specified. This may not be the case in many health areas, including those involving psychosocial factors, in which continuous time mediation analysis may be of interest.

In this paper, we develop a novel causal approach to mediation analysis that recognizes model variables as continuous processes over time. We begin by presenting a new causal differential equations (CDE) model. Using differential equations involving potential outcomes, the model is flexible in allowing a variety of variable types. We relate this model to standard regression models for longitudinal data, thus allowing intuitive and familiar approaches to data fitting as part of a continuous time mediation analysis. We further show how this approach allows one to extend the notion of mediation to that of 'partial mediation' whereby the mediator is activated over part, but not all, of the time range of interest. We apply the new approach to a longitudinal cohort study of dental caries in early childhood and conclude with a discussion of its advantages and limitations along with directions for future research.

## 2. CONTINUOUS TIME MEDIATION MODEL

### 2.1 Model

We introduce what we refer to as the causal differential equation (CDE) model, beginning with the following general form:

$$dV_t^{(k)}(A)/dt = g_k\left[t, \bar{V}_t(A)\right] \quad (1)$$

where $V_t^{(k)}(A)$ is the potential outcome of the $k$th response variable ($k=1,\ldots,K$) at time $t \in [0, T]$ under (possibly time-varying) intervention $A$; $\bar{V}_t(A) = \bar{V}_t(\bar{A}_t)$ is the history of $V = \left(V^{(1)}\ldots,V^{(K)}\right)'$, a column vector of the model variables, under intervention $A$, up to time $t$; and $g_k$ is a specified function for the $k$th variable. Note that the $V^{(k)}$ may be expected or latent versions of the corresponding observed variables. This model may be supplemented by assumed probability distribution functions for the variables.

The CDE model yields an (instantaneous) causal effect of an intervention, $A^1$, versus another intervention, $A^0$, at some time $t$, expressed as the following difference in potential outcomes for some variable, $V$ (dropping the superscript $(k)$ for now),

$$\frac{dV_t(A^1)}{dt} - \frac{dV_t(A^0)}{dt} = \lim_{\Delta \to 0}\left[\{V_{t+\Delta}(\overline{A}^1_{t+\Delta}) - V_t(\overline{A}^1_t)\} - \{V_{t+\Delta}(\overline{A}^0_{t+\Delta}) - V_t(\overline{A}^0_t)\}\right]/\Delta \quad (2)$$

In other words, expression (2) represents the difference in the instantaneous change in the potential outcome of variable $V$ at time $t$, due to intervention $A^1$ versus intervention $A^0$.

For concreteness, and applicability to our later causal mediation problem (in our dental data example), we will focus on the following special case of the CDE model:

$$\frac{d\widetilde{M}_t(A)}{dt} = g_M\left[t, X^M_t, \boldsymbol{L^M}; \alpha\right] \quad (3)$$

$$\frac{d\widetilde{Y}_t(A)}{dt} = g_Y\left[t, X^Y_t, \widetilde{M}_t(A), \boldsymbol{L^Y}; \beta\right] \quad (4)$$

where $\widetilde{Y}_t$ and $\widetilde{M}_t$ denote individual-level expected values of the outcome and the mediator, respectively, at time $t$; $X^Y_t$ and $X^M_t$ are functions indicating the exposure level (affecting $Y$ and $M$, respectively) at $t$; $A$ is the 'intervention' (effectively defined by $X^Y_t$ and $X^M_t$); $\boldsymbol{L^M}$ and $\boldsymbol{L^Y}$ are vectors of baseline covariates, possibly including latent variables (e.g., random effects) predicting $M$ and $Y$, respectively; and $\alpha$ and $\beta$ are vectors of unknown parameters (which may sometimes be dropped in the notation).

By integrating both sides of the above differential equation, we obtain the following integral equations:

$$\widetilde{M}_t(A) = \widetilde{M}_0(A_0) + \int_0^t g_M\left[s, X^M_s, \boldsymbol{L^M}\right]ds \quad (5)$$

$$\widetilde{Y}_t(A) = \widetilde{Y}_0(A_0) + \int_0^t g_Y\left[s, X^Y_s, \widetilde{M}_s(A), \boldsymbol{L^Y}\right]ds \quad (6)$$

Since $X^M_t$, $X^Y_t$, and $\widetilde{M}_t(A)$ are generally functions of time, t, it will be helpful to define the following composite functions:

$$h_A^M(t) \equiv g_M\left[t, X_t^M, \boldsymbol{L^M}\right], \; h_A^Y(t) \equiv g_Y\left[t, X_t^Y, \widetilde{M}_t(X_t^M), \boldsymbol{L^Y}\right]. \quad (7)$$

If $h_A^M(t)$ and $h_A^Y(t)$ are continuous over $t \in [0, \; T]$, we obtain the models,

$$\widetilde{M}_t(A) = H_A^M(t) = G_M\left[t, X_t^M, \boldsymbol{L^M}\right] \quad (8)$$

$$\widetilde{Y}_t(A) = H_A^Y(t) = G_Y\left[t, X_t^Y, \widetilde{M}_t(A), \boldsymbol{L^Y}\right] \quad (9)$$

where $H_A^M(t)$ and $H_A^Y(t)$ are the antiderivatives of $h_A^M(t)$ and $h_A^Y(t)$, respectively, and $G_M$ and $G_Y$ are functions (of the indicated arguments) yielding the composite functions ($H_A^M$ and $H_A^Y$).

As a simple example, the differential equations,

$$\frac{d\widetilde{M}_t(A)}{dt} = \alpha_2 \exp\left(\alpha_0 + \alpha_1 x' + \alpha_2 t\right)$$

$$\frac{d\widetilde{Y}_t(A)}{dt} = \left[\beta_2 + \beta_3 \alpha_2 \exp\left(\alpha_0 + \alpha_1 x' + \alpha_2 t\right)\right] \cdot \exp\left[\beta_0 + \beta_1 x + \beta_2 t + \beta_3 \widetilde{M}_t(A)\right]$$

with $X_t^Y = x$ and $X_t^M = x'$ (i.e., both fixed over time), correspond to the integral equations,

$$\widetilde{M}_t(A) = \exp\left[\alpha_0 + \alpha_1 x' + \alpha_2 t\right]$$

$$\widetilde{Y}_t(A) = \exp\left[\beta_0 + \beta_1 x + \beta_2 t + \beta_3 \widetilde{M}_t(A)\right].$$

We may wish to directly model the integral, as opposed to the differential, equations, that is, $G_M$ and $G_Y$ rather than $g_M$ and $g_Y$. Particularly for health and behavioral data, it may be easier to specify a model of the expected values rather than derivatives of the expected values. This will be the emphasis of the present paper, in which we use suitable longitudinal models for the $G_M$ and $G_Y$ functions. However, the integral equations ((5) and (6)) motivate a new class of longitudinal models that handles certain extensions such as that of the next section. In other contexts, available scientific theory may allow direct specification of the CDE (i.e., $g_M$ and $g_Y$ functions).

## 2.2 Extension for Treatment Discontinuities

We wish to extend (8) and (9) to allow for $h_A^M(t)$ and/or $h_A^Y(t)$ being discontinuous. Such discontinuities may result from discontinuous changes in exposure or treatment over time. In fact, we will focus on the case of discontinuities in $X_t^Y$ and/or $X_t^M$ while the functions $g_Y$ and $g_M$ are continuous. For mediation analysis, this extension will enable, for example, an assessment of mediation through a treatment implemented over part - but not all - of the time range. Thus, we now make the relaxed assumption that $h_A^M(t)$ and $h_A^Y(t)$ are piecewise continuous over $[0, T]$ (thus, $H_A^M(t)$ and $H_A^Y(t)$ piecewise differentiable) and construct models based on (5) and (6). Supposing that $h_A^Y(t)$ has $d$ discontinuity points, $t_1, .., t_d$ in $(0, T)$ (so that $H_A^Y(t)$, the piecewise antiderivative of $h_A^Y(t)$, is piecewise differentiable for the resulting intervals) we write,

$$\widetilde{Y}_t(A) = \widetilde{Y}_0(A) + \sum_{j=0}^{d} I(t > t_j) G_Y\left[s, X_s^Y, \widetilde{M}_s(A), L^Y\right]\Big|_{t_j}^{t_j^* + 1} \quad (10)$$

where $t_0 \equiv 0$, $t_{d+1} \equiv T$, $t_j^* = \min\left(t, t_j\right)$, $I(t > t_j) = 1$ if $t > t_j$, $I(t > t_j) = 0$ otherwise, and $G_Y\left[t, X_t^Y, \widetilde{M}_t(A), L^Y\right] = h_A^Y(t)$ is the antiderivative of $g_Y\left[t, X_t^Y, \widetilde{M}_t(X_t^M), L^Y\right]$ for any constant $X_t^Y$ and $X_t^M$.

Each $\widetilde{M}_t(A)$ term would be expressed in a similar way using the piecewise differentiable function $H_A^Y(t)$ and a partition based on discontinuities in $h_A^M(t)$. Namely, when $h_A^M(t)$ has $r$ discontinuity points, $t_{M,1}, \ldots, t_{M,r}$ in $(0, T)$, we would have,

$$\widetilde{M}_t(A) = \widetilde{M}_0(A) + \sum_{j=0}^{r} I(t > t_{M,j}) G_M\left[s, X_s^M, L^M\right]\Big|_{t_{M,j}}^{t_{M,j+1}^*} \quad (11)$$

where $t_{M,0} \equiv 0$, $t_{M,r+1} \equiv T$, $t_{M,j}^* = \min\left(t, t_{M,j}\right)$, and $G_M\left[t, X_t^M, L^M\right] = h_A^M(t)$ is the antiderivative of $g_M\left[t, X_t^M, L^M\right]$ for any constant $X_t^M$.

To illustrate the above expressions, we consider a couple of simple examples also of interest for our later application.

Example 1. Suppose a person is exposed but the mediator set as if not exposed over the entire time interval, $[0, T]$. This intervention, denoted as $A^{1,0}$, can be described in previous notation as follows:

$$A^{1,0} \equiv \{X_t^Y = 1, X_t^M = 0; t \in \left[0, T\right]\} \equiv \left\{X^Y = 1, X^M = 0\right\}$$

resulting in potential outcomes, $\widetilde{Y}_t\left(A^{1,0}\right) = \widetilde{Y}_t\left(X^Y = 1,\ X^M = 0\right)$, for $t \in [0,\ T]$; note that we may drop the subscript $t$ in the notation in the case of a constant exposure status.

Because there are no discontinuities for this intervention, we can use expression (9) (a special case of (10)) along with (8) (a special case of (11)) to get,

$$\widetilde{Y}_t\left(A^{1,0}\right) = G_Y\left[\,t, X^Y = 1, \widetilde{M}_t\left(X^M = 0\right), \boldsymbol{L}^{\boldsymbol{Y}}\right]$$

for $t \in [0,\ T]$, with

$$\widetilde{M}_t\left(X^M = 0\right) = G_M\left[t,\ X^M = 0, \boldsymbol{L}^{\boldsymbol{M}}\right]$$

Example 2. Consider a person who is exposed but whose mediator is set as if exposed until time $t_1 < T$, and then as if not exposed starting at time $t_1$. This intervention will be denoted as $A^{1,t_1} \equiv \{X_t^Y = 1,\ X_t^M = \left(1[0, t_1),\ 0[t_1, T]\right)\}$.

As there is no discontinuity in $X_t^Y$ we have,

$$\widetilde{Y}_t\left(A^{1,t_1}\right) = G_Y\left[t, X_t^Y, \widetilde{M}_t\left(A^{1,t_1}\right), \boldsymbol{L}^{\boldsymbol{Y}}\right]$$

for $t \in [0,\ T]$. However, there is a discontinuity in $X_t^M$, so we use (11) to get,

$$\widetilde{M}_t\left(A^{1,t_1}\right) = G_M\left[t^*, X_{t^*}^M = 0,\quad L^M\right] + I(t > t_1) G_M\left[s, X_s^M = 1, L^M\right]\Big|_{t_1}^t$$

where $t^* = \min(t,\ t_1)$, $t \in [0,\ T]$.

We note that the theory up to this point allows exposure processes, $X_t^Y$ and $X_t^M$, to be continuous in each interval. However, our focus will generally be on the case of a binary exposure as is applicable in our data example.

## 2.3  Mediation Estimands

Mediation (that is, natural direct and indirect) effects may be formulated as appropriate contrasts (for example, differences or ratios) of expected potential outcomes of $Y$. For example, the natural indirect effect of $X_t$ on $\widetilde{Y}_t$ through $\widetilde{M}_t$ could be expressed as the difference in expected potential outcomes of $\widetilde{Y}_t$ under the two interventions, $A^1 = \{X_t^Y = 1,\ X_t^M = 1;\ t \in [0, T]\}$ and $A^{1,0} = \{X_t^Y = 1,\ X_t^M = 0;\ t \in [0, T]\}$ respectively, keeping in mind that the relevant part of an intervention for the outcome $\widetilde{Y}_t$ occurs up to time $t$. This contrast may be written as,

$$I_t\left(A^{1,0}\right) = E\left\{\widetilde{Y}_t\left(A^1\right)\right\} - E\left\{\widetilde{Y}_t\left(A^{1,0}\right)\right\}$$

$$= \int_l \left\{\widetilde{Y}_t\left(A^1\right) - \widetilde{Y}_t\left(A^{1,0}\right)\right\} f_L(l) d\mu_L(l) \quad (12)$$

where $f_L$ denotes the joint density function of the baseline covariate vector, $L$, obtained as the union of $L^Y$ and $L^M$. The estimand $I_t(A^{1,0})$ thus represents an indirect effect of exposure $X$ (continuously maintained at level 1 versus level 0) on $\widetilde{Y}$ at time $t$ though mediator $\widetilde{M}$, considered as a process, up to time $t$.

The corresponding natural direct effect, which also involves the intervention $A^0 \equiv \{X_t^Y = 0, \ X_t^M = 0; \ t \in [0, T]\}$, is,

$$D_t\left(A^{1,0}\right) = E\left\{\widetilde{Y}_t\left(A^{1,0}\right)\right\} - E\left\{\widetilde{Y}_t\left(A^0\right)\right\}$$

$$= \int_l \left\{\widetilde{Y}_t\left(A^{1,0}\right) - \widetilde{Y}_t\left(A^0\right)\right\} f_L(l) d\mu_L(l). \quad (13)$$

The preceding two effects yield the decomposition, $T_t = D_t(A^{1,0}) + I_t(A^{1,0})$, where $T_t = E\left\{\widetilde{Y}_t\left(A^1\right)\right\} - E\left\{\widetilde{Y}_t\left(A^0\right)\right\}$ is the total exposure effect at time $t$.

Alternative versions of these effects are obtained by defining the intervention, $A^{0,1} = \{X_t^Y = 0, \ X_t^M = 1; \ t \in [0, T]\}$. We then have $I_t\left(A^{0,1}\right) = E\left\{\widetilde{Y}_t\left(A^{0,1}\right)\right\} - E\left\{\widetilde{Y}_t\left(A^0\right)\right\}$ and $D_t\left(A^{0,1}\right) = E\left\{\widetilde{Y}_t\left(A^1\right)\right\} - E\left\{\widetilde{Y}_t\left(A^{0,1}\right)\right\}$, which comprise the alternative decomposition, $T_t = D_t(A^{1,0}) + I_t(A^{1,0})$. The discussion in Albert et al. (2018)[8] is relevant to the choice of estimands/decomposition.

As an extension alluded to in the previous section, we may wish to consider the indirect effect of exposure through the mediator over a reduced portion of the time period, thus a 'partial' indirect effect. For example, the intervention $A^{1,t_1}$ involves exposed individuals where the effect of the exposure on the mediator is removed starting at some time $t_1$. A corresponding indirect effect may be defined as,

$$I_t\left(A^{1,t_1}\right) = E\left\{\widetilde{Y}_t\left(A^1\right)\right\} - E\left\{\widetilde{Y}_t\left(A^{1,t_1}\right)\right\}. \quad (14)$$

This estimand is relevant to contexts, such as in our data example, where the start of the (future) intervention would correspond to stopping the exposure. An alternative estimand may be of interest when the intervention would involve initiation of exposure or treatment. Another possibility is an intervention that stops (rather than starts) at some time $t_1$.

The above estimands at any given time $t$ have a cross-sectional interpretation. We may also wish to assess the *overall* direct and indirect effects over the time range of interest. A useful summary measure is the 'area between the curves' (ABC). Specifically, this is the area between the curves given by the plots of the expected value of $\widetilde{Y}$ versus time for the two interventions implied by an indirect effect of interest. For example, the ABC for the total indirect effect (that is, corresponding to an intervention affecting $M$ over the whole time range) is given by

$$ABC\left(A^{1,0}\right) = \int_0^T \left[ E\left\{\widetilde{Y}_t\left(A^1\right)\right\} - E\left\{\widetilde{Y}_t\left(A^{1,0}\right)\right\}\right] dt. \quad (15)$$

Alternative versions are defined in an obvious manner, for example, $ABC\left(A^{0,1}\right) = \int_0^T \left[ E\left\{\widetilde{Y}_t\left(A^{0,1}\right)\right\} - E\left\{\widetilde{Y}_t\left(A^0\right)\right\}\right] dt$. The ABC measure is easily generalized for 'partial' indirect effects using (14); namely, we define,
$ABC\left(A^{1,t_1}\right) = \int_0^T \left[ E\left\{\widetilde{Y}_t\left(A^1\right)\right\} - E\left\{\widetilde{Y}_t\left(A^{1,t_1}\right)\right\}\right] dt$.

Alternative scales for natural direct and indirect effects may be used. A popular alternative to the mean difference scale given above is the mean ratio scale. For example, ratio-scale analogs to the mediation estimands, (12) and (13), are as follows:

$$I_t^r\left(A^{1,0}\right) = E\left\{\widetilde{Y}_t\left(A^1\right)\right\} / E\left\{\widetilde{Y}_t\left(A^{1,0}\right)\right\}$$

$$D_t^r\left(A^{1,0}\right) = E\left\{\widetilde{Y}_t\left(A^{1,0}\right)\right\} / E\left\{\widetilde{Y}_t\left(A^0\right)\right\}$$

and we obtain a decomposition of the total (ratio scale) effect at time $t$ as $T_t^r \equiv E\left\{\widetilde{Y}_t\left(A^1\right)\right\} / E\left\{\widetilde{Y}_t\left(A^0\right)\right\} = I_t^r\left(A^{1,0}\right) \bullet D_t^r\left(A^{1,0}\right)$. The alternative versions for the ratio scale are denoted in obvious notation as $I_t^r\left(A^{0,1}\right)$ and $D_t^r\left(A^{0,1}\right)$ and provide the alternative decomposition, $T_t^r = I_t^r\left(A^{0,1}\right) \bullet D_t^r\left(A^{0,1}\right)$. In models in which the alternative forms are equal we may write, for example, $I_t^r = I_t^r\left(A^{1,0}\right) = I_t^r\left(A^{0,1}\right)$; further, we may drop the subscript when the effect is constant over time, for example, $D^r = D^r(A^{1,0}) = D^r(A^{0,1})$.

For the area between the curves, a natural definition corresponding to the ratio scale, while maintaining the interpretation as an area, is based on the log-transformed expected values, for example, $\mathrm{ABC}^r\left(A^{1,0}\right) = \int_0^T \left[ \log E\left\{\widetilde{Y}_t\left(A^1\right)\right\} - \log E\left\{\widetilde{Y}_t\left(A^{1,0}\right)\right\}\right] dt$, and similarly for the

alternative version denoted as $\text{ABC}^r(A^{0,1})$. These definitions are readily generalized to the interventions starting at time $t_1$, for example,

$$ABC^r\left(A^{1,t_1}\right) = \int_0^T \left[\log E\left\{\widetilde{Y}_t\left(A^1\right)\right\} - \log E\left\{\widetilde{Y}_t\left(A^{1,t_1}\right)\right\}\right]dt.$$ In the case where $I_t^r\left(A^{1,0}\right) = I_t^r\left(A^{0,1}\right)$ for $t$

$\in [0, T]$, we use the abbreviated notation, $\text{ABC}^r = \text{ABC}^r(A^{1,0}) = \text{ABC}^r(A^{0,1})$.

For the indirect effects (including ABCs) for either scale we may also define mediation (or indirect effect) proportions. For instance, the indirect effect proportions (for intervention $A$ and at time $t$) for the difference and ratio scales are defined as $I_t(A)(\text{prop}) \equiv I_t(A)/(D_t(A) + I_t(A))$ and $I_t^r(A)(\text{prop}) \equiv \log(I_t^r(A))/\{\log(I_t^r(A)) + \log(D_t^r(A))\}$, respectively. Similarly, for areas between the curves we define proportions on the difference and ratio scales as, for example,

$\text{ABC}\left(A^{1,t_1}\right)(\text{prop}) = \text{ABC}\left(A^{1,t_1}\right)/\text{ABC}\left(A^0\right)$ and $\text{ABC}^r\left(A^{1,t_1}\right)(\text{prop}) = \text{ABC}^r\left(A^{1,t_1}\right)/\text{ABC}^r\left(A^0\right)$,

where $\text{ABC}\left(A^0\right) \equiv \int_0^T \left[E\left\{\widetilde{Y}_t\left(A^1\right)\right\} - E\left\{\widetilde{Y}_t\left(A^0\right)\right\}\right]dt$ and

$\text{ABC}^r\left(A^0\right) \equiv \int_0^T \left[\log E\left\{\widetilde{Y}_t\left(A^1\right)\right\} - \log E\left\{\widetilde{Y}_t\left(A^0\right)\right\}\right]dt$.

## 2.4  Identification and Inference

The estimands given above involve expected values of $\widetilde{Y}_t(A)$ for various interventions, that is, particular specifications of $A$. We will demonstrate the identifiability of these expected potential outcomes by showing that they can be written as functions of (estimable) association model parameters under certain assumptions. Our approach is an extension of the mediation formula approaches of Pearl (2001)[28] and Imai et al., 2010[3] (see also Albert and Nelson, 2011[6], and Daniel et al., 2015[7]).

Our assumptions are as follows, starting with a continuous version of the standard consistency assumption:

Assumption 1. Consistency:

$$\begin{aligned} \widetilde{Y}_t(x, m) &= \widetilde{Y}_t \text{ if } X_t = x \text{ and } \widetilde{M}_t = m, \text{ and} \\ \widetilde{M}_t(x') &= \widetilde{M}_t \text{ if } X_t = x' \end{aligned}$$

for all $x$, $x'$, $m$ and $t \in [0, T]$. That is, the potential outcomes for $\widetilde{Y}_t$ and $\widetilde{M}_t$ where the levels of the causal variables are set to those observed (for a given individual), are equal to the observed (or actual latent) values for $\widetilde{Y}_t$ and $\widetilde{M}_t$, respectively.

We also assume a continuous time version of the sequential ignorability assumption:

Assumption 2. Sequential ignorability in continuous time:

$$\left\{\widetilde{Y}_t(x, m), \widetilde{M}_t(x')\right\} \coprod X_t \Big| L = l \quad (16)$$

$$\widetilde{Y}_t(x, m) \coprod \widetilde{M}_t(x') \Big| X_t = x', L = l. \quad (17)$$

for all $x$, $x'$, $m$, and $l$, and any $t \in [0, T]$.

Expression (16) states that potential outcomes for $\widetilde{Y}$ and $\widetilde{M}$ at a given time, $t$, are independent of the observed exposure ($X$) at time $t$, given the baseline variables $L$; similarly, (17) states that potential outcomes for $\widetilde{Y}$ and $\widetilde{M}$ at time $t$ are independent given $X_t$ and $L$. In other words, it is assumed that there are no unobserved confounders among the model variables ($X$, $M$, and $Y$) at any time. An accompanying assumption is that of positivity, that is, $P(X_t = x | L = l) > 0$ and $P(\widetilde{M}_t(x) = m | X_t = x, \quad L = l) > 0$, for $x = 0, 1$, and $m$ and $l$ in their respective support sets. While this sequential ignorability assumption may appear to be strong (as it often is considered to be in the discrete time case), we note that the vector of baseline confounders ($L$) may include latent variables, as utilized in our data example discussed later.

In Web Appendix A, we show, given the above assumptions, that the expected potential outcome for a constant intervention can be expressed using the following version of the mediation formula:

$$E\left\{\widetilde{Y}_t\left(x, \widetilde{M}_t(x')\right)\right\} = \int_l \int_m E\left\{\widetilde{Y}_t \Big| X = x, \widetilde{M}_t = m, L = l\right\} f_{\widetilde{M}_t \big| X = x', L = l}(m|x', l) f_L\left(l\right) d\mu_{\widetilde{M}_t}\left(m\right) d\mu_L\left(l\right)$$

$$(18)$$

for $t \in [0, T]$. More general (including discontinuous) intervention functions require additional assumptions. Web Appendix A also shows for such cases that expected potential outcomes for $Y$ can be identified under the causal differential equations model, that is, (5)-(9).

Estimation of the expected potential outcomes using the continuous time mediation formula (18) is done in conjunction with association models for $M$ and $Y$. As a concrete example, we consider the following model used for the dental data described in the next section:

$$\widetilde{M}_{ij} \equiv E\{M_{ij} | t_{ij}, x_i, c_i, u_i^M\} = \frac{\exp[\alpha_0 + \alpha_1 x_i + \boldsymbol{\alpha}_2 \boldsymbol{c_i} + u_i^M]}{\left[1 + \exp\{-\alpha_3 t_{ij}\}\right]^{1/\nu_M}} \quad (19)$$

$$\widetilde{Y}_{ij} \equiv E\{Y_{ij}|t_{ij}, x_i, c_i, \widetilde{M}_{ij}, u_i^Y\} = \frac{\exp[\beta_0 + \beta_1 x_i + \boldsymbol{\beta}_2 \boldsymbol{c_i} + \beta_3 \log(\widetilde{M}_{ij}) + u_i^Y]}{\left[1 + \exp\{-(\beta_4 t_{ij} + \beta_5 t_{ij} \bullet \log(\widetilde{M}_{ij}))\}\right]^{1/\nu}} \quad (20)$$

Equations (19) and (20) are nonlinear regression models based on the generalized logistic function[29]; for convenience, we refer to them as generalized logistic models. Although (19) and (20) have specific forms pertaining to our data example, variations (for example, including other interaction terms, a time-dependent exposure, $X$, and different measurement times for $Y$ and $M$) are possible. We also utilize natural continuous-time extensions of models (19) and (20). Thus we write $\widetilde{M}_{it}$ and $\widetilde{Y}_{it}$ to refer to expected values for $M$ and $Y$, respectively, for individual $i$ at any given time $t$.

Note the important role of the random effects in (19) and (20) in increasing the plausibility of the sequential ignorability assumptions (in particular, (17)). From the DAG on the right-hand side of Figure 1, it is apparent that previous observations of the mediator ($M$) represent confounders of the relationship between $M$ and $Y$ at a later time. We assume that the random effects explain any associations (further, any casual relationships) among the repeated measures for $M$, and likewise for $Y$. The assumed causal model is thus represented in Figure S1 (Supporting Information, Appendix A). We see for this DAG (under the nonparametric structural equation model (NPSEM) interpretation[18]) that assumptions (16) and (17), with $u^M$ and $u^Y$ included in $L$, are satisfied.

The association models for $M$ and $Y$ may be fit jointly, but we consider a computationally faster two-step approach, making the additional assumption of independent $u_i^M$ and $u_i^Y$. A Monte Carlo approach is used in lieu of integration over $\widetilde{M}_t$ in the mediation formula. Note, as indicated above, that the baseline covariate vector, $L$, is considered to include both the observable baseline variables and random effects; that is, $L = (c, u^M, u^Y)$. The algorithm for estimation of expected potential outcomes is as follows:

1.  Fit association models for $M$ and $Y$.

    a.  Step 1: Fit (e.g., via maximum likelihood) the $M$ model (19) based on an assumed distribution for $M$ and the data: repeated responses, $M_{ij}$, the observed exposure ($x_i$) and baseline covariates ($c_i$) for $i = 1,\ldots,N, j = 1,\ldots,n_i$. Obtain predicted values for the $u_i^M$ yielding predicted values for $\log(\widetilde{M}_{ij})$ (denoted as $\widehat{M}_{ij}$) for each individual $i$ and time $t_{ij}$.

    b.  Step 2. Fit the $Y$ model (20) under an assumed distribution for $Y$ using $Y_{ij}$ (the repeated measures) $x_i, c_i$ and $\widehat{M}_{ij}$ for $i = 1,\ldots,N, j = 1,\ldots,n_i$.

        We thus obtain estimates of regression parameters (the $\alpha$'s and $\beta$'s from (19) and (20), respectively) as well as the random effects variances, with estimates denoted with hats (e.g., $\widehat{\sigma}_M^2$).

2. For a given intervention, compute estimated expected potential outcomes via the mediation formula.

Do the following over a grid of values of time, $t \in [0, T]$, by some small interval size $\delta$:

For a specified intervention, $A$ (say with $X_t^Y = x, X_t^M = x'$), and time, $t$, do the following independently for each person $i = 1$ to $N$ (with covariate vector $c_i$):

    a.    draw $\hat{u}_i^M$ from $N\left(0, \hat{\sigma}_M^2\right)$ and (independently) $\hat{u}_i^Y$ from $N\left(0, \hat{\sigma}_Y^2\right)$;

    b.    compute predicted value (denoted as $\widehat{M}_{it}(A)$) of $\log\left(\widetilde{M}_{it}(A)\right)$ given $\hat{u}_i^M$ and $c_i$ with $x_i = x'$ and time $t$ using continuous time version of (19) with estimates plugged in for parameters;

    c.    compute predicted value (denoted as $\widehat{Y}_{it}(A)$) of $\widetilde{Y}_{it}(A)$ given $\hat{u}_i^Y$, $\widehat{M}_{it}(A)$, and $c_i$, with $x_i = x$ and time $t$ using continuous time version of (20) with estimates plugged in for parameters;

    d.    An estimate of the marginal effect is obtained by summing over $i$; namely, $\widehat{E}\left\{\widetilde{Y}_t(A)\right\} = \sum_{i=1}^{N} \widehat{Y}_{it}(A)$.

To reduce Monte Carlo error, one may 'clone' the sample (or other 'reference group') using a chosen multiplier as suggested in previous work[7,8]. The above describes the approach for fixed $X_t^M$ and $X_t^Y$ (which would be similar for continuous but not fixed exposures). For discontinuous $X_t^M$ and $X_t^Y$ the above steps can also be implemented where the appropriate $\widehat{Y}_{it}(A)$ and $\widehat{M}_{it}(A)$ are used based on expressions (10) and (11). Web Appendix B provides an illustrative derivation of formulae for expected potential outcomes for a discontinuous intervention, namely, $A^{1,t_1}$ (defined in Section 2.2), in the context of the generalized logistic model.

The estimated potential outcomes are then used to obtain estimated mediation effects for a chosen scale. We note that for the ratio scale, in contrast to the difference scale, we have $D_t^r\left(A^{0,1}\right) = D_t^r\left(A^{1,0}\right)$ and $I_t^r\left(A^{0,1}\right) = I_t^r\left(A^{1,0}\right)$, which obtains when there is not an $X$ by $M$ interaction in the $Y$ model. Further, in the present model, the ratio-scale direct effect is fixed over time (and given by $\exp(\beta_1)$), thus, denoted simply as $D^r$, while the indirect effect (written as $I_t^r$) varies over time due to the inclusion of the $M$ by $t$ interaction term. Once indirect effects are computed over a grid of times, the area between the curves (ABC) corresponding to a given type of indirect effect of interest may be obtained using the trapezoid method.

Confidence intervals for the specified direct and indirect effects (at any given time) can be obtained via bootstrap resampling. In particular, we use the bootstrap percentile method to obtain 95 percent confidence intervals.

## 3.  SIMULATION STUDY

We conducted a simulation study to evaluate properties of our mediation effect estimates. Our main scenario roughly mimics the dental data (to be presented and analyzed in Section 4) based on the generalized logistic model described above. However, using the same model we also wish to learn the implications of varying the number of observed measurements per person.

We simulated data using models (19) and (20), with both $Y$ and $M$ as negative binomial, $X$ as Bernoulli, and a single baseline covariate $C$ as normally distributed. As $C$ represents a confounder it is involved in the generation of $X$ (using a logistic regression model). Random effects for $M$ and $Y$ (namely, $u^M$ and $u^Y$) were generated independently from normal distributions. A time variable, $t$, was also included. Models (19) and (20) were then used to generate $M$ (as a function of specified $t$ and generated $x$, $c$, and $u^M$) and $Y$ (as a function of $t$, $x$, $c$, $u^M$, and $u^Y$). A total sample size of 200 was used. The number of measurements per person (equally spaced in the time interval [0, 40]) were 3, 6, and 11. The parameter values used in the simulations are provided in Web Appendix C, Table S1. Five hundred independent replicates (datasets) were generated and analyzed as described below.

We included both ratio and difference scale mediation effect estimators. The method and formulae are given in Section 2. A multiplier of 10 ('cloning' the sample) was used to reduce Monte Carlo error. Bootstrap (percentile method) 95 percent confidence intervals were obtained using 300 bootstrap samples (sampling data vectors corresponding to subjects in each dataset). Occasionally, generated values (usually in combination with extreme estimates from certain bootstrap samples) would result in a large (nonevaluable) argument for the exponential function or power operation, which we remedied through truncation of the relevant argument. In addition, some datasets resulting in non-convergence in the fit of either the $M$ or $Y$ model resulting in exclusions. From the 500 replicates minus the exclusions (the numbers of which were recorded), the statistics listed below were computed.

The true values for each estimand were obtained by applying the continuous time mediation formula (18) using the true values for the regression parameters, the empirical distribution of the covariates for the given dataset, and drawing random effects (in a Monte Carlo approach) from their true distributions (yielding generated $\log\left(\widetilde{M}_{it}(A)\right)$ for given $A$, $c_i$, and $t$). Note that the generated (empirical) covariate distribution for a dataset was considered as the true covariate distribution and dataset-specific true values were used in computing the biases and coverage probabilities. For each scenario, we computed (averaging over replications): bias (average estimate minus the true value), relative bias (average ratio of the bias and the true value), standard error of the bias, coverage (percent of 95% confidence intervals that cover the true value) and power (percent of 95% confidence intervals that do not cover 0).

Tables 1 and 2 provides the simulation results for the ratio-scale estimands. Note that on the ratio scale the (natural) direct effect is constant over time and the two versions for each effect (e.g., $D(A^{1,0})$ and $D(A^{1,0})$) are equal.

From Table 1, we see that relative biases are low (less than 6%) for direct and indirect effect (including ABC) estimates even with as few as 3 time points. Proportions (including that for the ABC) tend to be less stable and have high (simulation estimated) relative biases (up to 65%) and high standard error of bias at $m$=3. However, these biases are reduced considerably with higher $m$. The estimated relative biases for the ABC proportion are 10.2% and 5.5% for 6 and 11 time points, respectively.

Table 2 provides simulation results for coverage and power for the same (ratio-scale) estimands. Coverage is found to be good (at least 92%) even for $m = 3$, though coverage for indirect effect proportions are somewhat conservative for smaller $m$. Closer to nominal (95%) coverage is found for $m = 6$ and this is closer still for $m$=11. Power for all estimands is seen to increase with increasing $m$, particularly in going from 3 to 6 time points.

The equalities noted above for the alternative versions for the ratio-scale estimands do not hold on the difference scale; consequently, there are a greater number of distinct difference-scale estimands. Nevertheless, the overall conclusion for these (see results provided in Web Appendix C, Tables S2, S3) are generally consistent with those of the ratio scale.

## 4.   DATA EXAMPLE

The data for the present example are from a longitudinal study of dental caries in a cohort of very low birth weight (VLBW) and normal birth weight (NBW) children followed from birth[30]. In this study 468 child-caregiver dyads (234 VLBW; 234 NBW) were enrolled and assessed at child ages 8, 18, and 36 months on oral health outcomes, as well as behavioral and demographic variables, including socioeconomic status (SES).

A secondary finding of this study was that there is a relationship between SES and dental caries as measured by the number of decayed, filled and missing teeth (DMFT). As an exploratory question, we sought to learn the extent to which the effect of SES (considered as a fixed binary variable, 1 for low SES, 0 for high SES) on DMFT is mediated by the child's consumption of sugary drinks. The latter was obtained via a caregiver questionnaire which included questions about frequency of the child's consumption of such sugary drinks as soft drinks and juice. The caregiver was asked to respond to each question on a five-level Likert scale (1=none to 5=very often). We calculated a sugary drink score (denoted as SDRK) as the mean score over the relevant questions. A table of descriptive statistics for the model variables is provided in Web Appendix D (Table S4). The analysis was based on 440 child-caregiver dyads that had complete data for the included covariates (noted below) and at least one measurement (over the three time points) each of SDRK and DMFT. Similar results (not shown) were obtained using the complete cases ($n$=195).

Although SDRK and DMFT were measured at only three time points, these measurements are considered as realizations of underlying (latent) continuous time processes. We therefore wished to determination the extent of mediation as a function of age and to assess the overall mediational effect over the age range of interest.

For the longitudinal association models for $M$ (SDRK) and $Y$ (DMFT) we used the generalized exponential models given by (19) and (20), respectively. For both models, the

vector of baseline covariates, $c$, included birth group (1 for very low birth weight group, 2 for normal birth weight group), and sex (1 for male, 0 for female). Both $Y$ and $M$ were assumed to be distributed as negative binomial.

The generalized logistic model is appropriate for the dental data in part because it describes the means for $M$ and $Y$ (conditional on the covariates and random effects) as monotonically increasing over time, as would be expected for sugary drink use and DMFT in young children. Initial models for $M$ and $Y$ included time by SES ($t \times X$) interactions in the numerator; however, these terms did not appear to improve the model fit based on AIC (and were not found to be statistically significant) and were dropped from the models.

The estimates (with standard errors) for the $M$ and $Y$ model parameters are given in Table 3. Figure 3 shows the predicted means of potential outcomes for DMFT, including those corresponding to the total indirect and partial indirect effects as a function of time ($t$, age in months). The natural direct and indirect effects were computed on both the difference and ratio scales. Estimates and 95% confidence intervals for mediation effects on the ratio scale are provided in Table 4. Bootstrap percentile confidence intervals were based on 499 bootstrap samples, with possible exclusions due to non-convergence. A multiplier of 10 ('cloning' the sample) was used to reduce Monte Carlo error in the mediation formula computations. Results on the mean difference scale (for which there a larger number of estimands) are provided in Web Appendix D (Table S5).

As an example, we see from Table 4 that the (time independent) ratio-scale estimate of the natural direct effect (and 95% confidence interval (CI)) is 2.05 (1.02, 4.19). Thus, at any given time over the observed time range, there is an estimated factor increase of 2.1 in mean DMFT for low SES versus high SES when the mediator, SDRK, takes values as if everyone were high (or low) SES. The ratio-scale estimates of the natural indirect effect (and 95% CI), at 18 and 36 months, respectively, are 0.91 (0.72, 1.76) and 1.31 (1.04, 2.11). The latter result indicates an estimated 1.3 factor increase in mean DMFT at 36 months of age if everyone were high (equivalently, low) SES, and SDRK (over the whole time range) were at the level each person would have if low SES versus the level each person would have if high SES. The corresponding estimated mediation proportions (95% CIs) are −0.14 (−0.88, 0.85) and 0.27 (0.03, 0.93) for 18 and 36 months respectively. Thus, an estimated 27 percent of the effect of SES on DMFT at 36 months is through SDRK (considered continuously up to that time); from the 95% confidence interval this is seen to be nominally statistically significant at the 0.05 $\alpha$ level.

The estimated mediation proportions on the difference scale (see Web Supporting Information, Appendix D, Table S5) are similar overall, and the conclusions regarding statistical significance are the same for both scales. Our contention is that the ABCs based on the difference scale are more meaningful than those of the ratio scale as the former corresponds to areas based on means (as in Figure 3) rather than log means. For the ABC on the difference scale, in particular ABC($A^{1,0}$), corresponding to a total mediation effect, the estimated proportion (95% CI) is 0.19 (−0.16, 0.94). The lack of statistical significance for the ABC, in contrast to the indirect effect at 36 months, may be due to the greater variability

of the former; also the proportion is lower (for $A^{1,0}$, 0.19 versus 0.37 (difference scale)), presumably due to the lower (and even negative) estimated indirect effects at younger ages.

For partial mediation where the mediator is affected (by a future intervention) starting at 8 months, the estimated overall mediation proportion (ABC – difference scale, 95% CI) is 0.11 (−0.16, 0.70). If instead the intervention begins at 18 months the estimated ABC proportion (95% CI) is 0.01 (−0.053, 0.18). Thus an estimated 11% (1%) of the effect of SES on DMFT up to age 36 months is due to SDRK at 8 (18) months and later. As expected, these proportions are smaller than that of the complete mediation situation (where an intervention affecting SDRK starts at birth) and are also not statistically significant. However, it is interesting that starting the intervention affecting SDRK at 18 months (to a lesser extent, 8 months) provides a dramatic reduction in the predicted mediation proportion relative to starting the intervention at time 0 (birth).

For comparison, we also analyzed these data using a roughly analogous discrete time approach (Bind et. al., 2016).[31] Note that this approach provides estimates of *overall* natural direct and indirect effects, assumed as constant over the measurement times. This is in contrast to our continuous time approach which considers natural direct and indirect effects as functions of time. We found that results from the discrete time approach are consistent with those of the continuous time approach, in the sense that the former gives estimates that are similar to those found from the continuous time approach at around the midpoint of the time range for the data. Details and further discussion are given in the Web Supporting Information, Appendix F.

To critically evaluate a particular data analysis using the proposed continuous time mediation analysis approach, it will be important to carefully consider the assumptions of the method. In the Web Supporting Information, Appendix E, we provide an elaborated discussion of key assumptions, both causal and for the association model, and discuss their plausibility for the dental data. For the untestable continuous time sequential ignorability assumption, just as with the established discrete version, it will be desirable in practice to perform a sensitivity analysis. Unfortunately, a sensitivity analysis expressly designed for the continuous time mediation model is not yet available. However, a rough approach is possible using a recently developed sensitivity analysis method for a discrete time causal mediation analysis with outcomes (*Y* and *M*) following generalized linear models.[32] The results of the sensitivity analysis suggest that the conclusion of a statistically significant natural direct effect is sustained, while conclusions regarding natural indirect effects change, over a plausible range of sensitivity parameter values. Further details are provided in the Web Supporting Information, Appendix G.

## 5. DISCUSSION

In this paper we present an approach to causal mediation analysis for longitudinal data using continuous time models. In particular, we introduce a causal (potential outcome based) differential equations (CDE) model to account for underlying mediator (*M*) and final outcome (*Y*) processes that are continuous over time. We consider an easily integrable class of functions so that expected outcomes can be related to standard (albeit nonlinear)

longitudinal regression models. Extension to wider classes of functions for differential equations models will be of interest for the future.

Our methodology is in the spirit of the mediation formula approach for estimation of natural direct and indirect effects; identification of these effects is obtained under a continuous time version of the commonly used sequential ignorability assumption. We specified nonlinear longitudinal regression (association) models based on the generalized logistic function for both the mediator ($M$) and final outcome ($Y$). A two-step approach to estimation of the association model parameters is proposed for computational ease. Alternative association models may be considered and it is possible that results will be sensitive to model misspecification. Thus, the use of model diagnostics and criticism is an important preliminary step. Nonparametric or semiparametric models will be of interest to better assure robustness of results.

While it may appear that the CDE model is not involved in the constant exposure case once the longitudinal models (such as (19) and (20)) are specified, in fact this model still provides the underpinning for the mediation process and resulting inference (which becomes clearer in the discontinuous exposure case). For example, the predicted values for $Y$ at some time $t$ are not entirely determined by the current values for $X$ and $M$ (that is, $\log(\widetilde{M})$) as seems to be specified by (19) and (20); rather those predicted values are dependent on the exposure history (in the present case, exposure being constant at the indicated levels, $x$ and $x'$ for $Y$ and $M$, respectively) from time 0, as revealed by the CDE (or its integral form).

As elaborated in Section 2.4, using the ratio scale in the context of the generalized logistic model, provides intuitive estimators of natural direct and indirect effects, in which time-independence (or dependence) follows directly from the specification of the corresponding association models (that is, the inclusion or not of certain interaction terms). Thus, a recently expressed criticism of the mediation formula methodology[33], namely the complicated and unintuitive nature of interaction effects, is circumvented in this case.

In addition, the selected association models use random effects to explain correlations among repeated measures. The random effects for the mediator and final outcome are assumed to be independent in our two-stage modelling approach. Another limitation is that individual trajectories are fixed conditional on the random effects, constraining the dynamic potential of the model. Thus, while our use of random effects allows for some randomness in outcome trajectories, the current model may not be adequate for certain types of causal relationships, for example, where there are lagged effects. On the other hand, at least in our example, we believe that the lack of a lag is plausible as we are describing the relationship among the underlying processes; for example, average (smoothed) sugary drink usage may have an immediate, albeit infinitesimal, effect on the underlying dental caries process. In future work, we will seek to expand the flexibility of the approach by employing stochastic differential equations to describe mediation in the context of dynamic changes over time.

Our simulation studies showed good properties of the estimators, indicating that the two-step approach works reasonably well. This is true despite the approach not accounting for measurement error due to the use of estimated $\log(\widetilde{M}_{it}(A))$ in the $Y$ model. There is the

potential for more refined approaches, for example, using joint modeling, though computational challenges would need to be overcome. Our simulations also show the potential for gains in precision with increasing numbers of (equally spaced) measurements. We found the method to be useful in a dental data example despite the relatively small number of (three) time points. The data used in our example and a SAS macro to implement the method are provided in the Web Supplementary Information.

Novel features of our continuous time method include the use of the 'area between the curves' (ABC) as a measure of the overall mediation effect. In addition, we are able to consider alternative ('partial') mediation effects corresponding to interventions that affect the mediator starting (or stopping) at selected post-baseline times. In this way, the continuous time methodology allows for prediction of effects of more refined and realistic future interventions. Further variations of conceived interventions, for example, with additional starts and/or pauses of treatment, may be readily implemented.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## ACKNOWLEDGEMENTS

## REFERENCES

1. Robins JM, Greenland S. Identifiability and exchangeability for direct and indirect Effects. Epidemiology. 1992;3:143–155. [PubMed: 1576220]

2. Albert JM. Mediation analysis via potential outcomes models. Stat Med. 2008;27:1282–1304. [PubMed: 17691077]

3. Imai K, Keele L, Yamamoto T. Identification, Inference and Sensitivity Analysis for Causal Mediation Effects. Stat Sci. 2010;25:51–71.

4. Wang W, Nelson S, Albert JM. Estimation of causal mediation effects for a dichotomous outcome in multiple-mediator models using the mediation formula. Stat Med. 2013;32:4211–4228. [PubMed: 23650048]

5. Taguri M, Featherstone J, Cheng J. Causal mediation analysis with multiple causally non-ordered mediators. Stat Methods Med Res. 2015;27:3–19. [PubMed: 26596350]

6. Albert JM, Nelson S. Generalized Causal Mediation Analysis. Biometrics. 2011;67:1028–1038. [PubMed: 21306353]

7. Daniel RM, De Stavola BLD, Cousens SN, Vansteelandt S. Causal mediation analysis with multiple mediators. Biometrics. 2015;71:1–14. [PubMed: 25351114]

8. Albert JM, Cho JI, Liu Y, Nelson S. Generalized causal mediation and path analysis: Extensions and practical considerations. Stat Methods Med Res. 2018:1–15.

9. VanderWeele TJ, Tchetgen Tchetgen EJ. Mediation analysis with time varying exposures and mediators. J R Stat Soc Ser B. 2017;79:917–938.

10. Pearl J The causal mediation formula-a guide to the assessment of pathways and mechanisms. Prev Sci. 2012;13:426–436. [PubMed: 22419385]

11. Robins J A new approach to causal inference in mortality studies with a sustained exposure period-application to control of the healthy worker survivor effect. Math Model. 1986;7:1393–1512.

12. Lange T, Rasmussen M, Thygesen LC. Assessing natural direct and indirect effects through multiple pathways. Am J Epidemiol. 2014;179:513–518. [PubMed: 24264291]

13. Bollen KA. Structural Equations with Latent Variables. John Wiley & Sons, Inc; 1989.

14. Farrell AD. Structural equation modeling with longitudinal data: Strategies for examining group differences and reciprocal relationships. J Consult Clin Psychol. 1994;62:477–487. [PubMed: 8063974]

15. Rogosa D A critique of cross-lagged correlation. Psychol Bull. 1980;88:245–258.

16. Gu F, Preacher KJ, Ferrer E. A State Space Modeling Approach to Mediation Analysis. J Educ Behav Stat. 2014;39:117–143.

17. Blood EA, Cheng DM. The use of mixed models for the analysis of mediated data with time-dependent predictors. J Environ Public Health. 2011:1–12.

18. Pearl J Causality: Models, Reasoning, and Inference (2nd Edition). Cambridge University Press; 2009.

19. Moore SM, Borawski EA, Cuttler L, Ievers-Landis CE, Love TE. IMPACT: A multi-level family and school intervention targeting obesity in urban youth. Contemp Clin Trials. 2013;36:574–586. [PubMed: 24008055]

20. Deboeck PR, Nicholson JS, Bergeman CS, Preacher KJ. From modeling long-term growth to short-term fluctuations: Differential equation modeling is the language of change Springer Proc Math Stat. 2013:427–447. Springer New York.

21. Aalen OO, Røysland K, Gran JM, Kouyos R, Lange T. Can we believe the DAGs? A comment on the relationship between causal DAGs and mechanisms. Stat Methods Med Res. 2016;25:2294–2314. [PubMed: 24463886]

22. Muthén B, Asparouhov T. Causal Effects in Mediation Modeling: An Introduction With Applications to Latent Variables. Struct Equ Model A Multidiscip J. 2015;22:12–23.

23. Gill RD, Robins JM. Causal inference for complex longitudinal data: The continuous case. Ann Stat. 2001;29:1785–1811.

24. Lok JJ. Statistical modeling of causal effects in continuous time. Ann Stat. 2008;36:1464–1507.

25. Zhang M, Joffe MM, Small DS. Causal inference for continuous-time processes when covariates are observed only at discrete times. Ann Stat. 2011;39:131–173.

26. Zhang M, Small DS. Effect of vitamin a deficiency on respiratory infection: Causal inference for a discretely observed continuous time non-stationary Markov process. Can J Stat. 2012;40:646–662.

27. Deboeck PR, Preacher KJ. No need to be discrete: A method for continuous time mediation analysis. Struct Equ Model A Multidiscip J. 2016;23:61–75.

28. Pearl J. Direct and indirect effects Proc seventeenth Conf Uncertain Artif Intell. 2001:411–420. Morgan Kaufmann Publishers Inc http://dl.acm.org/citation.cfm?id=2074073.

29. Richards FJ. A flexible growth function for empirical use. J Exp Bot. 1959;10:290–301.

30. Nelson S, Albert JM, Soderling E, et al. Increased number of teeth predict acquisition of mutans streptococci in infants. Eur J Oral Sci. 2014;122:346–352. [PubMed: 25183438]

31. Bind MAC, Vanderweele TJ, Coull BA, Schwartz JD. Causal mediation analysis for longitudinal data with exogenous exposure. Biostatistics. 2016;17:122–134. [PubMed: 26272993]

32. Albert JM, Wang W. Sensitivity analyses for parametric causal mediation effect estimation. Biostatistics. 2015;16:339–351. [PubMed: 25395683]

33. Lange T, Vansteelandt S, Bekaert M. A simple unified approach for estimating natural direct and indirect effects. Am J Epidemiol. 2012;176:190–195. [PubMed: 22781427]
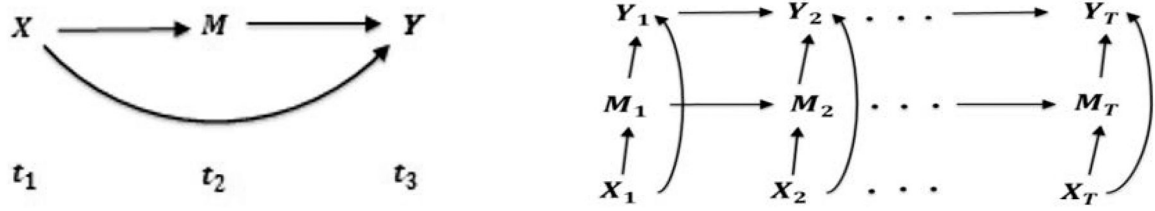
Author Manuscript Author Manuscript Author Manuscript Author Manuscript

**Figure 1.**
Discrete mediation models (left: single measurement, right: repeated measures)
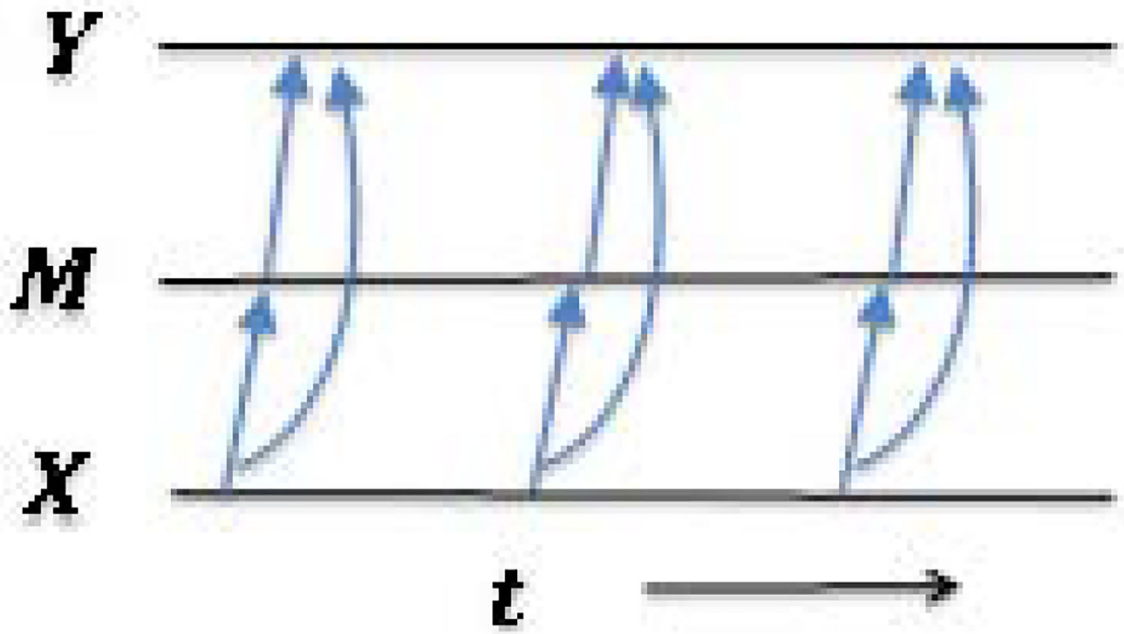
**Figure 2.**
Mediation in continuous time

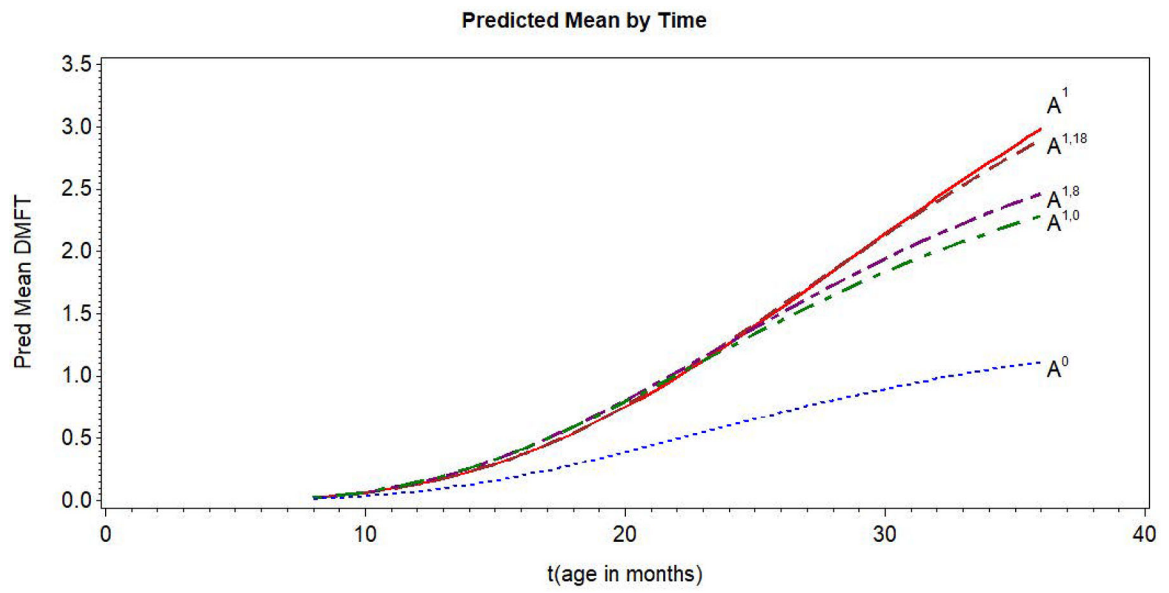**Predicted Mean by Time**



**Figure 3.**
Plot of mean predicted DMFT by age of child (t) using the generalized logistic model, (18) and (19), for alternative interventions (indicated by different $A$'s).

**Table 1.**

Simulation estimated bias, relative bias, and standard error of bias (based on 500 replicates mimicking the dental data, $n$=200) for mediation (natural direct ($D$) and indirect ($I$)) effects on ratio scale; subscripts (20 and 40) indicate times for indirect effects; $m$ indicates the number of (equally spaced) measurements for simulated data. Note: all 500 replicates were used for $m = 6$ and 11; one resulted in non-convergence (499 were used) for $m = 3$

| Estimand | True | Bias | | | Rel Bias | | | SE Bias | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | $m = 3$ | 6 | 11 | $m = 3$ | 6 | 11 | $m = 3$ | 6 | 11 |
| $D^r$ | 1.350 | 0.068 | 0.036 | 0.022 | 0.051 | 0.026 | 0.017 | 0.021 | 0.015 | 0.012 |
| $I_{20}^r$ | 1.653 | −0.033 | 0.003 | −0.018 | −0.018 | 0.003 | −0.009 | 0.013 | 0.012 | 0.011 |
| $I_{40}^r$ | 1.279 | 0.060 | 0.053 | 0.038 | 0.048 | 0.042 | 0.031 | 0.009 | 0.008 | 0.007 |
| ABC$^r$ | 13.403 | −0.480 | −0.821 | −0.578 | −0.030 | −0.057 | −0.037 | 0.237 | 0.189 | 0.175 |
| $I_{20}^r$(prop) | 0.625 | 0.107 | 0.028 | 0.027 | 0.174 | 0.047 | 0.045 | 0.056 | 0.034 | 0.010 |
| $I_{40}^r$(prop) | 0.449 | 0.220 | 0.140 | 0.065 | 0.445 | 0.315 | 0.157 | 0.321 | 0.051 | 0.026 |
| ABC$^r$ (prop) | 0.598 | −0.368 | 0.060 | 0.032 | −0.650 | 0.102 | 0.055 | 0.483 | 0.021 | 0.011 |

**Table 2.**

Simulation estimated coverage and power (based on 500 replicates mimicking the dental data, $n$=200, from percentile method 95% confidence intervals, 300 bootstrap samples) for mediation (natural direct ($D$) and indirect ($I$)) effects on ratio scale; subscripts (20 and 40) indicate time points for indirect effects; $m$ indicates the number of measurements. All 500 replicates used for $m$ = 6 and 11; 499 used for $m$ = 3

| Estimand | Coverage (%) | | | Power (%) | | |
|---|---|---|---|---|---|---|
| | $m$ = 3 | 6 | 11 | $m$ = 3 | 6 | 11 |
| $D^r$ | 0.954 | 0.948 | 0.952 | 0.164 | 0.254 | 0.286 |
| $I^r_{20}$ | 0.948 | 0.924 | 0.942 | 0.862 | 0.978 | 0.982 |
| $I^r_{40}$ | 0.948 | 0.932 | 0.932 | 0.808 | 0.958 | 0.978 |
| ABC$^r$ | 0.920 | 0.916 | 0.942 | 0.794 | 0.916 | 0.972 |
| $I^r_{20}$(prop) | 0.970 | 0.958 | 0.946 | 0.575 | 0.864 | 0.952 |
| $I^r_{40}$(prop) | 0.990 | 0.972 | 0.958 | 0.415 | 0.654 | 0.782 |
| ABC$^r$(prop) | 0.972 | 0.958 | 0.946 | 0.589 | 0.826 | 0.918 |

**Table 3.**

Estimates, standard errors, and p-values (Wald test) of parameters in generalized logistic models for $M$ (SDRK) and $Y$ (DMFT) fit to dental data

| | **$M$ Model** | | | | **$Y$ Model** | | |
|---|---|---|---|---|---|---|---|
| **Parameter** | **Estimate** | **SE** | **$p$-value** | **Parameter** | **Estimate** | **SE** | **$p$-value** |
| $a_0$ (Int) | 0.18 | 0.28 | 0.53 | $\beta_0$ (Int) | −2.50 | 0.81 | 0.002 |
| $a_1$ (SES) | 0.77 | 0.15 | <0.001 | $\beta_1$ (SES) | 0.69 | 0.46 | 0.13 |
| $a_{21}$ (Birth) | 0.22 | 0.14 | 0.12 | $\beta_{21}$ (Birth) | 0.29 | 0.37 | 0.44 |
| $a_{22}$ (Sex) | −0.077 | 0.14 | 0.59 | $\beta_{22}$ (Sex) | −0.046 | 0.37 | 0.90 |
| $a_3$ (Time) | 0.17 | 0.027 | <0.001 | $\beta_3$ (SDRK) | 1.15 | 0.38 | 0.003 |
| $k_M$ (Disp) | 0.38 | 0.080 | <0.001 | $\beta_4$ (Time($t$)) | 0.21 | 0.057 | <0.001 |
| $\nu_M$ (Damp) | 0.12 | 0.023 | <0.001 | $\beta_5$ ($t \times$ SDRK) | −0.057 | 0.022 | 0.012 |
| $\sigma_M$ (SD) | 0.57 | 0.11 | <0.001 | $k$ (Disp) | 1.64 | 0.67 | 0.015 |
| | | | | $\nu$ (Damp) | 0.047 | 0.023 | 0.047 |
| | | | | $\sigma_Y$ (SD) | 2.85 | 0.92 | 0.002 |

**Table 4.**

Results for analysis of dental data – inference for mediation (natural direct ($D$) and indirect ($I$)) estimands on mean ratio scale. Estimates and, for indirect effects, proportions are given along with bootstrap percentile method 95% confidence intervals (499 bootstrap samples generated, 489 used). Subscripts indicate time (age in months) for (cross-sectional) indirect effects. ABC (area between the curves) provides overall indirect effect for specified intervention (starting at indicated time)

| Estimand | Est | 95% CI | Prop. | 95% CI |
|---|---|---|---|---|
| $D^r$ | 2.05 | (1.02, 4.19) | - | - |
| $I_8^r$ | 0.98 | (0.53, 1.61) | −0.03 | (−4.35, 1.87) |
| $I_{18}^r$ | 0.91 | (0.72, 1.76) | −0.14 | (−0.88, 0.85) |
| $I_{36}^r$ | 1.31 | (1.04, 2.11) | 0.27 | (0.03, 0.93) |
| $\text{ABC}^r(A^{1,0})$ | 0.64 | (−6.27, 16.39) | 0.03 | (−0.38, 0.89) |
| $\text{ABC}^r(A^{1,8})$ | 0.01 | (−4.40, 11.83) | 0.0007 | (−0.26, 0.56) |
| $\text{ABC}^r(A^{1,18})$ | −0.03 | (−0.63, 1.58) | −0.002 | (−0.03, 0.08) |