Statistics in Medicine

WILEY

Statistics in Medicine

RESEARCH ARTICLE OPEN ACCESS

A Path-Specific Effect Approach to Mediation Analysis With Time-Varying Mediators and Time-to-Event Outcomes Accounting for Competing Risks

Arce Domingo-Relloso[1,2,3] | Yuchen Zhang[1] | Ziqing Wang[1] | Astrid M. Suchy-Dicey[4] | Dedra S. Buchwald[5] | Ana Navas-Acien[2] | Joel Schwartz[6] | Kiros Berhane[1] | Brent A. Coull[7] | Linda Valeri[1,8]

[1]Department of Biostatistics, Columbia University Mailman School of Public Health, New York, New York, USA | [2]Department of Environmental Health Sciences, Columbia University Mailman School of Public Health, New York, New York, USA | [3]School of Science and Technology, IE University, Madrid, Spain | [4]Department of Epidemiology, Boston University Chobanian & Avedisian School of Medicine, Boston, Massachusetts, USA | [5]Neurosciences Institute, Department of Neurological Surgery, University of Washington, Seattle, Washington, USA | [6]Department of Environmental Health Sciences, Harvard TH Chan School of Public Health, Boston, Massachusetts, USA | [7]Department of Biostatistics, Harvard TH Chan School of Public Health, Boston, Massachusetts, USA | [8]Department of Epidemiology, Harvard TH Chan School of Public Health, Boston, Massachusetts, USA

Correspondence: Arce Domingo-Relloso (ad3531@cumc.columbia.edu)

Received: 6 March 2025 | **Revised:** 15 January 2026 | **Accepted:** 19 January 2026

Keywords: competing risks | g-formula | longitudinal data | mediation analysis | path-specific effects | survival analysis

ABSTRACT

Not accounting for competing events in survival analysis can lead to biased estimates, as individuals who die from other causes do not have the opportunity to develop the event of interest. Formal definitions and considerations for causal effects in the presence of competing risks have been published, but not for the mediation analysis setting when the exposure is not separable and both the outcome and the mediator are nonterminal events. We propose, for the first time, an approach based on the path-specific effects framework to account for competing risks in longitudinal mediation analysis with time-to-event outcomes. We do so by considering the pathway through the competing event as another mediator, which is nested within our longitudinal mediator of interest. We provide a theoretical formulation and related definitions of the effects of interest based on the mediational g-formula, as well as a detailed description of the algorithm. We also present a simulation study and an application of our algorithm to data from the Strong Heart Study, a prospective cohort of American Indian adults. In this application, we evaluated the mediating role of the blood pressure trajectory (measured in three visits) on the association of arsenic and cadmium with time to cardiovascular disease, accounting for competing risks by death. Identifying the effects through different paths enables us to evaluate the impact of metals on the outcome of interest, as well as through competing risks, more transparently.

1 | Introduction

In survival settings, competing events refer to any event that makes it impossible for the event of interest to occur. For example, death is a competing event for any other event, as when an individual dies, he/she cannot experience any further events. Not accounting for competing events in survival analysis can lead to bias caused by the fact that individuals that die from other causes do not have the opportunity to develop the event of interest [1]. This has been clearly shown, for example, in the context of aging-related diseases such as dementia. As reported in Rojas-Saunero et al. [2], counterintuitive findings have been

observed when factors such as smoking [3] or cancer history [3] showed protective effects on dementia when ignoring death. This does not mean that those factors are protective of dementia, but because smokers and cancer patients tend to die earlier, this premature mortality makes them less likely to develop dementia. There is an urgent need to develop new frameworks to effectively acknowledge competing risks, and to report causal effects that take their impact into account.

Various statistical estimands have been proposed for competing risks in failure-time settings including marginal cumulative incidence, cause-specific cumulative incidence, marginal hazard, subdistribution hazard and cause-specific hazard [4]. However, Young et al. [5] was the first paper to present a formal definition of causal effects in the presence of competing risks using counterfactuals. In that paper, the authors described two different estimands that can be considered in the presence of competing risks: the direct effect, defined as the risk under elimination of competing events, and the total effect, defined as the risk without elimination of competing events. Martinussen and Stensrud [6] proposed to address the issue of competing events by considering a hypothetical scenario in which we have a treatment separable in two components, the first only affects the event of interest and the second only affects the competing event. Huang et al. approached the problem of mediation analysis in the presence of semi-competing risks [7], while Tai et al. used principal stratification to define the effects of interest in the presence of survival outcomes and truncated mediators [8]. However, those works considered death as the primary outcome. In our setting, the primary outcome is nonterminal, with potential competing risks by death or other outcomes.

In this work, we propose a framework for the causal mediation setting in presence of mediators measured repeatedly over time. We do this by considering the competing event as an additional mediator that is affected by past values of the mediator of interest (i.e., nested within the primary mediator process of interest), and in turn affects future values of both the mediator and the outcome deterministically. Thus, we quantify the strength of the competing event by evaluating the indirect effect through death. This path-specific effect structure has previously been described in the literature in the context of nested mediators (i.e., not in the context of competing risks) [9], but the authors did not consider time-varying nested mediators in that work. One of the key differences between our approach and previously considered approaches to deal with competing events is that we do not consider the framework in which causal effects are calculated under elimination of the competing event nor a framework in which the exposure can be separated into components that only affect the terminal event, as these scenarios are not realistic in many settings in which competing events by death simply cannot be eliminated or separated.

The paper is organized as follows: in Section 2, we introduce the data structure and notation related to our causal setting. In Section 3, we introduce the counterfactual estimands of causal mediation analysis in the presence of competing events. In Section 4, we describe the identifiability assumptions. Section 5 describes the estimation of the effects. Section 6 is a full description of our algorithm. Section 7 includes the simulation study. In Section 8, we apply the approach to data from the Strong Heart Study, a prospective cohort study of American Indian adults. This analysis evaluated the potential mediated effect of blood pressure (longitudinal mediator) on the association between arsenic and cadmium (exposures, in separate models) and cardiovascular disease incidence (CVD) (outcome), accounting for competing events by death. Finally, Section 8 includes the discussion.

2 | Data Structure and Notation

Consider $i = 1, \ldots, n$ individuals exposed to different levels of a continuous exposure A at baseline. Individuals are assumed to be independent and identically distributed. Let $M_{i1}, \ldots, M_{iK}$ denote longitudinal measurements of a mediator for the i-th individual at visits $1, \ldots, K$. For simplicity, we denote this vector as $M_1, \ldots, M_K$ hereinafter. In addition, let L_0 represent a vector of baseline covariates, and Y_k, D_k, and S_k denote indicators of the event of interest (e.g., CVD), a competing event (e.g., death from all causes other than CVD) and survival by time interval k, respectively. Please note that $D_k + S_k = 1$. Our causal framework is depicted in the directed acyclic graph (DAG) shown in Figure 1. By definition, $D_0 \equiv Y_0 \equiv 0$ and $S_0 \equiv 1$, because the study population is restricted to those who have not yet experienced the event of interest or the competing event at baseline. We assume, without loss of generality, the temporal ordering (A, M_k, D_k, S_k, Y_k) within each follow-up visit $k > 0$.

Following the notation used in Young et al. [5], we denote the history of a random variable using overbars, for example, $\overline{M}_k = (M_1, \ldots, M_k)$ is the history of the mediator of interest through interval k. Note that, if an individual is known to experience the competing event by visit $k > 0$ without having experienced the event of interest ($Y_{k-1} = 0, D_k = 1$), then all future indicators of the event of interest will be observed and will be deterministically zero because, by definition, individuals that experience the competing event cannot experience the event of interest. Please also note that, if the event of interest Y happens at time k, repeated measures of the mediator of interest are only considered until time $k - 1$, as the first occurrence of Y is considered our final event. Following Figure 1, the competing event indicator at a certain time point k (D_k) can be considered as another mediator which is nested within the vector for the mediator process M, in the sense that it is influenced by past values of the mediator of interest and in turn influences future values of the mediator of interest (in a deterministic way, because if an individual dies, the mediator is death-truncated).

3 | Counterfactual Estimands of Causal Mediation Analysis in Presence of Competing Events

We used the Lin-Ying additive hazards model [10], a special case of the Aalen additive hazards model with time-invariant coefficients, to model our time-to-event outcome. While Cox proportional hazards is a more widely used model in the survival setting, the additive model has the advantage that it avoids the noncollapsibility of the hazard ratio, which has made this model more popular for the mediation analysis setting [11]. Thus, most of the prior work in mediation analysis has been focused on additive models [12], which are collapsible. This additive model allows

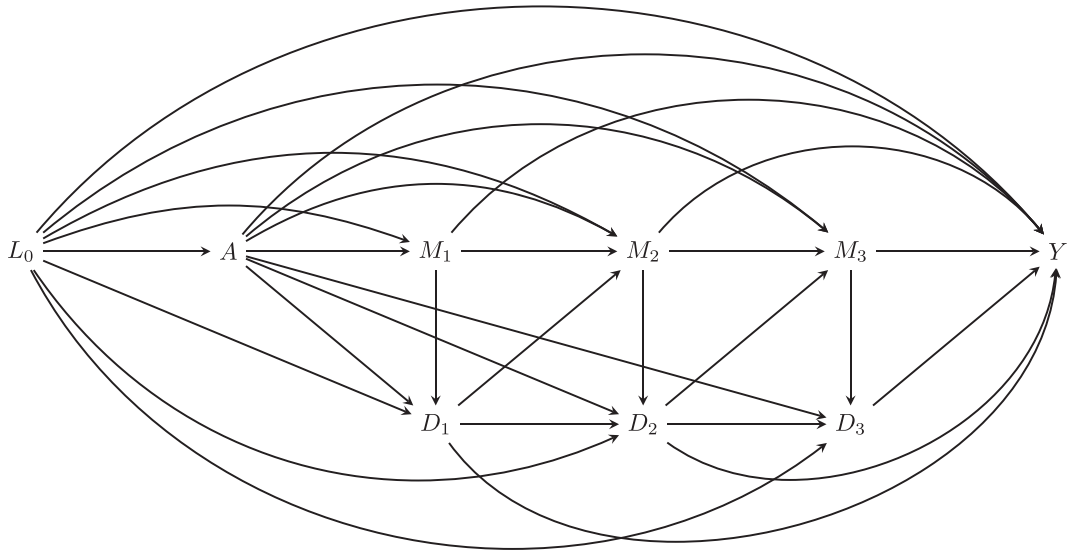

FIGURE 1 | Directed acyclic graph for $K = 3$. A represents the exposure, L_0 represents a vector of baseline confounders, M is the mediator of interest, measured at three time points (M_1, M_2, M_3), D is the competing risk indicator, measured at three time points (D_1, D_2, D_3), and Y is the time-to-event outcome. Please note that no causal associations are depicted between M_1 and D_3, M_2 and D_3, M_1 and M_3, M_1 and D_2, and D_1 and D_3, because we assume that our process is Markovian and therefore each longitudinal value of M and D is only affected by immediately prior values of M and D. This assumption can be relaxed within the proposed approach.

us to define our counterfactual outcome in terms of absolute attributable risks (i.e., number of cases of the outcome of interest per a number of person-years), which is an absolute measure relevant for public health, but also in terms of differences in survival probability. Thus, our effects are defined either in differences in hazards, which represent differences in instantaneous risk of the disease, or in differences in survival probability, which represent differences in cumulative probability of being free of the disease over a period. While our choice is to use an additive model in order to have these two different measures of effect, our model could be extended to Cox and other survival models. For times $k = 0, \ldots, K$, consider the counterfactual history of the mediator of interest $\overline{M}_{k+1}\left(a, \overline{S}_k\left(a, \overline{M}_k(\ldots)\right), \overline{M}_k\left(a, \overline{S}_{k-1}(\ldots)\right)\right)$, where $\overline{M}_k(\ldots)$ and $\overline{S}_{k-1}(\ldots)$ denote recursive dependency on previous values of M and S. This counterfactual history refers to the value the history of the mediator would take had the individual, possibly contrary to fact, been exposed to $A = a$ and to the survival status $\overline{S}_k(a)$. Consider also the counterfactual history of the survival status $\overline{S}_k\left(a, \overline{M}_k\left(a, \overline{S}_{k-1}(\ldots)\right), \overline{S}_{k-1}\left(a, \overline{M}_{k-1}(\ldots)\right)\right)$, which is considered as another mediator, and is denoted as $\overline{S}_k\left(a, \overline{M}_k(a)\right)$ for simplicity. In addition, consider the counterfactual outcome $\gamma\left(a, \overline{M}_{k+1}\left(a, \overline{S}_k(a)\right), \overline{S}_k\left(a, \overline{M}_k(a)\right)\right)$, which refers to the number of cases attributable to the exposure per a number of person-years (typically set to 10 000 or 100 000) at time k; and the counterfactual outcome $P\left(Y_{k+1}\left(a, \overline{M}_k\left(a, \overline{S}_{k-1}(a)\right), \overline{S}_k\left(a, \overline{M}_k(a)\right)\right) = 0\right)$, which refers to the probability of being free of the outcome at time $k + 1$. In additive hazards models, the hazard or rate function γ for a given time $k + 1$ is given by:

$$\gamma\left(k+1; a, M_k\left(a, S_{k-1}(a)\right), S_{k-1}\left(a, M_{k-1}(a)\right)\right)$$
$$= \lim_{dk \to 0} \frac{1}{dk} P\Big(Y\left(a, M_k\left(a, S_{k-1}(a)\right), S_{k-1}\left(a, M_{k-1}(a)\right)\right) \in]k$$
$$+1, k+1+dk] \,\Big|\, Y\left(a, M_k\left(a, S_{k-1}(a)\right), S_{k-1}\left(a, M_{k-1}(a)\right)\right) > k+1\Big)$$

The temporal dependencies between the mediators are detailed in the DAG presented in Figure 1. For instance, if $D_2 = 1$, then $S_2 = 0$, and therefore M_3 is not observed. Thus, S_2 deterministically impacts M_3 and in general, M_k is dependent on S_{k-1}. Similarly, M_2 will affect the probability of dying at visit 2, thus, S_k is dependent on M_k.

We propose to decompose our causal mediation effects of interest in a path-specific effects approach to account for competing risks. Therefore, we define, for certain levels a and a^* of the exposure of interest, the following path-specific effects on the rate difference, as well as survival probability difference, scales:

1. Direct effect of the exposure on the outcome (DE): effect that operates neither through mortality nor through the longitudinal mediator of interest M (Figure S1).

$$\gamma\left(a^*, \overline{M}_k\left(a, \overline{S}_{k-1}(a)\right), \overline{S}_k\left(a, \overline{M}_k(a)\right)\right)$$
$$- \gamma\left(a, \overline{M}_k\left(a, \overline{S}_{k-1}(a)\right), \overline{S}_k\left(a, \overline{M}_k(a)\right)\right),$$

   being γ the rate obtained from an additive hazards model, and

$$P\left(Y_{k+1}\left(a^*, \overline{M}_k\left(a, \overline{S}_{k-1}(a)\right), \overline{S}_k\left(a, \overline{M}_k(a)\right)\right) = 0\right)$$
$$- P\left(Y_{k+1}\left(a, \overline{M}_k\left(a, \overline{S}_{k-1}(a)\right), \overline{S}_k\left(a, \overline{M}_k(a)\right)\right) = 0\right).$$

2. Indirect effect through the history of the mediator of interest (IEM), when the death process behaves under the reference level of exposure (Figure S2).

$$\gamma\left(a^*, \overline{M}_k\left(a^*, \overline{S}_{k-1}(a)\right), \overline{S}_k\left(a, \overline{M}_k(a^*)\right)\right)$$
$$- \gamma\left(a^*, \overline{M}_k\left(a, \overline{S}_{k-1}(a)\right), \overline{S}_k\left(a, \overline{M}_k(a)\right)\right),$$

and

$$P\left(Y_{k+1}\left(a^*, \overline{M}_k\left(a^*, \overline{S}_{k-1}(a)\right), \overline{S}_k\left(a, \overline{M}_k(a^*)\right)\right) = 0\right) - P\left(Y_{k+1}\left(a^*, \overline{M}_k\left(a, \overline{S}_{k-1}(a)\right), \overline{S}_k\left(a, \overline{M}_k(a)\right)\right) = 0\right).$$

This is the effect that operates by impacting the mediator of interest directly, which in turn might impact the outcome (i) directly or (ii) via the mortality process. Please note that separating (i) and (ii) is impossible unless we make additional modeling (parametric) assumptions (see Discussion section).

3. Indirect effect through the history of the competing event (IED): effect that operates by impacting the death process directly, which has a deterministic relationship with the mediator of interest and the outcome, in the sense that the mediator is unobserved, and the outcome does not happen if death happens before (Figure S3).

$$\gamma\left(a^*, \overline{M}_k\left(a^*, \overline{S}_{k-1}(a^*)\right), \overline{S}_k\left(a^*, \overline{M}_k(a^*)\right)\right) - \gamma\left(a^*, \overline{M}_k\left(a^*, \overline{S}_{k-1}(a)\right), \overline{S}_k\left(a, \overline{M}_k(a^*)\right)\right),$$

and

$$P\left(Y_{k+1}\left(a^*, \overline{M}_k\left(a^*, \overline{S}_{k-1}(a^*)\right), \overline{S}_k\left(a^*, \overline{M}_k(a^*)\right)\right) = 0\right) - P\left(Y_{k+1}\left(a^*, \overline{M}_k\left(a^*, \overline{S}_{k-1}(a)\right), \overline{S}_k\left(a, \overline{M}_k(a^*)\right)\right) = 0\right).$$

4. TE:

$$\gamma\left(a^*, \overline{M}_k\left(a^*, \overline{S}_{k-1}(a^*)\right), \overline{S}_k\left(a^*, \overline{M}_k(a^*)\right)\right) - \gamma\left(a, \overline{M}_k\left(a, \overline{S}_{k-1}(a)\right), \overline{S}_k\left(a, \overline{M}_k(a)\right)\right),$$

and

$$P\left(Y_{k+1}\left(a^*, \overline{M}_k\left(a^*, \overline{S}_{k-1}(a^*)\right), \overline{S}_k\left(a^*, \overline{M}_k(a^*)\right)\right) = 0\right) - P\left(Y_{k+1}\left(a, \overline{M}_k\left(a, \overline{S}_{k-1}(a)\right), \overline{S}_k\left(a, \overline{M}_k(a)\right)\right) = 0\right).$$

Sum of direct effect and indirect effect through the mediator of interest (Figure S4):

$$\gamma\left(a^*, \overline{M}_k\left(a^*, \overline{S}_{k-1}(a)\right), \overline{S}_k\left(a, \overline{M}_k(a^*)\right)\right) - \gamma\left(a, \overline{M}_k\left(a, \overline{S}_{k-1}(a)\right), \overline{S}_k\left(a, \overline{M}_k(a)\right)\right),$$

and

$$P\left(Y_{k+1}\left(a^*, \overline{M}_k\left(a^*, \overline{S}_{k-1}(a)\right), \overline{S}_k\left(a, \overline{M}_k(a^*)\right)\right) = 0\right) - P\left(Y_{k+1}\left(a, \overline{M}_k\left(a, \overline{S}_{k-1}(a)\right), \overline{S}_k\left(a, \overline{M}_k(a)\right)\right) = 0\right)$$

Please note that the sum of the three path-specific effects leads to the TE:

$$TE = DE + IEM + IED$$

Figures S1–S4 show the DAGs with the highlighted paths corresponding to each of our effects of interest. A comparison of the interpretation of these estimands with those obtained in traditional mediation analysis can be found in Table 1. Please note that, in presence of competing risks, our definition of the TE, $\mathbb{E}\left(Y_{k+1}\left(a^*, \overline{S}_k(a^*)\right)\right) - \mathbb{E}\left(Y_{k+1}\left(a, \overline{S}_k(a)\right)\right)$ is different to the definition of the TE in classical survival models, $\mathbb{E}\left(Y_{k+1}(a^*)\middle|\overline{S}_k = 1\right) - \mathbb{E}\left(Y_{k+1}(a)\middle|\overline{S}_k = 1\right)$, in the sense that classical survival models do not consider death as an informative censoring event [13], and they calculate the effect conditioning on survival up to time k, whereas the TE accounting for competing risks calculates the effect when setting survival to the level it would take at the level of exposure a versus a^*. Conversely, the sum of the direct effect and the indirect effect through M, $\mathbb{E}\left(Y_{k+1}\left(a^*, \overline{S}_k(a)\right)\right) - \mathbb{E}\left(Y_{k+1}\left(a, \overline{S}_k(a)\right)\right)$, does not condition on survival, but sets survival to the level it would take under the reference level of exposure (see Table 1). The differences between these effects are also illustrated in our analysis of the Strong Heart Study data in Section 8.

4 | Effect Estimation and Identifiability Assumptions

The following assumptions need to be fulfilled for the effects described in the previous section to be identifiable.

4.1 | Exchangeability

No unmeasured confounding of the relationship between the exposure and the history of the outcome, and the exposure and the survival status:

$$\left(Y_{k+1}(a^*, \overline{s}_k = 1, \overline{m}_k), \overline{S}_k(a^*, \overline{m}_k)\right) \perp A | L_0.$$

1. No unmeasured confounding of the relationship between the history of the mediator and both the outcome and the survival status at time $k+1$: $\left(Y_{k+1}(a^*, s_k = 1, m_k), S_k(a^*, m_k)\right) \perp M_k | A, L_0, \overline{Y}_k = 0, \overline{S}_{k-1} = 1, \overline{M}_{k-1} = \overline{m}_{k-1}$
2. No unmeasured confounding of the relationship between the mediator of interest and the survival status at each time point: $S_{k+1}(a, m) \perp M_k | A, L_0, \overline{Y}_k = 0, \overline{S}_k = 1, \overline{M}_{k-1} = \overline{m}_{k-1}$
3. No unmeasured confounding of the relationship between the history of the mediator and the exposure: $\overline{M}_{k+1}(a, \overline{s} = 1) \perp A | L_0$
4. Cross-world assumption: no confounders of the mediators-outcome relationship affected by the exposure. $Y_{k+1}(a, \overline{m}, \overline{s} = 1) \perp \left(\overline{M}_k(a^*, \overline{s} = 1), \overline{S}_k(a^*, \overline{m})\right) | L_0, \overline{Y}_k = 0, \overline{S}_{k-1} = 1, \overline{M}_{k-1} = \overline{m}_{k-1}$
5. Conditioning on the mediator and the competing event induces no additional backdoor paths between the exposure and the outcome: $Y_{k+1}(a, \overline{m}, \overline{s} = 1) \perp A | \overline{M}_k, \overline{S}_k, L_0$

Please note that the cross-world assumption would be violated if time-varying confounders exist. However, our approach can be extended to the setting in which we can consider time-varying confounders by adding additional paths to the path-specific

TABLE 1 | Interpretation of the path-specific effects and comparison to effects obtained from traditional mediation analysis.

	Interpretation	Formula
Total effect (classical survival analysis)	Effect of the exposure on the outcome conditional on survival under the observed exposure: *we treat death as a noninformative censoring event*	$\mathrm{E}\left(Y_{k+1}(a^*)\middle\vert\overline{S}_{k+1}=1\right)-\mathrm{E}\left(Y_{k+1}(a)\middle\vert\overline{S}_{k+1}=1\right)$[a]
Total effect (accounting for competing risks)	Effect of the exposure on the outcome: *we treat death as an informative censoring event*. Intervention on both exposure and survival	$\mathrm{E}\left(Y_{k+1}\left(a^*,\overline{S}_{k+1}(a^*)\right)\right)-\mathrm{E}\left(Y_{k+1}\left(a,\overline{S}_{k+1}(a)\right)\right)$
Sum of direct + indirect effect through the mediator of interest	Effect of the exposure on the outcome fixing survival to the value it would take under the exposure $A=a$ (for survival under low levels of exposure): no intervention on survival	$\mathrm{E}\left(Y_{k+1}\left(a^*,\overline{S}_{k+1}(a)\right)\right)-\mathrm{E}\left(Y_{k+1}\left(a,\overline{S}_{k+1}(a)\right)\right)$
Indirect effect through the mediator of interest	Effect of the exposure on the outcome through the mediator when survival is fixed to the value it would take under the reference exposure level	$\mathrm{E}\left(Y_{k+1}\left(a^*,\overline{S}_{k+1}(a),\overline{M}_k(a^*)\right)\right)-\mathrm{E}\left(Y_{k+1}\left(a^*,\overline{S}_{k+1}(a),\overline{M}_k(a)\right)\right)$
Indirect effect through death	Effect of the exposure on the outcome through the death process when the mediator is fixed to the value it would take under the reference exposure level	$\mathrm{E}\left(Y_{k+1}\left(a^*,\overline{S}_{k+1}(a^*),\overline{M}_k(a^*)\right)\right)-\mathrm{E}\left(Y_{k+1}\left(a^*,\overline{S}_{k+1}(a),\overline{M}_k(a^*)\right)\right)$
Direct effect	Effect of the exposure on the outcome through other pathways (when both the mediator and the death process are fixed to the value they would take under the reference exposure level)	$\mathrm{E}\left(Y_{k+1}\left(a^*,\overline{S}_{k+1}(a),\overline{M}_k(a)\right)\right)-\mathrm{E}\left(Y_{k+1}\left((a),\overline{S}_{k+1}(a),\overline{M}_k(a)\right)\right)$

[a]The traditional total effect assumes that death is random censoring, thus, $E\left(Y_{k+1}(a^*)\middle\vert\overline{S}_{k+1}=1\right)-E\left(Y_{k+1}(a)\middle\vert\overline{S}_{k+1}=1\right)=E\left(Y_{k+1}(a^*)\right)-E\left(Y_{k+1}(a)\right)$.

effects involving the time-varying confounder, similar to the approach implemented in Le Bourdonnec et al. [14] (see the Discussion section).

4.2 | Positivity

$$P\left(A=a\middle\vert L_0=l_0\right)>0,$$

and

$$P\left(\overline{M}_{k+1}=\overline{m}_{k+1},\overline{S}_{k+1}=\overline{s}_{k+1}\middle\vert A=a,L_0=l_0,\overline{Y}_k=0,\overline{S}_k=1\right)>0,$$
$$\forall k=1,\ldots,K.$$

4.3 | Consistency

If $A=a$ and $S_k=1$, then, $\overline{M}_{k+1}=\overline{M}_{k+1}(a,\overline{s}=1)$, $S_{k+1}=\overline{S}_{k+1}(a,\overline{m})$, $\overline{Y}_{k+1}=\overline{Y}_{k+1}(a,\overline{m},\overline{s}=1)$.

We additionally assume the stable unit treatment value assumption, that is, that there is no interference across units.

Under these identifiability conditions, the expected value of the outcome at time $k+1$, $\mathbb{E}\left[Y_{k+1}\left(a,\overline{M}_k,\overline{S}_k\right)\right]$ is identified by the following function of the observed data:

$$\begin{aligned}\phi(a,a,a)=&\int_{\overline{m}_k}\mathbb{E}\left(Y_{k+1}\middle\vert A=a,\overline{S}_k=1,\overline{M}_k=\overline{m}_k,L_0=l_0\right)\\&f\left(\overline{S}_k=1\middle\vert A=a,\overline{M}_k=\overline{m}_k,L_0=l_0\right)\\&f\left(\overline{M}_k=\overline{m}_k\middle\vert A=a,\overline{S}_{k-1}=1,L_0=l_0\right)d\overline{m}_k\end{aligned}$$

This expression is called the mediational g-formula [15–17].

Our path-specific effects can be computed using the mediational g-formula:

1. Direct effect of the exposure on the outcome:

$$\begin{aligned}&\phi(a^*,a,a)-\phi(a,a,a)\\&=\int_{\overline{m}_k}\mathbb{E}\left(Y_{k+1}\middle\vert A=a^*,\overline{S}_k=1,\overline{M}_k=\overline{m}_k,L_0=l_0\right)\\&\quad f\left(\overline{S}_k=1\middle\vert A=a,\overline{M}_k=\overline{m}_k,L_0=l_0\right)\\&\quad f\left(\overline{M}_k=\overline{m}_k\middle\vert A=a,\overline{S}_{k-1}=1,L_0=l_0\right)d\overline{m}_k\\&-\int_{\overline{m}_k}\mathbb{E}\left(Y_{k+1}\middle\vert A=a,\overline{S}_k=1,\overline{M}_k=\overline{m}_k,L_0=l_0\right)\\&\quad f\left(\overline{S}_k=1\middle\vert A=a,\overline{M}_k=\overline{m}_k,L_0=l_0\right)\\&\quad f\left(\overline{M}_k=\overline{m}_k\middle\vert A=a,\overline{S}_{k-1}=1,L_0=l_0\right)d\overline{m}_k\end{aligned}$$

2. Indirect effect through the history of the mediator of interest:

$$\begin{aligned}&\phi(a^*,a^*,a^*)-\phi(a^*,a^*,a)\\&=\int_{\overline{m}_k}\mathbb{E}\left(Y_{k+1}\middle\vert A=a^*,\overline{S}_k=1,\overline{M}_k=\overline{m}_k,L_0=l_0\right)\\&\quad f\left(\overline{S}_k=1\middle\vert A=a^*,\overline{M}_k=\overline{m}_k,L_0=l_0\right)\\&\quad f\left(\overline{M}_k=\overline{m}_k\middle\vert A=a^*,\overline{S}_{k-1}=1,L_0=l_0\right)d\overline{m}_k\end{aligned}$$

$$
\begin{aligned}
&- \int_{\overline{m}_k} \mathbb{E}\left(Y_{k+1} \middle| A = a^*, \overline{S}_k = 1, \overline{M}_k = \overline{m}_k, L_0 = l_0\right) \\
&\quad f\left(\overline{S}_k = 1 \middle| A = a^*, \overline{M}_k = \overline{m}_k, L_0 = l_0\right) \\
&\quad f\left(\overline{M}_k = \overline{m}_k \middle| A = a, \overline{S}_{k-1} = 1, L_0 = l_0\right) d\overline{m}_k
\end{aligned}
$$

3. Indirect effect through the history of the competing event:

$$
\begin{aligned}
&\phi(a^*, a^*, a) - \phi(a^*, a, a) \\
&= \int_{\overline{m}_k} \mathbb{E}\left(Y_{k+1} \middle| A = a^*, \overline{S}_k = 1, \overline{M}_k = \overline{m}_k, L_0 = l_0\right) \\
&\quad f\left(\overline{S}_k = 1 \middle| A = a^*, \overline{M}_k = \overline{m}_k, L_0 = l_0\right) \\
&\quad f\left(\overline{M}_k = \overline{m}_k \middle| A = a, \overline{S}_{k-1} = 1, L_0 = l_0\right) d\overline{m}_k \\
&\quad - \int_{\overline{m}_k} \mathbb{E}\left(Y_{k+1} \middle| A = a^*, \overline{S}_k = 1, \overline{M}_k = \overline{m}_k, L_0 = l_0\right) \\
&\quad f\left(\overline{S}_k = 1 \middle| A = a, \overline{M}_k = \overline{m}_k, L_0 = l_0\right) \\
&\quad f\left(\overline{M}_k = \overline{m}_k \middle| A = a, \overline{S}_{k-1} = 1, L_0 = l_0\right) d\overline{m}_k
\end{aligned}
$$

4. TE:

$$
\begin{aligned}
&\phi(a^*, a^*, a^*) - \phi(a, a, a) \\
&= \int_{\overline{m}_k} \mathbb{E}\left(Y_{k+1} \middle| A = a^*, \overline{S}_k = 1, \overline{M}_k = \overline{m}_k, L_0 = l_0\right) \\
&\quad f\left(\overline{S}_k = 1 \middle| A = a^*, \overline{M}_k = \overline{m}_k, L_0 = l_0\right) \\
&\quad f\left(\overline{M}_k = \overline{m}_k \middle| A = a^*, \overline{S}_{k-1} = 1, L_0 = l_0\right) d\overline{m}_k \\
&\quad - \int_{\overline{m}_k} \mathbb{E}\left(Y_{k+1} \middle| A = a, \overline{S}_k = 1, \overline{M}_k = \overline{m}_k, L_0 = l_0\right) \\
&\quad f\left(\overline{S}_k = 1 \middle| A = a, \overline{M}_k = \overline{m}_k, L_0 = l_0\right) \\
&\quad f\left(\overline{M}_k = \overline{m}_k \middle| A = a, \overline{S}_{k-1} = 1, L_0 = l_0\right) d\overline{m}_k
\end{aligned}
$$

The proofs of these expressions are presented in the Supporting Information. Please note that assumption 4.1.6 is not needed for the proof of the TE or the indirect effect through the mediator of interest, but it is needed for the proofs of the direct effect and the indirect effect through death.

5 | Mediational g-Formula Algorithm With Time-Varying Mediators and Competing Risks

Based on the DAG presented in Figure 1, our mediational g-formula algorithm proceeds as follows. We choose two exposure values a and a^* (e.g., percentiles 25th and 75th). For each bootstrap sample $1, \ldots, J$:

1. Fit sequential parametric linear and logistic regression models, respectively, for each time point $k = 1, \ldots, K$ for both M and S: $M_k \sim M_{k-1} + A + L_0$, $S_k \sim M_k + A + L_0$.
2. Fit a parametric time-to-event additive hazards model for the outcome Y using a counting process format [18] to adjust for M in all time points: $Y \sim M + A + L_0$.
3. Predict the counterfactual values of M_k and D_k for each time point k: $M_k(a) = M_k\left(a, M_{k-1}(a), S_{k-1}(a)\right)$; $M_k(a^*) = M_k\left(a^*, M_{k-1}(a^*), S_{k-1}(a^*)\right)$; $M_k(a, a^*) = M_k(a, M_{k-1}(a), S_{k-1}(a^*))$, and $S_k(a) = S_k\left(a, M_k(a), S_{k-1}(a)\right)$; $S_k(a^*) = M_k\left(a^*, M_k(a^*), S_{k-1}(a^*)\right)$; $S_k(a^*, a) = S_k(a^*, M_{k-1}(a), S_{k-1}(a^*))$.

   The predictions for the mediator of interest are taken from the regression models. The death status of each participant is predicted by generating random values from a binomial distribution with the predicted probabilities of the logistic regression model. Thus, different individuals will be predicted to die/survive for different counterfactual scenarios.

4. Predict the counterfactual values of the outcome Y:
   - $Y(a) = Y(a, \overline{M}(a, \overline{S}(a)), \overline{S}(a, \overline{M}(a)))$
   - $Y(a^*) = Y\left(a^*, \overline{M}\left(a^*, \overline{S}(a^*)\right), \overline{S}\left(a^*, \overline{M}(a^*)\right)\right)$
   - $Y\left(a^*, \overline{M}(a), \overline{S}(a)\right) = Y\left(a^*, \overline{M}(a, \overline{S}(a)), \overline{S}(a, \overline{M}(a))\right)$
   - $Y\left(a^*, \overline{M}(a, a^*), \overline{S}(a^*, a)\right) = Y\left(a^*, \overline{M}\left(a, \overline{S}(a^*)\right), \overline{S}\left(a^*, \overline{M}(a)\right)\right)$.
5. Calculate the effects of interest:
   - Direct effect: $\mathbb{E}[Y(a^*, M(a), S(a))] - \mathbb{E}[Y(a)]$.
   - Indirect effect through M: $\mathbb{E}[Y(a^*)] - \mathbb{E}[Y(a^*, M(a, a^*), S(a^*, a))]$.
   - Indirect effect through D: $\mathbb{E}[Y(a^*, M(a, a^*), S(a^*, a))] - \mathbb{E}[Y(a^*, M(a), S(a))]$.
   - Total effect: $\mathbb{E}[Y(a^*)] - \mathbb{E}[Y(a)]$.

We use a quantile bootstrap procedure to calculate confidence intervals (CIs). We take the 50th percentile of our bootstrap sample as the effect of interest, and the 2.5th and 97.5th percentiles as the 95% lower and upper CIs. While the theoretical study of these estimators, and hence the theoretical properties of the bootstrap are beyond the scope of this work, the bootstrap has been recommended for inference in effect estimation employing direct counterfactual imputation [6]. We explored the finite sample performance of the bootstrap in Section 6. Our simulation study shows that the bootstrap provides valid inferential results from having appropriate coverage and appropriate type I error for the scenarios considered.

Please note that we make a Markovian assumption and only adjust each model for the mediator in the preceding time point to avoid multicollinearity. These assumptions can be relaxed to better account for the whole history of the mediator in the model, for example, by adjusting the models for a linear combination of past values of the mediator. Our models allow for exposure-mediator interactions and for nonlinear effects, and can accommodate any type of mediator model for the mediator of interest, as well as any type of model suitable for binary outcomes for the death models. While our current version of the algorithm focuses on additive hazards models for the outcome model, our approach could easily be extended to other types of survival models such as Cox proportional hazards or accelerated failure time models. The code of the algorithm is publicly available and can be found in the following Github repository: https://github.com/arcedr/CM_CR_gFormula.

6 | Simulation Study

A simulation study was conducted to evaluate the performance of our proposed estimators across distinct scenarios. The initial scenario, termed "Mimic," was designed to replicate the real data application with three time points, detailed in Section 8 (Figure 1). Another scenario including five time points was also considered. The "No Mediated Effect" (NOME) scenario assessed estimator performance in the absence of a mediated effect, while the "No Competing Risks" (NODE) assessed performance in the absence of competing events. The "g-Formula Not Considering the Pathway Through Death" (NDP) evaluated the estimator sensitivity to model misspecification by applying the g-formula to the initial scenario ("Mimic") without accounting for the death pathway. Descriptions of the different scenarios can be found in Table S1. As a sensitivity analysis, we included an additional scenario that mimics our real data application using the Weibull distribution instead of the exponential distribution to simulate survival data.

To match the sample size of our real application (see Section 8), we simulated 500 datasets, each with a sample size of 3000. The performance of each estimator under each scenario was assessed using bias, root mean squared error (RMSE), CI coverage probability (i.e., percentage of CIs including the true effect), and power (i.e., percentage of simulations in which CIs did not include the null value whenever the true effect was not 0). Further details of the simulation strategy are provided in the Supporting Information.

Table S2 presents the true causal effects for all simulated scenarios. Each effect is displayed both in its original scale and scaled to number of cases per 100 000 person-years for ease of interpretation. For example, a true direct effect of 0.0018 corresponds to 181 cases per 100 000 person-years. For the first scenario, 3424 cases of the event of interest are attributable to the exposure (TE), of which 1589 happen through pathways different to the mediator of interest or death (direct effect), 2103 happen through longitudinal changes on the mediator of interest (indirect effect through the mediator of interest), and 268 are avoided due to the fact that people die before being able to experience the event of interest (indirect effect through death). Thus, this is a scenario with a strong direct effect and a strong indirect effect through the mediator of interest, with moderate competing risks for death. Figure 2 shows the bias, RMSE, power or type I error, and CI coverage for each simulation scenario. The scenarios "Mimic," "NOME," "NODE," and "5timepoint" generally exhibit low bias, low RMSE, and high CI coverage probability for all effects. This suggests that the estimators perform reliably and accurately when closely mimicking real data, as well as in the absence of a mediated effect or competing risks. Figure S5 shows the bias, RMSE, power or type I error, and CI coverage for the "Mimic" scenario comparing our approach to the g-formula without accounting for the death pathway ("NDP"). The "NDP" method, which excludes the pathway through death, shows higher bias and lower CI coverage, indicating reduced estimator performance and reliability when ignoring competing risks. In particular, the fact that the CI coverage is 0 for the indirect effect through death is a consequence of running an algorithm that does not account for competing risks in a setting in which competing risks indeed exist. The power is above 85% under all scenarios for all effects, except for the indirect effect through death. In the scenario with five time points, the power for the indirect effect through death is lower, which suggests that increasing variability (due to a high number of time points) might make it more difficult to detect competing risks. Results were highly consistent when using the Weibull distribution instead of the exponential distribution to generate the survival data (Figure S6). Overall, the results obtained using our algorithm closely align with the true values across various scenarios. These simulations show that ignoring the pathway through competing risks lowers the performance of the estimators.

7 | Application: Urinary Metals, Blood Pressure, and Cardiovascular Disease in the Strong Heart Study

We evaluated the potential mediating role of the systolic blood pressure trajectory (longitudinal mediator) on the association between arsenic and cadmium (exposures, in separate models) and incident CVD (outcome), accounting for competing risks by death. Both arsenic and cadmium have been shown to be associated with elevated blood pressure [19, 20] and with incident CVD [21, 22] in the SHS.

7.1 | Study Population: The Strong Heart Study

A total of 4549 men and women aged 45–74 years were recruited for the SHS from 13 American Indian tribes across three study centers in South Dakota, North Dakota, Oklahoma, and Arizona with baseline visits occurring between 1989 and 1991 [23]. In 2016, one tribal community ($N = 1033$) withdrew consent to use their data for research, leaving 3516 participants. Among the remaining participants, baseline urine samples were analyzed for participants who had sufficient urine for analysis and who were free of diabetes and CVD at baseline. After removing missing values in all covariates, 2925 participants were included in this analysis. Details of sociodemographic characteristics, systolic blood pressure and urinary arsenic and cadmium measurements, as well as CVD ascertainment, can be found in the Supporting Information. Briefly, baseline sociodemographic, lifestyle and anthropometric information were obtained through interview and physical examination. Metals were measured at the baseline visit using inductively coupled plasma mass spectrometry (ICP-MS). Incident CVD was defined as the first occurrence of fatal or nonfatal coronary heart disease, stroke, or congestive heart failure, or other nonfatal CVD. Brachial artery blood pressure was measured at three time points using a mercury sphygmomanometer.

7.2 | Statistical Analysis

We used Multivariate Imputation by Chained Equations (MICE) [24] to impute 605 missing values in systolic blood pressure in the second visit, and 890 missing values in the third visit, based on the values of the other visits. Details of the MICE procedure, as well as results for all MICE iterations, can be found in the Supporting Information and in Table S3. We ran the mediational g-formula algorithm to evaluate the potential mediating role of blood pressure (measured in three visits) on the association between both cadmium and arsenic (in separate models) and incident CVD,

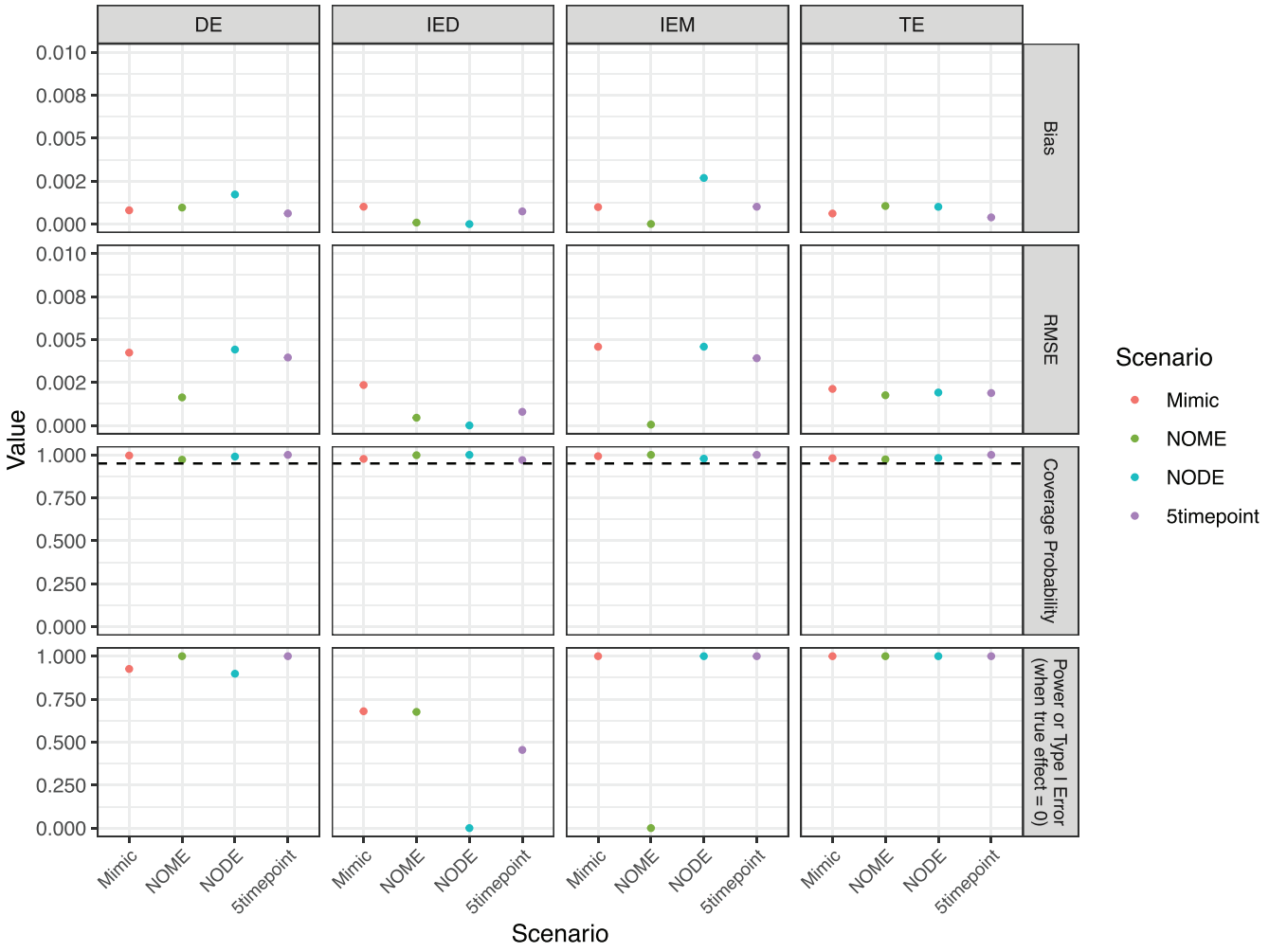

FIGURE 2 | Performance evaluation (bias, root mean squared error, confidence interval coverage, and power) of the different simulation scenarios. Rows represent metrics, and columns represent each causal effect. In each plot, the x-axis represents scenarios, and the y-axis represents the quantity of metrics. The dashed line for the coverage probability is located at 0.95 (representing 95% coverage probability). 5timepoint, five time points; CI, confidence interval; DE, direct effect; IED, indirect effect through death; IEM, indirect effect through the mediator of interest; Mimic, scenario that mimics Strong Heart Study data; NODE, no indirect effect through death (no competing risks); NOME, no indirect effect through the mediator of interest; RMSE, root mean squared error; TE, total effect.

accounting for competing risks by death introducing death as another mediator. To do so, we created a death indicator for each visit (between baseline and the second visit, between the second and the third visit, and between the third visit and the end of follow-up), which we used as the outcome to fit parametric logistic models. From those models, we obtained the probability of dying by that visit using random sampling from the binomial distribution. All models were adjusted for the metal of interest, age, sex, smoking status, BMI, and estimated glomerular filtration rate. Thus, we had two sets of mediator models:

1. Models of the mediator of interest (blood pressure) in each visit. These models were adjusted for blood pressure in the previous visit.
2. Models of the competing risk indicator (death) in each visit. These models were adjusted for blood pressure in the same visit.

To fit the outcome model, we used the counting process format [18], in which each individual contributes several rows to the database (one row per time interval, until the event happens). This format can be used in survival analysis without additional specifications [25]. We used a semi-parametric additive hazards model with CVD incidence as the outcome, adjusted for the metal of interest, blood pressure (in all visits), age, sex, smoking status, BMI, and estimated glomerular filtration rate.

We ran the mediational g-formula algorithm with 1000 bootstrap iterations, comparing the metals' 75th percentile to their 25th percentile. We thus have four effects of interest:

- *Total effect*: number of CVD cases per 100 000 person-years, or probability of not developing CVD after 20 years, attributable to an interquartile range (IQR) change in each metal concentration.
- *Direct effect*: number of CVD cases per 100 000 person-years, or probability of not developing CVD after 20 years, attributable to an IQR increase in each metal concentration not operating through the blood pressure or death pathways.

- *Indirect effect through blood pressure* (the mediator of interest): number of CVD cases per 100 000 person-years, or probability of not developing CVD after 20 years, attributable to an IQR increase in each metal concentration that occur through the blood pressure trajectory either causing CVD or causing death (which would make it impossible for CVD to occur). Please note that this pathway can operate through death only through changes in the mediator of interest, not through direct changes in death resulting from changes in the exposure.
- *Indirect effect through death* (the competing event): number of CVD cases per 100 000 person-years, or probability of not developing CVD after 20 years, attributable to an IQR increase in each metal concentration avoided by the fact that the individual died before the CVD event happening. Please note that this pathway can operate through death only through changes in the exposure, not through changes in the mediator of interest.

For comparative purposes, we are also interested in the following effects:

- *Sum of the direct effect and the indirect effect through blood pressure*: number of CVD cases per 100 000 person-years, or probability of not developing CVD after 20 years, attributable to an IQR increase in each metal concentration that occur through the blood pressure trajectory fixing the survival status to the value it would take under the reference level of the exposure (25th percentile).
- *Direct, indirect, and total effects conditional on survival*: effects that are traditionally calculated in mediation analysis, not accounting for competing risks.

7.3 | Results

Participant characteristics are shown in Table 2. There were 977 CVD cases. Participants who had a CVD event were older, more likely to be current smokers, had higher systolic blood pressure levels, and both higher urinary arsenic and cadmium concentrations. Table 3 and Figure S7 show the results of the longitudinal mediation analysis for arsenic, both without accounting for competing risks (i.e., not considering the pathway through death for other causes different to CVD) and accounting for competing risks. Results are presented both in differences in hazards and survival probability differences. The effects in differences in hazards show that, of 357 CVD cases attributable to an IQR increase in log-arsenic exposure, 49 would happen through the blood pressure trajectory, and eight CVD cases would be avoided by the participant dying from other causes before having the chance to develop CVD per 100 000 person-years. The effects in differences in survival probability show that an IQR increase in log-arsenic exposure would lead to a 4 percentage-point reduction in the probability of being free of CVD after 20 years. From those 4 percentage points, 0.54 would be attributable to longitudinal changes in blood pressure. The indirect effect through blood pressure is statistically significant in the survival probability difference scale, but not in the difference in hazards scale.

Table 4 and Figure S7 show the results for cadmium. The effects in differences in hazards show that, of 152 CVD cases attributable to an IQR increase in log-cadmium exposure, 14 would happen through the blood pressure trajectory, and nine CVD cases would be avoided by the participant dying before having the chance to develop CVD. The effects in differences in survival probability show that an IQR increase in log-cadmium exposure would lead to a 1.76 percentage points reduction in the probability of being free of CVD after 20 years. From those 1.76 percentage points, 0.16 would be attributable to longitudinal changes in blood pressure. The indirect effect through the blood pressure trajectory is only statistically significant in the survival probability scale, both accounting and not accounting for competing risks. The TE is not statistically significant and is attenuated when accounting for competing risks. The indirect effect through death has an impact on the TE, but is not statistically significant for either cadmium or arsenic.

8 | Discussion

In this work, we developed an extension of the mediational g-formula which can deal with competing events. This is, to our knowledge, the first method that incorporates competing events to the mediation analysis setting using the path-specific effects framework. Our algorithm is publicly available on Github and reports effects in terms of attributable cases per a number of person-years, which is a measure of public health impact, in addition to the survival probability difference scale, which has a causal interpretation. The results obtained in our simulation study show optimal performance of the algorithm.

Traditional Cox proportional hazards models censor people that die as if they had dropped out of the study, and therefore consider they could develop the disease at the same rate as those who remain in the study, leading to potential biases of the associations. Other approaches for competing risks in mediation analysis include the survivor average causal effect [26] and the separable effects framework [27]. These methods, however, are focused on a population that is either unrealistic (all survivors) or not observed (those who would survive no matter the exposure level). Also, effects cannot be separable in all settings. For example, in the context of the epidemiological study that motivates this new approach, we have intertwined mediators; thus, we would not be able to consider separate nonintertwined paths for the mediator of interest and the competing event. Instead, the indirect effect through the mediator of interest can also operate through the death pathway, provided that the death trajectory changes because of changes in the mediator of interest, and not because of direct changes in the exposure. Our method decomposes the contribution of each pathway to the association between an exposure and a health outcome while accounting for people that die during follow-up. This provides a better opportunity to investigate different pathways involved in the adverse health effects of elevated urinary metal levels.

Both exposure-mediator interactions and nonlinear effects can be considered in our algorithm. The algorithm can be extended to include time-varying confounders by considering an additional path-specific effect that evaluates the effect of the exposure on the

TABLE 2 | Participant characteristics by cardiovascular disease status.

	Overall ($N = 2932$)	Noncases ($N = 1955$)	Cardiovascular disease cases ($N = 977$)
Age (years)	55.3 (49.3, 62.6)	54.1 (48.6, 61.4)	57.4 (51.2, 64.5)
Sex (%)			
Female	1698 (57.9)	1151 (58.9)	547 (56)
Male	1234 (42.1)	804 (41.1)	430 (44)
Study center (%)			
AZ	397 (13.5)	294 (15)	103 (10.5)
OK	1275 (43.5)	891 (45.6)	384 (39.3)
ND/SD	1260 (43)	770 (39.4)	490 (50.2)
Smoking (%)			
Never	851 (29)	593 (30.3)	258 (26.4)
Former	979 (33.4)	656 (33.6)	323 (33.1)
Current	1102 (37.6)	706 (36.1)	396 (40.5)
BMI (kg/m^2)	29.7 (26.4, 33.8)	29.4 (25.9, 33.5)	30.3 (27.1, 34.2)
eGFR ($mL/min/1.73m^2$)	99.9 (90.6, 107.3)	100.7 (91.6, 107.9)	98.2 (89.2, 105.8)
SBP (visit 1)	124 (113, 136)	122 (111, 134)	128 (118, 141)
SBP (visit 2)	126 (114, 140)	124 (113, 136)	131 (119, 144)
SBP (visit 3)	129 (118, 142)	128 (117, 141)	132 (120, 146)
Arsenic μg/g	8.4 (5.1, 14.3)	8.4 (4.9, 13.9)	8.6 (5.4, 14.8)
Cadmium μg/g	0.97 (0.62, 1.5)	0.95 (0.61, 1.47)	1 (0.64, 1.55)

Note: Median (interquartile range) for continuous variables, percentages for categorical variables.
Abbreviations: AZ, Arizona; eGFR, estimated glomerular filtration rate; ND/SD, North Dakota and South Dakota; OK, Oklahoma; SBP, systolic blood pressure.

TABLE 3 | Direct, indirect, and total effects of one interquartile range increase in urinary arsenic on incident CVD through the systolic blood pressure trajectory and death.

	Without accounting for competing risks		Accounting for competing risks	
	Hazard difference[a]	Difference in survival probability in 20 years (%)	Hazard difference[a]	Difference in survival probability in 20 years (%)
Direct effect	311.08 (−83.25, 733.80)	−3.60 (−6.84, −0.38)	316.85 (−90.12, 748.50)	−3.56 (−6.92, −0.38)
Indirect effect through the blood pressure trajectory	48.85 (−1.53, 109.51)	−0.48 (−0.94, −0.10)	49.64 (−1.35, 115.54)	−0.54 (−1.04, −0.10)
Indirect effect through the death process	—	—	−8.53 (−24.02, 6.09)	0.10 (−0.08, 0.20)
Total effect	362.69 (−34.59, 797.39)	−4.04 (−7.36, −0.80)	357.80 (−51.27, 796.68)	−4.00 (−7.40, −0.76)
Sum of direct effect and indirect effect through blood pressure	359.93 (−84.79, 843.31)	−4.08 (7.78, −0.48)	366.50 (−91.47, 864.05)	−4.10 (−7.96, −0.48)

Note: Models adjusted for age, sex, study center (Arizona, Oklahoma, or North Dakota and South Dakota), estimated glomerular filtration rate, BMI and smoking status (never, former, or current smoking). The procedure was repeated five times for each of the imputed databases generated using Multivariate Imputation by Chained Equations (MICE), and means were taken for effect estimates and confidence intervals.
[a]The hazard difference is reported in number of cases per 100 000 person-years.

TABLE 4 | Direct, indirect, and total effects of one interquartile range increase in urinary cadmium on incident CVD through the systolic blood pressure trajectory and death.

	Without accounting for competing risks		Accounting for competing risks	
	Hazard difference[a]	Difference in survival probability in 20 years (%)	Hazard difference[a]	Difference in survival probability in 20 years (%)
Direct effect	149.4 (−80.7, 374.4)	−1.68 (−1.70, −1.67)	146.2 (−83.2, 374.9)	−1.69 (−1.71, −1.67)
Indirect effect through the blood pressure trajectory	13.1 (−17.7, 46.1)	−0.152 (−0.153, −0.150)	13.8 (−19.8, 49.1)	−0.16 (−0.32, −0.02)
Indirect effect through the death process	—	—	−8.8 (−22.6, 4.7)	0.1 (−0.04, 0.23)
Total effect	161.8 (−72.1, 387.8)	−1.83 (−1.85, −1.81)	152.2 (−80.2, 382.9)	−1.76 (−1.9, −1.63)
Sum of direct effect and indirect effect through blood pressure	162.5 (−98.4, 420.5)	−1.83 (−1.85, −1.82)	160.0 (−103.0, 424.0)	−1.85 (−2.03, −1.69)

Note: Models adjusted for age, sex, study center (Arizona, Oklahoma, or North Dakota and South Dakota), estimated glomerular filtration rate, BMI and smoking status (never, former, or current smoking). The procedure was repeated five times for each of the imputed databases generated using Multivariate Imputation by Chained Equations (MICE), and means were taken for effect estimates and confidence intervals.

[a]The hazard difference is reported in number of cases per 100 000 person-years.

outcome through those time-varying confounders [14], nested with the path through the mediator of interest and through the competing event. In addition, we currently model the trajectory of the mediator of interest by adjusting the mediator model at each time point for the immediate previous time point of the mediator, that is, we do not adjust for all the previous mediator measurements to avoid multicollinearity. This modeling approach could be modified to include more specific measures of the longitudinal trajectory of the mediator such as the rate of change.

Using our mediational g-formula algorithm, we identified a statistically significant indirect effect of the trajectory of blood pressure over time on the association between arsenic and CVD, and cadmium and CVD, in the survival probability difference scale. Our results also show the importance of taking competing risks into account for cadmium, as many people typically die from other causes attributable to cadmium before being able to develop CVD. The wider CIs in both the indirect and total effects show that the statistical uncertainty is greater when accounting for competing risks. The total effect not accounting for competing risks does not consider any variation in the survival status, as it is conditional on survival. Conversely, in our algorithm, we do account for the difference in the survival trajectory under different exposure scenarios, which makes this approach likely to enable modeling of a more realistic scenario. The effects were more statistically significant when using the survival probability difference in 20 years, which has a causal interpretation, as compared to using differences in hazards.

This work has several limitations. First, we consider discrete time points for death indicators, rather than modeling death as time-to-event, which might potentially lead to some measurement errors due to not being able to incorporate the exact time in which the event happened. Also, while we tested our algorithm in a setting (both the real data application and the simulation study) in which competing risks were quite strong (proportion of deaths: 46.3%), our simulation study shows that the power to detect competing risks might be insufficient in settings with a large number of time points. An extension to multistate models [28], in which our primary event of interest (CVD, in this case) and death could be considered as a composite event, represents relevant future work in order to increase power, minimize the loss of granularity and enable the consideration of recurrent events. Another limitation is the parametric modeling approach. It would be of interest to extend this work to the machine learning setting in order to be able to consider several exposures or mediators simultaneously. In addition, some of the assumptions might not always be easy to fulfill in real-world settings. In settings in which positivity or consistency violations might exist, an interesting approach is to consider stochastic interventions, instead of deterministic (or fixed) interventions for the exposure. That way, reasonable interventions within each subject's exposure range can be considered, which leads to a more realistic approach. Moreover, previous work has considered stochastic interventions on mediators when the cross-world assumption is violated. This approach has also been adopted in settings of mediation analysis with competing risks [28]. Alternatively, time-varying confounders can be incorporated into our method as additional mediators for which other path-specific effects are defined to ensure that the cross-world assumption is not violated. We have pursued this approach in previous work [14]. Finally, the assumption of no unmeasured confounding is not verifiable in practice in observational studies [29]. Thus, the results of this study need to be interpreted with caution. The extension of existing sensitivity analysis methods for mediator-outcome unmeasured confounders to the setting of longitudinal mediators and survival outcomes represents relevant future work.

In conclusion, this work shows the importance of considering competing events in mediation analysis with survival outcomes. Identifying the effects through different paths sheds light on the impact of metals on the outcome of interest, as well as through competing risks, more transparently.

Acknowledgments

The authors thank the investigators, the staff, and the participants of the Strong Heart Study (SHS) for their valuable contributions. This paper has been reviewed and approved by the SHS Publications and Presentations Committee. The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Funding

This work was supported by the National Heart, Lung, and Blood Institute (75N92019D00027, 75N92019D00028, 75N92019D00029, 75N92019D00030, R01HL109315, R01HL109301, R01HL109284, R01HL109282, R01HL109319, U01HL41652, U01HL41654, U01HL65520, U01HL65521, U01HL41642), the National Institute of Environmental Health Sciences (P42ES033719, P30ES009089, T32ES007322), and the National Institutes of Health (5R01AG07751802).

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

Strong Heart Study data are shared with researchers following Resource and Data Sharing Policies. Researchers can request data by following the instructions (https://strongheartstudy.org/Research/Study-Data-and-Study-Samples/Study-Data#391231915-data-and-summary-statistics-request-policy, accessed on March 6, 2024).

References

1. P. K. Andersen, R. B. Geskus, T. De witte, and H. Putter, "Competing Risks in Epidemiology: Possibilities and Pitfalls," *International Journal of Epidemiology* 41, no. 3 (2012): 861.

2. L. P. Rojas-Saunero, J. G. Young, V. Didelez, M. A. Ikram, and S. A. Swanson, "Considering Questions Before Methods in Dementia Research With Competing Events and Causal Goals," *American Journal of Epidemiology* 192, no. 8 (2023): 1415.

3. J. A. Driver, "Inverse Association Between Cancer and Neurodegenerative Disease: Review of the Epidemiologic and Biological Evidence," *Biogerontology* 15, no. 6 (2014): 547–557.

4. "Data Analysis With Competing Risks and Intermediate States," Ronald B. Geskus, Google Books, accessed November 14, 2023, https://books.google.com/books?hl=en&lr=&id=sswdCgAAQBAJ&oi=fnd&pg=PP1&ots=brp6Xz6xWr&sig=gHbUPA4fxH6N2rETG3XtQWfJCFs#v=onepage&q&f=false.

5. J. G. Young, M. J. Stensrud, E. J. Tchetgen Tchetgen, and M. A. Hernán, "A Causal Framework for Classical Statistical Estimands in Failure-Time Settings With Competing Events," *Statistics in Medicine* 39, no. 8 (2020): 1199–1236.

6. T. Martinussen and M. J. Stensrud, "Estimation of Separable Direct and Indirect Effects in Continuous Time," *Biometrics* 79, no. 1 (2023): 127–139.

7. Y. T. Huang, "Causal Mediation of Semicompeting Risks," *Biometrics* 77, no. 4 (2021): 1143–1154.

8. A. S. Tai, C. A. Tsai, and S. H. Lin, "Survival Mediation Analysis With the Death-Truncated Mediator: The Completeness of the Survival Mediation Parameter," *Statistics in Medicine* 40, no. 17 (2021): 3953–3974.

9. Y. T. Huanga and H. I. Yangc, "Causal Mediation Analysis of Survival Outcome With Multiple Mediators," *Epidemiology* 28, no. 3 (2017): 370–378.

10. D. Y. Lin and Z. Ying, "Semiparametric Analysis of the Additive Risk Model," *Biometrika* 81, no. 1 (1994): 61.

11. T. Martinussen and S. Vansteelandt, "On Collapsibility and Confounding Bias in Cox and Aalen Regression Models," *Lifetime Data Analysis* 19, no. 3 (2013): 279–296.

12. T. Lange and J. V. Hansen, "Direct and Indirect Effects in a Survival Context," *Epidemiology* 22, no. 4 (2011): 575–581.

13. J. D. Kalbfleisch and R. L. Prentice, "Chapter 7," in *The Statistical Analysis of Failure Time Data*, 2nd ed. (John Wiley & Sons, 2002), 167–175.

14. L. Bourdonnec Kateline, V. Linda, and P.-L. Cécile, "Continuous-Time Mediation Analysis for Repeatedly Measured Mediators and Outcomes," *Biometrics* 81, no. 2 (2025): ujaf062.

15. S.-H. Lin, J. Young, R. Logan, E. J. Tchetgen Tchetgen, and T. VanderWeele, "Parametric Mediational g-Formula Approach to Mediation Analysis With Time-Varying Exposures, Mediators, and Confounders," *Epidemiology* 28, no. 2 (2017): 266–274.

16. A.-S. Tai, S.-H. Lin, Y.-C. Chu, T. Yu, M. A. Puhan, and T. VanderWeele, "Causal Mediation Analysis With Multiple Time-Varying Mediators," *Epidemiology* 34, no. 1 (2023): 8–19.

17. S. H. Lin, J. G. Young, R. Logan, and T. VanderWeele, "Mediation Analysis for a Survival Outcome With Time-Varying Exposures, Mediators, and Confounders," *Statistics in Medicine* 36, no. 26 (2017): 4153–4166.

18. Z. Zhang, J. Reinikainen, K. A. Adeleke, M. E. Pieterse, and C. G. M. Groothuis-Oudshoorn, "Time-Varying Covariates and Coefficients in Cox Regression Models," *Annals of Translational Medicine* 6, no. 7 (2018): 121.

19. N. Franceschini, R. C. Fry, P. Balakrishnan, et al., "Cadmium Body Burden and Increased Blood Pressure in Middle-Aged American Indians: The Strong Heart Study," *Journal of Human Hypertension* 31, no. 3 (2017): 225–230.

20. J. A. Kaufman, C. Mattison, A. M. Fretts, et al., "Arsenic, Blood Pressure, and Hypertension in the Strong Heart Family Study," *Environmental Research* 195 (2021): 110864.

21. K. A. Moon, E. Guallar, J. G. Umans, et al., "Association Between Exposure to Low to Moderate Arsenic Levels and Incident Cardiovascular Disease," *Annals of Internal Medicine* 159, no. 10 (2013): 649.

22. M. Tellez-Plaza, E. Guallar, B. V. Howard, et al., "Cadmium Exposure and Incident Cardiovascular Disease," *Epidemiology* 24, no. 3 (2013): 421–429.

23. E. T. Lee, T. K. Welty, R. Fabsitz, et al., "The Strong Heart Study. A Study of Cardiovascular Disease in American Indians: Design and Methods," *American Journal of Epidemiology* 132, no. 6 (1990): 1141–1155.

24. M. J. Azur, E. A. Stuart, C. Frangakis, and P. J. Leaf, "Multiple Imputation by Chained Equations: What Is It and How Does It Work?," *International Journal of Methods in Psychiatric Research* 20, no. 1 (2011): 40–49.

25. F. Madadizadeh, A. Ghanbarnejad, V. Ghavami, M. Zare Bandamiri, and M. Mohammadianpanah, "Applying Additive Hazards Models for Analyzing Survival in Patients With Colorectal Cancer in Fars Province, Southern Iran," *Asian Pacific Journal of Cancer Prevention* 18, no. 4 (2017): 1077–1083.

26. E. J. Tchetgen Tchetgen, "Identification and Estimation of Survivor Average Causal Effects," *Statistics in Medicine* 33, no. 21 (2014): 3601–3628.

27. M. J. Stensrud, M. A. Hernán, E. J. Tchetgen Tchetgen, J. M. Robins, V. Didelez, and J. G. Young, "A Generalized Theory of Separable Effects in Competing Event Settings," *Lifetime Data Analysis* 27, no. 4 (2021): 588.

28. L. Valeri, C. Proust-Lima, W. Fan, et al., "A Multistate Approach for the Study of Interventions on an Intermediate Time-To-Event in Health Disparities Research," *Statistical Methods in Medical Research* 2023, no. 8 (2023): 1445–1460.

29. M. J. Valente, W. E. Pelham, H. Smyth, and D. P. MacKinnon, "Confounding in Statistical Mediation Analysis: What It Is and How to Address It," *Journal of Counseling Psychology* 64, no. 6 (2017): 659.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Data S1:** sim70425-sup-0001-Supinfo.docx.