Longitudinal mediation analysis of time-to-event endpoints in the presence of competing risks

Tat-Thang Vo[1,2] · Hilary Davies-Kershaw[3] · Ruth Hackett[4] ·
Stijn Vansteelandt[1,5]

Abstract

This proposal is motivated by an analysis of the English Longitudinal Study of Ageing (ELSA), which aims to investigate the role of loneliness in explaining the negative impact of hearing loss on dementia. The methodological challenges that complicate this mediation analysis include the use of a time-to-event endpoint subject to competing risks, as well as the presence of feedback relationships between the mediator and confounders that are both repeatedly measured over time. To account for these challenges, we introduce path-specific effect proportional (cause-specific) hazard models. These extend marginal structural proportional (cause-specific) hazard models to enable effect decomposition on either the cause-specific hazard ratio scale or the cumulative incidence function scale. We show that under certain ignorability assumptions, the path-specific direct and indirect effects indexing this model are identifiable from the observed data. We next propose an inverse probability weighting approach to estimate these effects. On the ELSA data, this approach reveals little evidence that the total effect of hearing loss on dementia is mediated through the feeling of loneliness, with a non-statistically significant indirect effect equal to 1.01 (hazard ratio (HR) scale; 95% confidence interval (CI) 0.99 to 1.05).

✉ Tat-Thang Vo
  tatthang@wharton.upenn.edu

1   Department of Applied Mathematics, Computer Science and Statistics, Ghent University, Ghent, Belgium

2   Present Address: Department of Statistics and Data Science, The Wharton School, University of Pennsylvania, Philadelphia, USA

3   Department of Population Health, London School of Hygiene and Tropical Medicine, London, UK

4   Health Psychology Section, Department of Psychology, King's College London, London, UK

5   Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London, UK

1 Introduction

This article is motivated by an analysis of the English Longitudinal Study of Ageing, a longitudinal cohort study of individuals aged 50 and older living in the community in England, which follows participants biennially since 2002/03 (Steptoe et al. 2013; Davies et al. 2017). In previous studies, it was shown that both self-reported hearing loss and loneliness are significantly associated with a higher risk of dementia (Davies et al. 2017; Davies-Kershaw et al. 2018; Rafnsson et al. 2020). The question of interest, which we shall address here, is whether loneliness mediates the impact of hearing loss on incident physician-diagnosed dementia.

The focus on time-to-event endpoints (dementia) complicates the planned mediation analysis. It renders the popular difference- and product-of-coefficient methods inappropriate (VanderWeele 2016; Robins and Greenland 1992; Pearl 2001), necessitating the use of more complex causal mediation analysis methods. While such methods have been developed for the analysis of time-to-event endpoints, most ignore that the mediator is not assessed at baseline and that subjects may therefore experience the event prior to the mediator being assessed(Lange et al. 2013; Huang and Yang 2017; Vandenberghe et al. 2018). Further complications arise from the mediator, loneliness, being repeatedly measured. While useful to better capture mediation via the entire longitudinal mediator process (Vansteelandt et al. 2019), this also gives rise to complex time-varying confounding patterns whereby mediators (e.g. loneliness) and confounders (e.g. comorbidities) mutually influence each other over time. These complications have been addressed in a number of recent works (Zheng and van der Laan 2017; Lin et al. 2017; Vansteelandt et al. 2019).

A further complication that we must consider, and that we have not previously found being addressed in the mediation analysis literature, is the presence of competing risks by death. With observed - as opposed to counterfactual - event times, the modelling of cause-specific hazards is well known to simplify the handling of competing risks. However, the modelling of (cause-specific) hazards is not readily possible in the previous mediation analysis works (Zheng and van der Laan 2017; Lin et al. 2017; Vansteelandt et al. 2019), which instead focus on the analysis of survival chances. In view of this, in this paper, we will introduce so-called path-specific effect proportional (cause-specific) hazard models, which directly parameterise the direct and indirect effects of a given exposure on the cause-specific hazard of the considered event. We will extend the weighting-based approach proposed by Mittinty and Vansteelandt (2020) for longitudinal natural effect models to enable the planned mediation analysis of an exposure via repeatedly measured mediators on a time-to-event endpoint subject to competing risks.

We proceed as follows. In the next section, we first describe the setting of interest. We then extend the path-specific effect proportional hazard model to take into account the longitudinal nature of the mediator. In the same section, we discuss the assumptions under which the path-specific direct and indirect effects derived from such model

are identifiable from data, and provide a step-by-step procedure for estimating these effects. In sect. 3, we apply the proposed approach to analyze data from the English Longitudinal Study of Ageing (Steptoe et al. 2013; Davies et al. 2017). We end with some final remarks and a discussion.

2 Proposal

2.1 Path-specific models for longitudinal mediators and a time-to-event endpoint subject to competing risks
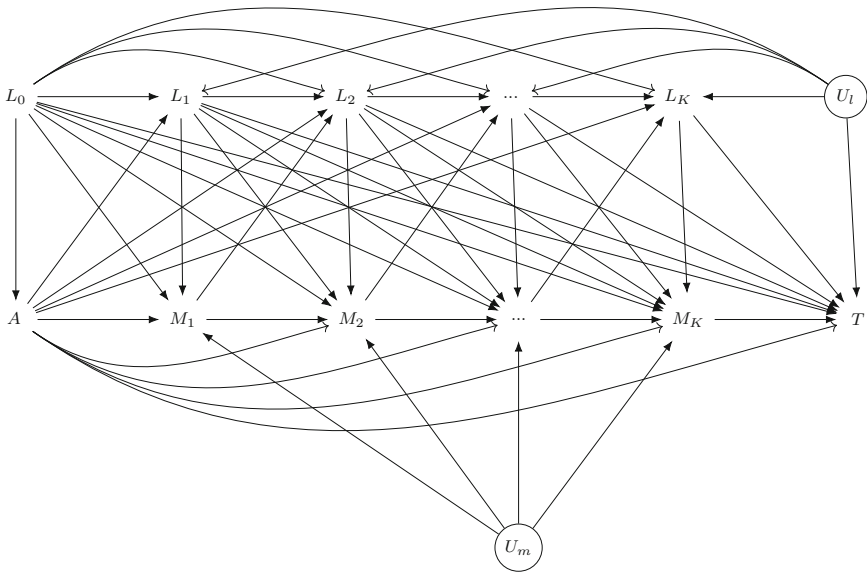
Consider an observational study in which independent individuals $i = 1, \ldots, n$ are exposed to a categorical factor A_i coded as $0, 1, \ldots, P-1$ for P different categories. Longitudinal measurements of the mediator $M_{i0}, M_{i1} \ldots, M_{iK}$ and of the covariates $L_{i0}, L_{i1} \ldots, L_{iK}$ are subsequently recorded at baseline (subscript 0) and at visits $1, \ldots, K$, along with (i) a time-to-event endpoint T_i and (ii) an index D_i specifying whether the main event ($D_i = 1$) or the competing one ($D_i = 2$) happens. Denote $t_{(k)}$, $k = 1 \ldots K$ the fixed time point after the onset of the exposure at which the measurements of M_k and L_k are pre-planned for all patients. Assume that these measurements are only recorded until the last visit K or until event $D_i = 1$ or $D_i = 2$ happens, whichever comes first. The time-to-event endpoint may be censored administratively or due to loss to follow-up, in which case $D_i = 0$.

The causal diagram in Fig. 1 depicts the relationships between the variables over time. In the diagram, L_k includes the indicator $I(T \geq t_{(k)})$ of having survived visit k. Throughout, we will denote the history of measurements up to visit k using a bar, i.e. $\overline{M}_k = (M_1, \ldots M_k)$ and $\overline{L}_k = (L_1, \ldots L_k)$.

To define the direct and indirect effects of interest, we will make use of so-called path-specific effects, expressed as a (cause-specific) hazard ratio. In particular, we define the counterfactual variables T_{a,a^*} and D_{a,a^*} as the time to the main or competing event (whichever comes first) and the corresponding event index that would be observed if the exposure A were set to a and the mediator levels changed to the levels that we would have seen if the exposure were set to a^* and the levels of the time-varying confounders were as observed under this joint intervention on A and $\overline{M}_K$, respectively. The cause-j-specific hazard $\lambda^j_{a,a^*}(t)$ observed under such intervention can then be expressed as:

$$\lambda^j_{a,a^*}(t) = \lim_{\delta t \to 0^+} \frac{P(t \leq T_{a,a^*} \leq t + \delta t, D_{a,a^*} = j \mid T_{a,a^*} \geq t)}{\delta t}$$

The total causal effect (TE) on the cause-j-specific hazard when the exposure changes from a to a^* is thus $HR^j_{TE}(t) = \frac{\lambda^j_{a^*,a^*}(t)}{\lambda^j_{a,a}(t)}$, which can be decomposed into the direct effect (DE) $HR^j_{DE}(t) = \frac{\lambda^j_{a^*,a}(t)}{\lambda^j_{a,a}(t)}$ and the indirect effect (IE) $HR^j_{IE}(t) = \frac{\lambda^j_{a^*,a^*}(t)}{\lambda^j_{a^*,a}(t)}$, where $HR^j_{TE}(t) = HR^j_{IE}(t) \times HR^j_{DE}(t)$. The indirect effect $HR^j_{IE}(t)$ hence reflects

Fig. 1 Causal diagram. U_l: unmeasured confounders affecting L and (T, D). U_m: unmeasured confounders affecting different measurements of M over time. Note that in this figure, $L_k = M_k = 0$ if $T < t_{(k)}$. This is to avoid that future measurements of L and M after the time-to-event T can influence T

the part of the treatment effect (on the cause-j-specific hazard) that is mediated via the pathways $A \to M_k \to \ldots \to T$, where $k = 1, 2, \ldots$. On Fig. 1, these pathways start from the treatment A and go directly to one of the mediators before getting to the event time T by any intermediate path. In contrast, the direct effect $HR_{DE}^j(t)$ reflects the part of the treatment effect (on the cause-j-specific hazard) that does not go through any of the above pathways.

There are many reasons why we choose to focus on the aforementioned pathways. As visualized in Fig. 1, one may nearly always expect the mediators and confounders to influence each other mutually over time. Due to this, many pathways will involve both mediators and confounders. This blurs a good understanding which pathways can be viewed as representing a mechanism via the mediator versus a mechanism via confounders. The decomposition that corresponds best with our intuition of a mechanism via a given mediator, is arguably the one we choose to focus on. The treatment or exposure first affects the mediator, which may in turn generate a cascade of effects, possibly involving confounders and mediators at later time points, to subsequently affect the outcome. Pathways whereby the exposure first influences confounders before in turn affecting mediator and then outcome, correspond better with our intuition of capturing (part of) the mechanism via that confounder. As will be shown below, this combination of effects, whereby the exposure first affects the mediator (at any time) and then in turn outcome by no matter what pathways, is also what can be identified from the observed data under reasonable assumptions.

One further challenge when defining the causal indirect effect via the proposed pathway is the truncation-by-death problem. This arises as a result of the difference in

a person's time-to-death with versus without the considered exposure. Indeed, even if a given person would survive a given time point t in the absence of exposure ($A = 0$), that person's counterfactual mediator value in the presence of exposure ($A = 1$) is ill-defined at time t, if that person would have died by that time in the presence of the exposure. In that case, it is wise to interpret the identified effects differently as expressing the effect when setting the mediator at each time point at a random draw from its distribution in the exposed who are still alive and have the same considered observed data history. A more detailed discussion on this can be found elsewhere (Vansteelandt et al. 2019).

Finally, note that the total, direct and indirect effects on the (cause-specific) hazard ratio scale have a subtle interpretation. In the causal inference literature, the hazard ratio is often criticized due to the fact that it may be affected by selection effects caused by unobserved heterogeneity over time, and hence does not generally have a causal interpretation (Young et al. 2020; Martinussen et al. 2020). As an example, consider the cause-j specific hazard ratio HR_{TE}^j defined above as:

$$HR_{TE}^j = \frac{\lambda_{a^*,a^*}^j(t)}{\lambda_{a,a}^j(t)} = \frac{\lim_{\delta t \to 0^+} P(t \le T_{a^*,a^*} \le t + \delta t, D_{a^*,a^*} = j | T_{a^*,a^*} \ge t)/\delta t}{\lim_{\delta t \to 0^+} P(t \le T_{a,a} \le t + \delta t, D_{a,a} = j | T_{a,a} \ge t)/\delta t}$$

The righthand expression makes clear that HR_{TE}^j contrasts the cause-j specific hazard functions with and without the exposure for two separate groups of individuals, namely (i) those who survive time t if exposed (i.e. $T_{a^*,a^*} \ge t$) and (ii) those survive time if unexposed t. (i.e. $T_{a,a} \ge t$). These two groups are in general not comparable if the exposure has a non-null causal effect on (T, D) (except at baseline if the exposure is 'randomized'). This selection effect likewise impacts the two causal contrasts HR_{DE}^j and HR_{IE}^j.

To overcome the above limitation, the total, direct and indirect effect could be alternatively defined on the cause-specific cumulative incidence scale. For instance, under the joint intervention on A and $\overline{M}_K$ described above, one then has:

$$F_{a,a^*}^j(t) := P(T_{a,a^*} \le t, D_{a,a^*} = j) = \int_0^t \exp\left(-\int_0^u \sum_j \lambda_{a,a^*}^j(s)ds\right)\lambda_{a,a^*}^j(u)du, \tag{1}$$

which expresses the incidence of the occurrence of event j at time t, while taking other competing risks into account. Contrasts of $F_{a^*,a^*}^j(t)$ and $F_{a,a}^j(t)$ (e.g. on the relative scale) hence quantify the total effect of the exposure on the event j at time t. Similarly, contrasts of $F_{a^*,a^*}^j(t)$ and $F_{a^*,a}^j(t)$ (or $F_{a^*,a}^j(t)$ and $F_{a,a}^j(t)$) quantify the indirect effect of A on the event j via the pathways $A \to M_k \to \ldots \to T$ (or the direct effect of A on the event j not via these pathways) at time t. Unlike hazard ratios, these contrasts are marginally defined and hence are causally interpretable.

Even so, exposure effects on the cumulative incidence functions also have a subtle interpretation because of the exposure effect on competing causes. For instance, the exposure may seem to be protective of the event of interest merely because it increases

the likelihood of the competing event (Putter et al. 2007). Young et al. (2020) therefore evaluate effects under hypothetical interventions which eliminate the exposure effect on the competing events. We here avoid this as it demands additional assumptions, which we judged to be implausible in our study.

We are now ready to define the cause—j—specific path-specific effect proportional hazard model, accounting for longitudinal mediators, as follows:

$$\lambda_{a,a^*}^{j}(t) = \lambda_0^{j}(t)e^{\alpha_{1j}a+\alpha_{2j}a^*} \tag{2}$$

for all a, a^*, where $\lambda_0^{j}(t)$, α_{1j} and α_{2j} are unknown. This model is inspired by the so-called *natural effect models* proposed by Steen et al. (2017b), and subsequently extended by Mittinty and Vansteelandt (2020) to a longitudinal outcome setting. The difference, however, is that model (2) targets path-specific effects rather than natural indirect effects as in Steen et al. (2017b). Also, note that when $a = a^*$, the proposed model (2) reduces to the standard *marginal structural model* proposed by Robins et al. (2000). It can thus be viewed as a generalization of *marginal structural models* to investigate path-specific effects.

Under model (2), the total, direct and indirect effect can be expressed on the hazard ratio scale as $HR_{TE}^{j} = e^{(\alpha_{1j}+\alpha_{2j})(a^*-a)}$; $HR_{IE}^{j} = e^{\alpha_{2j}(a^*-a)}$ and $HR_{DE}^{j} = e^{\alpha_{1j}(a^*-a)}$, respectively. To assess the possibility of mediator-exposure interaction, one can alternatively consider model:

$$\lambda_{a,a^*}^{j}(t) = \lambda_0^{j}(t)e^{\alpha_{1j}a+\alpha_{2j}a^*+\alpha_{3j}a.a^*} \tag{3}$$

Under model (3), the total causal effect on the hazard ratio scale is expressed as $HR_{TE}^{j} = e^{(\alpha_{1j}+\alpha_{2j}+\alpha_{3j}(a^*+a))(a^*-a)}$ and is decomposed into the indirect effect $HR_{IE}^{j} = e^{(\alpha_{2j}+\alpha_{3j}a)(a^*-a)}$ and the direct effect $HR_{DE}^{j} = e^{(\alpha_{1j}+\alpha_{3j}a^*)(a^*-a)}$. The cumulative incidence function $F_{a,a^*}^{j}(t)$ at any time t can also be computed from the estimates of model (2) and (3) (see below). Finally, note that other models for survival outcomes, such as the Aalen model, can also be extended to the current context.

2.2 Identification

In this section, we discuss the assumptions that allow one to identify the distribution of $Y_{a,a^*} = (T_{a,a^*}, D_{a,a^*})$ from the observed data. For now, we will ignore the problem of censoring and will specifically focus on identifying the cause-j-specific cumulative distribution function $F_{a,a^*}^{j}(t)$ at a time t between t_3 and t_4, i.e. the last measurements of M and L are M_3 and L_3, respectively. All results given below generalize to any time t with measurements for more than three study visits.

First, we make use of the following recanting witness (or *cross-world*) assumption:

(i)

$$\left\{ Y(a, \bar{l}_3, \overline{m}_3), L_1(a), L_2(a, l_1, m_1), L_3(a, \bar{l}_2, \overline{m}_2) \right\} \perp\!\!\!\perp$$

$$\left\{ M_1(a^*, l_1), M_2(a^*, \bar{l}_2, m_1), M_3(a^*, \bar{l}_3, \overline{m}_2) \right\} \Big| L_0$$

This assumption is satisfied in the causal graph depicted in Fig. 1, provided that it represents a non-parametric structural equation model with independent errors. Under this assumption, $F^j_{a,a^*}(t)$ for $t_3 < t \le t_4$ can be written as:

$$F^j_{a,a^*}(t) = \int \Pr\{T(a, \overline{m}_3) \le t, D(a, \overline{m}_3) = j,$$

$$L_1(a) = l_1, L_2(a, l_1, m_1) = l_2, L_3(a, \bar{l}_2, \overline{m}_2) = l_3 | L_0 = l_0\}$$

$$\times f\{M_1(a^*, l_1) = m_1, M_2(a^*, \bar{l}_2, m_1) = m_2, M_3(a^*, \bar{l}_3,) = m_3, \overline{m}_2 | L_0 = l_0\}$$

$$\times f(L_0 = l_0) \, d\bar{l}_3 \, d\overline{m}_3 \tag{4}$$

where $f\{\cdot\}$ denotes the (joint) density. To link $F^j_{a,a^*}(t)$ to the observed data, we will further assume that the set of baseline covariates L_0 is sufficient to control for confounding of the relationship between A and $Y = (T, D)$, as well as between A and M_k or L_k at any time $t_{(k)}$.

(ii) $A \perp\!\!\!\perp L_k(a, \overline{m}_{k-1})|L_0$ and $A \perp\!\!\!\perp M_k(a, \bar{l}_k)|L_0$ for all $k = 1, 2, 3$

(iii) $A \perp\!\!\!\perp Y(a, \bar{l}_3, \overline{m}_3) \mid L_0$

Besides, we assume that conditional on the observed history, there are no unmeasured confounders of the relationship between the time-to-event outcome and the mediator, as well as between the mediator and the longitudinal confounder at any time. More precisely,

(iv) $M_k \perp\!\!\!\perp \{Y(a, \bar{l}_3, \overline{m}_3), L_{k+1}(a, \bar{l}_k, \overline{m}_k), \ldots, L_3(a, \bar{l}_k, \overline{m}_k)\}|A = a, \overline{L}_k = \bar{l}_k, \overline{M}_{k-1}$ for $k = 1, 2$ and $M_3 \perp\!\!\!\perp Y(a, \bar{l}_3, \overline{m}_3)|A = a, \overline{L}_3 = \bar{l}_3, \overline{M}_2 = \overline{m}_2, L_0$

(v) $L_k \perp\!\!\!\perp \{M_k(a, \bar{l}_k, \overline{m}_{k-1}), \ldots, M_3(a, \bar{l}_3, \overline{m}_2)\}|\overline{M}_{k-1} = \overline{m}_{k-1}, \overline{L}_{k-1} = \bar{l}_{k-1}, A = a, L_0$ for all $k = 1, \ldots, 3$

For instance, in Fig. 1, assumption (iv) is satisfied since conditioning on the exposure A and the history of the time-varying confounder L up to time t is sufficient to adjust for confounding of the relationship between the time-to-event outcome T and the mediator level at time t. Our development allows for the presence of unmeasured common causes of the mediators over time (i.e. denoted U_m in Fig. 1) and separate, independent unmeasured common causes of the baseline/time-varying confounders and the survival time (T, D) (i.e. denoted U_l).

Finally, we make use of the standard consistency assumption, i.e.:

(vi) $\Pr(L_1(a) = L_1|a) = \Pr(M_1(a, l_1) = M_1|a, l_1) = \Pr(L_2(a, l_1, m_1) = L_2|a, l_1, m_1) = \ldots = 1$

In Appendix 1, we show that under the above assumptions, $F^j_{a,a^*}(t)$ can be linked to the observed data as:

$$F^j_{a,a^*}(t) = \int \Pr\{T \le t, D = j|a, \bar{l}_3, \overline{m}_3, l_0\} \cdot \prod_{s:0<s\le 3} f(m_s|a^*, \overline{m}_{s-1}, \bar{l}_s, l_0, T \ge t_{(s)})\cdot$$

$$\cdot f(l_s|\bar{l}_{s-1}, \overline{m}_{s-1}, l_0, a, T \ge t_{(s)}) \cdot f(l_0) \, d\overline{m}_3 \, d\bar{l}_3 \, dl_0$$

2.3 Estimation

In what follows, we generalize the standard estimation procedure for the popular marginal structural models to fit models (2) and (3). For this, we make an additional assumption that censoring is non-informative, in the sense that $C \perp\!\!\!\perp (T, \overline{M}_{\lfloor t \rfloor}, \overline{L}_{\lfloor t \rfloor}, L_0)|A$ where C denote the time-to-censoring. At any timepoint, the instantaneous risk of the event among patients who then drop out of the study is hence not different (at all future times) from that of patients who remain, conditional on the exposure level. In the next section, we will relax this assumption to allow for more complicated censoring mechanisms. Under the aforementioned assumptions, Appendix 2 shows that consistent estimators of the parameters indexing model (2) can be obtained by solving the estimating equation:

$$
\int_0^\infty \sum_{i,a,a^*} \left\{ \binom{a}{a^*} - \frac{\sum_{a,a^*} \hat{E}\left[\binom{a}{a^*} R_i^j(t)I(C_i \geq t)W_i(\lfloor t \rfloor, a, a^*)e^{\alpha_1 ja + \alpha_2 ja^*} \right]}{\sum_{a,a^*} \hat{E}\left[R_i^j(t)I(C_i \geq t)W_i(\lfloor t \rfloor, a, a^*)e^{\alpha_1 ja + \alpha_2 ja^*} \right]} \right\} \cdot
$$

(5)

$$
\cdot R_i^j(t)I(C_i \geq t)W_i(\lfloor t \rfloor, a, a^*)\left(dN_i^j(t) - \lambda_0^j(t)e^{\alpha_1 ja + \alpha_2 ja^*} \right) = 0
$$

where $R_i^j(t) = I(T_i \geq t, D_i = j)$; $dN_i^j(t) = I(T_i = t, D_i = j)$ and $I(.)$ denotes the indicator function. Besides,

$$
W_i(\lfloor t \rfloor, a, a^*) = \frac{\prod_{s\in\mathbb{N}, t_s \leq \lfloor t \rfloor} \hat{\Pr}(M_{s,i}|A_i = a^*, \overline{M}_{s-1,i}, \overline{L}_{s,i}, L_{0,i}, T_i \geq t_{(s)})}{\prod_{s\in\mathbb{N}, t_s \leq \lfloor t \rfloor} \hat{\Pr}(M_{s,i}|A_i = a, \overline{M}_{s-1,i}, \overline{L}_{s,i}, L_{0,i}, T_i \geq t_{(s)})}
$$
$$
\times \frac{I(A_i = a)}{\hat{\Pr}(A_i = a|L_{0,i})}
$$

denotes the weight of individual i at time t and at exposure levels a and a^*, $\hat{\Pr}(\cdot)$ denotes a consistent estimator of the corresponding probability $\Pr(\cdot)$, and $\hat{E}(\cdot)$ denotes the sample average. Here, the notation $\lfloor t \rfloor$ is slightly different from its standard definition, to take into account the fact that if a patient experiences an event or leaves the study at time $t = t_{(k)}$ of visit k, no measurement of M_k and L_k is possible at that time. More precisely,

$$
\lfloor t \rfloor = \begin{cases} t_{(k-1)} & \text{if } t = t_{(k)} \\ t_{(k)} & \text{if } t_{(k)} < t < t_{(k+1)} \end{cases}
$$

where $k = 1, \ldots, K$. The second component of the weight ensures that the exposure-outcome association is adjusted for confounding by L_0. It creates a pseudo-population in which the exposure is no longer associated with L_0 and hence removes confounding by L_0 (Lange et al. 2012). The first component of the weight then distinguishes between the direct and indirect paths by correcting for the fact that the observed mediator value

at each time point before t may differ from the counterfactual value that is of interest at that time. From this, the fitting procedure is described as follows:

Step 1 Postulate and fit a suitable model for the exposure A conditional on the baseline confounders (L_0) based on the original data set. For instance, a multinomial logistic model can be used for a categorical exposure with P possible values:

$$\log \frac{\Pr(A = a | L_0 = l_0)}{\Pr(A = 0 | L_0 = l_0)} = \beta_{0,a} + \beta_{1,a} \, l_0,$$

where $a = 1, \ldots, P - 1$ and $\Pr(A = 0 | L_0 = l_0) = 1/(1 + \sum_{a=1}^{P-1} e^{\beta_{0,a} + \beta_{1,a} \, l_0})$.

Step 2 Convert the original dataset to a long or counting-process format, in which the observation period $[0, T_i]$ of individual i (where T_i is the observation time of this individual) is broken into $k' + 1$ intervals where $k' = \text{argmax}_k \{t_{(k)} \leq \lfloor T \rfloor\}$. Each interval k ($1 \leq k \leq k'$) will have the following information encoded:

(a) The beginning of the interval, which equals the time $t_{(k-1)}$ of visit $k - 1$, with $t_{(0)} = 0$.
(b) The end of the interval, which equals the time $t_{(k)}$ of visit k or the event/censoring time T_i for the last interval.
(c) The event status at the end of the interval.
(d) The exposure A_i and the baseline covariates L_{0i}, whose values remain unchanged across all intervals.
(e) The history of the mediator and the longitudinal confounders recorded up to the end of the interval. Note that the history of L and M for subject i up to the time T_i is similar to their history up to the last visit prior to time T_i.

Table 1 provides a toy example in which three patients receive a binary treatment and are followed up for a total of three years, with two visits pre-planned at the end of year 1 and 2. Patient 3 is free of event at the end of the study and hence has the mediator level fully recorded at the two intermediate visits. In contrast, patient 1 and 2 experience an event after 1.5 and 0.9 years, due to which they have no (i.e. patient 2) or only one (i.e. patient 1) mediator level recorded. Table 2 illustrates how the information of these three hypothetical individuals is encoded in a counting-process format.

Step 3 Postulate and fit a suitable model for the mediator at each time $t_{(k)}$, conditional on the exposure, the longitudinal confounder $\overline{L}_k$ and previous measurements of the mediator (i.e. $\overline{M}_{k-1}$), by using the long data set. For instance, one may assume multinomial logistic models for a categorical mediator M_k with possible values $0, \ldots, Q$,

Table 1 Illustrating example: A toy dataset in standard short format

Individual	Follow-up time (years)	Status	A	M_1	M_2	L_0	L_1	L_2
1	1.5	1	1	m_{11}	–	l_{01}	l_{11}	–
2	0.9	2	0	–	–	l_{02}	–	–
3	3.0	0	1	m_{13}	m_{23}	l_{03}	l_{13}	l_{23}

Here, m_{ij} and l_{ij} denote the mediator and longitudinal confounder level of individual j recorded at visit $t_{(i)}$, respectively. The treatment A is binary (0 vs. 1) and there are two competing events, coded as 1 and 2

Table 2 The counting-process format of the dataset in Table 1

Individual	Start	Stop	Status	A	M_k	M_{k-1}	L_k	L_{k-1}	L_0
1	0	1	0	1	0	0	0	0	l_{01}
1	1	1.5	1	1	m_{11}	0	l_{11}	0	l_{01}
2	0	0.9	2	0	0	0	0	0	l_{02}
3	0	1	0	1	0	0	0	0	l_{03}
3	1	2	0	1	m_{13}	0	l_{13}	0	l_{03}
3	2	3	0	1	m_{23}	m_{13}	l_{23}	l_{13}	l_{03}

that is:

$$\log \frac{\Pr(M_k = q | A = a, \overline{M}_{k-1} = \overline{m}_{k-1}, \overline{L}_k = \overline{l}_k, L_0 = l_0, T \geq t_{(k)})}{\Pr(M_k = 0 | A = a, \overline{M}_{k-1} = \overline{m}_{k-1}, \overline{L}_k = \overline{l}_k, L_0 = l_0, T \geq t_{(k)})}$$
$$= \gamma_{0q} + \gamma_{1q} t_{(k)} + \gamma_{2q} a + \gamma'_{3q} \overline{m}_{k-1} + \gamma'_{4q} \overline{l}_k + \gamma_{5q} l_0$$

where $q = 1, \ldots, Q$. Note that other non-linear terms such as two-by-two interactions between baseline covariates and previous measurements of M before k can also be added to the above model.

Step 4 A new data set is then constructed by copying the original data set (in long format) P times and including an additional variable A^* to capture the P possible values of the exposure relative to the indirect path. A^* is set to the actual value of the exposure A for the first replication, to the other potential values of A for the remaining replications. This step ensures that the estimating equation solved by standard software is summing over levels of a^* as in equation (5). Such a step is also standard in the estimation procedure of natural effect models and extensions thereof, which was first proposed by Lange et al. (2012) in the simple setting of single mediator. For the example discussed in Tables 1 and 2, the corresponding extended data set is provided in Table 3.

Step 5 Compute weights by applying the fitted models from steps 1 and 3 to the new data set. At visit k, the weight for the ith individual is $w_i(k, a, a^*) = w_i^{ttm}(a) \cdot w_i^{med}(k, a, a^*)$, where:

$$w_i^{ttm}(a) = \sum_{a=0}^{P-1} \frac{I(A_i = a)}{\hat{\Pr}(A_i = a | L_{0i} = l_{0i})}$$

and

$$w_i^{med}(k, a, a^*)$$
$$= \prod_{s \in \mathbb{N}, t_{(s)} \leq t_{(k)}} \frac{\hat{\Pr}(M_{s,i} = m_{s,i} | A_i = a^*, \overline{M}_{s-1,i} = \overline{m}_{s-1,i}, \overline{L}_{s,i} = \overline{l}_{s,i}, L_{0,i} = l_{0,i}, T_i \geq t_s)}{\hat{\Pr}(M_{s,i} = m_{s,i} | A_i = a, \overline{M}_{s-1,i} = \overline{m}_{s-1,i}, \overline{L}_{s,i} = \overline{l}_{s,i}, L_{0,i} = l_{0,i}, T_i \geq t_s)}$$

Table 3 Extended data set for the example in Table 1

Individual	Start	Stop	Status	A	A^*	M_k	M_{k-1}	L_k	L_{k-1}	L_0
1	0	1	0	1	1	0	0	0	0	l_{01}
1	1	1.5	1	1	1	m_{11}	0	l_{11}	0	l_{01}
2	0	0.9	2	0	0	0	0	0	0	l_{02}
3	0	1	0	1	1	0	0	0	0	l_{03}
3	1	2	0	1	1	m_{13}	0	l_{13}	0	l_{03}
3	2	3	0	1	1	m_{23}	m_{13}	l_{23}	l_{13}	l_{03}
1	0	1	0	1	0	0	0	0	0	l_{01}
1	1	1.5	1	1	0	m_{11}	0	l_{11}	0	l_{01}
2	0	0.9	2	0	1	0	0	0	0	l_{02}
3	0	1	0	1	0	0	0	0	0	l_{03}
3	1	2	0	1	0	m_{13}	0	l_{13}	0	l_{03}
3	2	3	0	1	0	m_{23}	m_{13}	l_{23}	l_{13}	l_{03}

where the superscript ttm and med denotes treatment and mediator, respectively. At the end of the follow-up time, the weight for a patient having $T_i = t_i$ is $w_i(t_i, a, a^*) = w_i(\lfloor t_i \rfloor, a, a^*)$.

Step 6 Fit the path-specific effect cause-specific proportional hazard model (2) and (3) by proportional hazard regression of the cause-specific event time on A and A^* on the basis of the expanded data set, using the weights computed in the previous step. Derive confidence intervals for the parameters in model (2) and (3) using the non-parametric bootstrap. For this, one first generates S bootstrap samples with replacement from the original dataset, then repeat all the above steps for each bootstrap sample. The 95% confidence interval for each parameter in model (2) and (3) is computed by using the 2.5% and 97.5% quantiles of the bootstrap distribution of the corresponding estimator.

Step 7 Establish the cause-specific cumulative incidence curves under different sets of a and a^*. To estimate the curves, consider the different event types as terminal states of a multi-state model where the transition from one state to the other is treated as an absorbing state, i.e. the one that subjects never exist (Putter et al. 2007).

2.4 Addressing complications due to censoring

As stated above, when the censoring is non-informative conditional on the exposure (Fig. 2a), the provided estimating procedure remains valid without further adjustment. When censoring is dependent upon the baseline covariate vector L_0 and the exposure A (i.e. $C \perp\!\!\!\perp (T, \overline{M}_{\lfloor t \rfloor}, \overline{L}_{\lfloor t \rfloor})|L_0, A$), one may adjust for censoring by alternatively focusing on the so-called conditional cause-specific path-specifci effect proportional hazard model, that is,

$$\lambda_{a,a^*}^j(t|L_0 = l_0) = \lambda_0^j(t)e^{\alpha_{1j}a + \alpha_{2j}a^* + \alpha'_{2j}l_0}$$

(a)

(b)
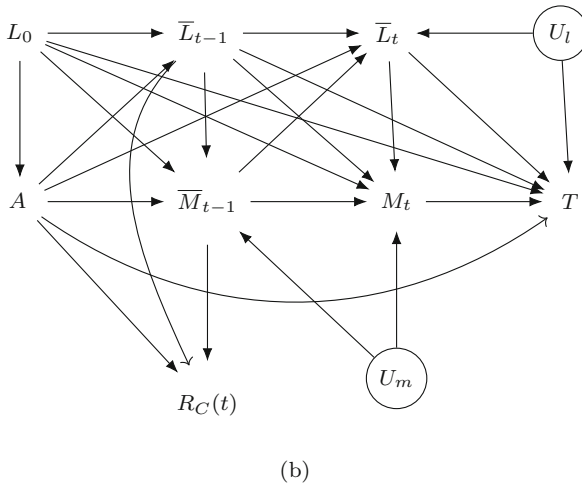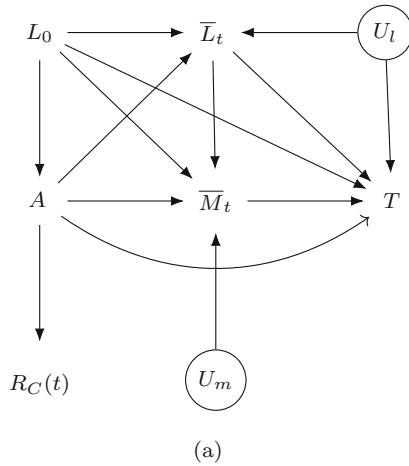
Fig. 2 (Simplified) causal diagram when censoring presents—**a** Censoring is non-informative conditional on the exposure and **b** Censoring is non-informative at time t conditional on the exposure and the history up to that time. $R_C(t) = I(C \geq t)$, where C is the time-to-censoring

Note that an interaction between a^* and l_0 could also be permitted in such model to assess the possibility of mediator-baseline covariate interaction. The procedure discussed in Sect. 2.2 can then be applied to estimate the parameters in this model, with a slight adjustment in step 6 where apart from A and A^*, the covariates L_0 (and the product of A^* and L_0 if mediator- baseline covariate interaction is assessed) are also included into the proportional hazard regression model. In practice, censoring may however also depend upon post-baseline factors such as the longitudinal mediator and confounder levels that are measured prior to censoring (Fig. 2b). In that case, progress can be made upon assuming that at any time t, the risk of dropping out of the study for patients not yet experiencing an event does not depend on when they will experience

an event in the future and of what type this event will be, given the history up to time t, i.e. $\lambda_C(t|T > t, T, D, A, \overline{M}_t, \overline{L}_t) = \lambda_C(t|T > t, A, \overline{M}_t, \overline{L}_t)$ where $\lambda_C(\cdot)$ denoting the cause-specific hazard function of the time-to-censoring. In that case, the so-called inverse probability of censoring weighting approach can be used to account for censoring. More precisely, the parameters indexing models (2) can be estimated by solving the estimating Eq. (5) with the following weight for each individual i (see appendix A2 for a formal proof):

$$W_i(\lfloor t \rfloor, a, a^*) = \frac{\prod_{s:t_{(s)} \leq \lfloor t \rfloor} \hat{\Pr}(M_{s,i}|A_i = a^*, \overline{M}_{s-1,i}, \overline{L}_{s,i}, L_{0,i}, T_i \geq t_{(s)}, C_i \geq t_{(s)})}{\prod_{s:t_{(s)} \leq \lfloor t \rfloor} \hat{\Pr}(M_{s,i}|A_i = a, \overline{M}_{s-1,i}, \overline{L}_{s,i}, L_{0,i}, T_i \geq t_{(s)}, C_i \geq t_{(s)})}$$

$$\cdot \frac{I(A_i = a)}{\hat{\Pr}(A_i = a|L_{0,i})} \cdot$$

$$\cdot \frac{1}{\prod_{0 < s \leq t} \left[1 - \hat{\lambda}_C(s|T_i > s, \overline{M}_{\lfloor s \rfloor,i}, \overline{L}_{\lfloor s \rfloor,i}, L_{0,i}, A_i = a)ds\right]}$$

Here, $\prod_s x_s$ is defined as a product limit and $\hat{\lambda}_C(\cdot)$ is a consistent estimator of $\lambda_C(\cdot)$. With the above weight $W_i(\lfloor t \rfloor, a, a^*)$, the additional component that accounts for the informative censoring can make the overall weight become unstable (e.g. when the censoring hazard is close to 1 in some strata). To overcome this, one can then use stabilized (censoring) weights which incorporate a numerator defined in the same way as the denominator but adjusting only for the exposure, that is:

$$\frac{\prod_{s:0 \leq s \leq t} \left[1 - \hat{\lambda}_C(s|A_i = a)ds\right]}{\prod_{s:0 \leq s \leq t} \left[1 - \hat{\lambda}_C(s|T_i > s, \overline{M}_{\lfloor s \rfloor,i}, \overline{L}_{\lfloor s \rfloor,i}, L_{0,i}, A_i = a)ds\right]}$$

One then needs to postulate two models for the censoring hazard at time t, with one conditioning on exposure and the other conditioning on exposure, baseline covariates and the history of the longitudinal mediator and confounders up to time t, where only the latter model needs to be correct. For instance,

$$\lambda_C(t|a_i) = \lambda'_{0C}(t)e^{\eta a_i}$$
$$\lambda_C(t|T_i > t, \overline{m}_{\lfloor t \rfloor,i}, \overline{l}_{\lfloor t \rfloor,i}, l_{0,i}, a_i) = \lambda_{0C}(t)e^{\theta_0 a_i + \theta_1 l_{0,i} + \theta'_2 \overline{m}_{\lfloor t \rfloor,i} + \theta'_3 \overline{l}_{\lfloor t \rfloor,i}} \tag{6}$$

Note that non-proportional hazards can be addressed in the standard way, e.g. the time-varying effect can be modeled parametrically or with a step function.

In step 5 of the estimation procedure, apart from computing the mediator and treatment weights, one needs to additionally derive the censoring weight. More precisely, the weight for the ith individual at visit k is now $w_i^{ttm}(a, a^*) \cdot w_i^{med}(k, a) \cdot w_i^{cen}(k, a)$,

where the superscript *cen* denotes censoring, i.e.,

$$w_i^{cen}(k, a) = \frac{\prod_{s:0 \leq s \leq t}\left[1 - \hat{\lambda}_C(s|A_i = a)ds\right]}{\prod_{s:0 \leq s \leq t}\left[1 - \hat{\lambda}_C(s|T_i > s, \overline{M}_{\lfloor s \rfloor,i}, \overline{L}_{\lfloor s \rfloor,i}, L_{0,i}, A_i = a)ds\right]}$$

while $w_i^{ttm}(a, a^*)$ and $w_i^{med}(k, a, a^*)$ are computed as above. Of note, some R packages such as ipcwswitch provide useful built-in functions to calculate the censoring weights (Graffeo et al. 2019). We thus refer the reader to these references for more practical guidance.

Once the individual weights are computed, the natural effect cause-specific proportional hazard model (2) and (3) can be fitted using these weights and the confidence intervals for the estimates can be derived via the nonparametric bootstrap, as described above.

3 Illustrating example

We illustrate the proposed approach on the ELSA dataset (https://www.elsa-project.ac.uk), which includes 4232 patients over 50 years of age. In this ongoing study, the first contact with the participants was in 2002/03 (wave 1). These participants were then followed up biennially, with measures collected via computer-assisted face-to-face interview and self-completion questionnaires. As stated above, the question of interest here is whether the feeling of loneliness mediates the impact of hearing loss on time-to-dementia (measured in years), accounting for mortality as a competing event.

For this analysis, we used the hearing measurement recorded at wave 2 (e.g. 2004/05) and dichotomized subjects into two groups, namely normal ($A = 0$) and limited ($A = 1$) hearing ability. Among 4232 patients at baseline, 772 (18.2%) had limited hearing ability. The longitudinal mediator "loneliness" was recorded from wave 3 (2006/07) to wave 7 (2014/15) and had two potential values, namely frequent ($M = 1$) vs. infrequent feeling of loneliness ($M = 0$). Alongside the mediator, four longitudinal confounders were recorded over time (i.e. wave 3 to wave 7), namely depression status (yes vs. no), mobility score (continuous), smoking status (non-smoker vs. current smoker) and alcohol status (non-drinker vs. current drinker). The baseline covariates consisted of 15 variables, including age at wave 2, gender, ethnicity (white vs. non-white), wealth (1=low, 5=high), education level (1 = no formal qualification, 2 = intermediate and 3 = higher education), marital status (yes vs. no), the use of hearing aids (yes vs. no), the presence of hypertension (yes vs. no), diabetes (yes vs. no), stroke (yes vs. no) and cancer (yes vs. no), alongside the baseline values of the four aforementioned time-varying confounders. A detailed description of these covariates, as well as the mediator distribution, the event and censoring rate over time is provided in the appendix and in previous works (Davies et al. 2017; Hackett et al. 2018; Davies-Kershaw et al. 2018; Rafnsson et al. 2020). Of note, all patients were followed up until one year after the last wave in 2014/15, which equals to a total of 11 years of follow-up.

We assumed that the relationship between the variables obeys the causal structure depicted in Fig. 1. Here, the mediators and confounders measured at time t_{k-1} are time-varying confounders of the relationship between the mediator measured at time t_k and the outcome. To estimate the exposure weights (step 1), we first considered a logistic (exposure) model adjusting for the main effects of all baseline covariates. To assess the potential of covariate-covariate interactions, we used a LASSO variable selection process (R package glmnet) to select the most important interaction terms from the set of all 105 possible two-by-two covariate interactions (i.e. a zero penalty was given to the main effect of all variables). The chosen interactions were then added into the exposure model. Of note, we selected the value of the tuning parameter that minimized the mean cross-validated error in a ten-fold cross-validation with deviance as loss function.

To estimate the mediation weights (step 3), we first considered a logistic (mediator) model adjusting for t_k, M_{k-1}, L_k and the main effects of all baseline covariates, i.e.:

$$\log \frac{\Pr(M_k = 1 | A = a, M_{k-1} = m_{k-1}, L_k = l_k, L_0 = l_0, T \geq t_{(k)})}{\Pr(M_k = 0 | A = a, M_{k-1} = m_{k-1}, L_k = l_k, L_0 = l_0, T \geq t_{(k)})}$$
$$= \gamma_0 + \gamma_1 t_{(k)} + \gamma_2 a + \gamma_3 m_{k-1} + \gamma_4 l_k + \gamma_5' l_0$$

We then improved this mediator model in a sequential manner. First, to assess whether the conditional distribution of M_k has a residual dependence upon the measurements of M and L that preceded M_{k-1} and L_k, we used the LASSO to determine the first measurement of M and L prior to M_{k-1} and L_k that were predictive for M_k, conditional on t_k, M_{k-1}, L_k and the main effects of all baseline covariates in L_0. This measurement and all measurements following this one up to M_{k-1} for M and L_k for L were included into the mediator model. Next, we assessed whether there were important (i) treatment-baseline/longitudinal covariate interactions, (ii) time-baseline covariate interactions and (iii) baseline covariate-covariate interactions that should be adjusted for. For each step, an independent LASSO variable selection process was performed to select the most important interaction terms from all possible interactions. The interactions that were chosen in the previous step were always included in the model of the subsequent steps (which implies no shrinkage on these terms in the subsequent steps). In each LASSO procedure, we selected the value of the tuning parameter that minimized the mean cross-validated error in a ten-fold cross-validation with deviance as loss function. The final model was refitted before calculating the mediation weights.

To estimate the censoring weights, we first considered a cause-specific proportional hazard model adjusting for M_k, L_k, the exposure A and the main effects of all baseline covariates L_0. We then improved this censoring model in a sequential manner. First, to assess whether the censoring hazard at time t had a residual dependence upon the history of M and L that preceded the time t_k for M and for L, we implemented a backward elimination process, using the Akaike information criterion to determine the first post-baseline measurement of M and L that were predictive for the censoring hazard at time t, conditional on the later measurements. This measurement and all measurements following this one were included into the censoring proportional hazard model. Next, as for the mediator model, we assessed whether there were

Table 4 Data analysis: estimation of the path-specific effect models (1) and (2)

Model	Coefficient	Estimate	95%CI	p-value
(1)	**Dementia** ($j = 1$)			
	α_{1j}	0.60	(0.09; 1.00)	0.01
	α_{2j}	0.01	(-0.01; 0.05)	0.44
	Death ($j = 2$)			
	α_{1j}	-0.13	(-0.48; 0.14)	0.46
	α_{2j}	0.00	(-0.01; 0.01)	0.91
(2)	**Dementia** ($j = 1$)			
	α_{1j}	0.59	(0.09; 1.00)	0.01
	α_{2j}	0.00	(-0.01; 0.02)	0.83
	α_{3j}	0.02	(-0.02; 0.08)	0.52
	Death ($j = 2$)			
	α_1	-0.12	(-0.48; 0.15)	0.49
	α_2	0.01	(-0.01; 0.02)	0.58
	α_3	-0.01	(-0.05; 0.02)	0.63

important (i) treatment-baseline/longitudinal covariate interactions and (ii) baseline covariate-covariate interactions that should be adjusted for. For each step, an independent backward elimination process was performed to select the most important interaction terms from all possible interactions of the same type. The interactions that were chosen in the previous step were always included in the model of the subsequent steps (which implies no exclusion of these terms in the subsequent steps). Note that we used backward elimination for the construction of the censoring models (as opposed to LASSO) due to the lack of prepackaged software that can apply LASSO or other advanced variable selection methods on a counting format survival dataset (Bien et al. 2013; Bickel et al. 2010). Results of the variable selection processes for the treatment, mediator and censoring models are reported in the Online Supplementary Materials.

The two path-specific effect proportional hazard models specific for dementia and for death (i.e. model (1) without interaction between a and a^*, and model (2) with this interaction) were then fitted using the calculated weights. The confidence intervals of the total, direct and indirect hazard ratios were derived by the non-parametric bootstrap method, with 5000 samples taken from the original data set by sampling with replacement. We then established the cumulative incidence curves of dementia and of death under different sets of a and a^*. As the interaction between a and a^* was not statistically significant at the 5% level, we only use the results of the path-specific effect model (1) (which omits the interaction between a and a^*) to establish the cumulative incidence curves.

Results of the estimation procedures are provided in Tables 4 and 5. As can be seen from these tables, the total effect of hearing loss on the time to dementia diagnosis was statistically significant (Model 1, HR = 1.84; 95%CI 1.10 to 2.71). Model (1) further suggested that this total effect was weakly mediated through feelings of loneliness, with a non-statistically significant indirect effect equal to 1.01 (HR scale; 95%CI

Table 5 Data analysis: the effect of limited vs. normal hearing ability on the time-to-event outcomes, mediated through the feeling of loneliness

Model	Effect	Hazard ratio	95%CI	Mediated proportion
(1)	**Dementia**			
	Total effect ($A = 1$ vs. $A = 0$)	1.84	(1.10; 2.71)	
	Direct effect	1.82	(1.10; 2.72)	
	Indirect effect	1.01	(0.99; 1.05)	2.0%
	Death			
	Total effect ($A = 1$ vs. $A = 0$)	0.88	(0.63; 1.15)	
	Direct effect	0.88	(0.62; 1.15)	
	Indirect effect	1.00	(0.99; 1.01)	− 0.8%
(2)	**Dementia**			
	Total effect ($A = 1$ vs. $A = 0$)	1.84	(1.10; 2.71)	
	Direct effect	1.84	(1.09; 2.72)	
	Indirect effect	1.00	(0.99; 1.02)	0.3%
	Total effect ($A = 0$ vs. $A = 1$)	0.54	(0.37; 0.91)	
	Direct effect	0.55	(0.37; 0.91)	
	Indirect effect	0.98	(0.93; 1.02)	3.2%
	Death			
	Total effect ($A = 1$ vs. $A = 0$)	0.88	(0.62; 1.15)	
	Direct effect	0.88	(0.61; 1.15)	
	Indirect effect	1.01	(0.99; 1.02)	− 4.0%
	Total effect ($A = 0$ vs. $A = 1$)	1.13	(0.87; 1.61)	
	Direct effect	1.13	(0.86; 1.62)	
	Indirect effect	1.00	(0.98; 1.04)	3.2%

The mediated proportion is calculated on log scale

0.99 to 1.05). This expresses that the hazard of dementia would increase by 1% if all patients were to have limited hearing ability but loneliness levels were switched from the values that would have been observed if they had normal hearing ability to the value observed under limited hearing. In contrast, the total effect of hearing loss on mortality was not statistically significant (Model 1, HR = 0.88; 95%CI 0.63 to 1.15). There was no statistical evidence of an indirect effect through feelings of loneliness (HR = 1.00, 95%CI 0.99 to 1.01). These findings did not change when considering model (2) with interaction (Table 4—p-value of the interaction coefficient equals 0.52 for dementia and 0.63 for death).

Figure 3 provides the estimated cumulative incidence curve of dementia for different a and a^*, which visualizes the weak indirect effect of hearing loss on dementia through the suggested longitudinal mediator. At some time points, there are large jumps in these curves due to the fairly high rate of interval censoring in the dataset (i.e. if the date for dementia diagnosis (or for death) was not known but the person had a new diagnosis (or passed away) from one visit to the next, then we considered the midpoint between two visits as the event date). Besides, note that the exposure (hearing impairment) was
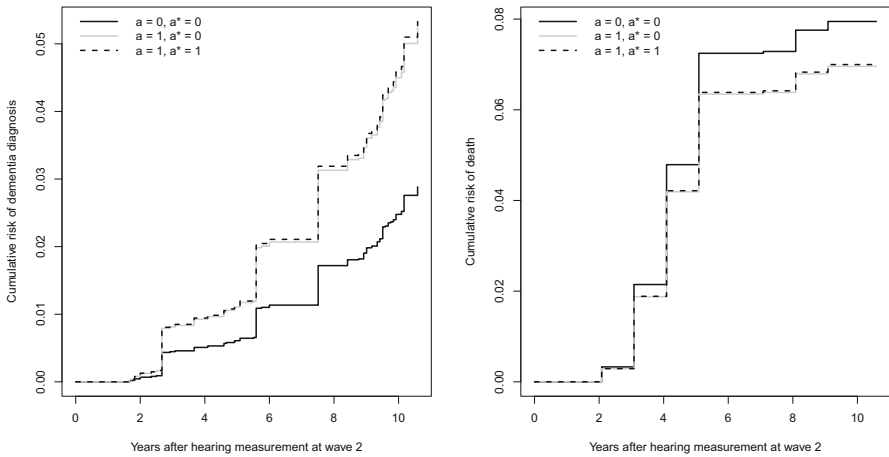
Fig. 3 Estimated cumulative incidence curves of dementia diagnosis and of death over time, if the whole population were to have hearing ability level a, but the loneliness levels $M_1, M_2, \ldots$ were switched from the values observed under hearing ability level a to the values observed under hearing ability level a^* (while the levels of the time-varying confounders were as observed under this joint intervention on A and M

not found to affect mortality. This implies that interpretation of the exposure effect on the cumulative incidence of dementia is not hindered by the exposure effect on the competing event.

The above findings should be interpreted with caution due to many potential limitations. First, there might be important baseline or time-varying confounders that were not measured and taken into account. For instance, genes affecting the likelihood of hearing loss may well be correlated with genes increasing the risk of dementia. Second, note that choosing time zero at wave 2 prevents bias due to left truncation, but renders interpretation slightly more subtle by ignoring age at study entry, which we did in order to display population-level results. Third, the obtained findings can be biased due to the involved models being incorrectly specified or censoring being informative (e.g. elderly patients who live alone might not come to the control visit due to dementia-related problems). Fourth, the data-adaptive methods used for constructing exposure, mediator and censoring models may affect the validity of the obtained confidence intervals and p-values. Addressing this complication is non-trivial and beyond the scope of this paper. In practice, a possible (heuristic) approach to overcome this could be to adjust for the union of all variables and/or interactions selected in all models, to ensure that no important confounders are omitted. This is inspired by the double selection principle proposed in the recent literature (Belloni et al. 2014). Further study might give insight into the extent to which this improves results, though we are not hopeful that a rigorous justification is possible.

4 Discussion

In this paper, we have generalized the weighting-based strategy proposed for natural effect models in single mediation analysis to the setting where the mediator of interest is repeatedly measured over time (hence subject to longitudinal confounders) and the primary outcome is a time-to-event endpoint, subject to competing risks. The proposed approach yields consistent estimates for the suggested path-specific direct and indirect effects if the causal assumptions hold, and the path-specific effect model and the conditional distribution of the exposure, mediator and censoring are correctly specified. As noted by Steen et al. (2017b), the mediator model needs careful consideration, especially when the exposure (and baseline covariates) are highly predictive of the mediator, for then even minor misspecification can have a major impact on the weights, lead to biased results and large variance due to extreme weights. While the presence of extreme weights might appear as a limitation at first, it may also diagnose severe model extrapolation that often goes unnoticed when using a repeated regression approach proposed for the same setting (Steen et al. 2017b). Simple weighting-based approaches also tend to yield larger standard errors (compared to imputation or regression-based approaches) due to lack in efficiency. This can be especially problematic when the mediator is continuous. In that case, the weight-based approaches tend to be unstable even under proper model specification. The repeated regression approach may be more appropriate for continuous mediators (Steen et al. 2017a; Vansteelandt et al. 2012, 2019).

Like in other settings, the presence of competing risks complicates the interpretation of the results. This is because the cumulative incidence of the event of interest is determined by the cause-specific hazards for both competing events, which may both be impacted by the exposure. Further insight may sometimes be gained by studying how the direct and indirect effects of exposure on the cumulative incidence of the event of interest would be like if we could intervene to hold the cause-specific hazard for the competing event fixed (Gran et al. 2015). While of interest, we have chosen not to do this in our study because it is unlikely that we have access to all common causes of dementia and mortality to render such further adjustment trustworthy.

Several proposals can be made to improve the suggested approach. For instance, instead of modeling the cause-specific hazard as discussed above, one may alternatively consider the Fine and Gray subdistribution hazard model to directly quantify the impact of covariates on the cumulative incidence function (Fine and Gray 1999). Future research might also focus on the development of doubly or multiply robust estimators (Bang and Robins 2005) to improve the robustness and efficiency of the current weight-based approach. The proposed strategy can also be easily extended to take into account multiple mediators $M^{(1)}, \ldots, M^{(V)}$ that are repeatedly measured over time. When these mediators are causally ordered then as suggested by VanderWeele and Vansteelandt (2014), one can first evaluate the effect mediated through $M^{(1)}$, then examine how much this changes when $M^{(1)}$ and $M^{(2)}$ are jointly considered as mediators. This then reveals the additional contribution of $M^{(2)}$ beyond $M^{(1)}$ alone. The process is then carried on by sequentially adding one mediator at a time until all V mediators are included. By accounting for multiple, repeatedly measured mediators, results of the analysis may allow one to get closer to evaluating the entire mediation

process that underlies the treatment mechanism in practice. Finally, future research should also extend the proposed approach to account for continuous exposures, which are quite common in epidemiology and social science.

Supplementary Information   The online version contains supplementary material available at https://doi.org/10.1007/s10985-022-09555-7.

Declarations

Conflict of interest   The authors declare that they have no conflict of interest.

References

Bang H, Robins JM (2005) Doubly robust estimation in missing data and causal inference models. Biometrics 61(4):962–973

Belloni A, Chernozhukov V, Hansen C (2014) Inference on treatment effects after selection among high-dimensional controls. Rev Econ Stud 81(2):608–650

Bickel PJ, Ritov Y, Tsybakov AB (2010) Hierarchical selection of variables in sparse high-dimensional regression. In: Borrowing strength: theory powering applications-a Festschrift for Lawrence D. Brown, Institute of Mathematical Statistics, pp 56–69

Bien J, Taylor J, Tibshirani R (2013) A lasso for hierarchical interactions. Ann Stat 41(3):1111

Davies HR, Cadar D, Herbert A, Orrell M, Steptoe A (2017) Hearing impairment and incident dementia: findings from the english longitudinal study of ageing. J Am Geriatr Soc 65(9):2074–2081

Davies-Kershaw HR, Hackett RA, Cadar D, Herbert A, Orrell M, Steptoe A (2018) Vision impairment and risk of dementia: findings from the english longitudinal study of ageing. J Am Geriatr Soc 66(9):1823–1829

Fine JP, Gray RJ (1999) A proportional hazards model for the subdistribution of a competing risk. J Am stat Assoc 94(446):496–509

Graffeo N, Latouche A, Le Tourneau C, Chevret S (2019) ipcwswitch: an r package for inverse probability of censoring weighting with an application to switches in clinical trials. Comput Biol Med 111:103339

Gran JM, Lie SA, Øyeflaten I, Borgan Ø, Aalen OO (2015) Causal inference in multi-state models-sickness absence and work for 1145 participants after work rehabilitation. BMC Public Health 15(1):1–16

Hackett RA, Davies-Kershaw H, Cadar D, Orrell M, Steptoe A (2018) Walking speed, cognitive function, and dementia risk in the english longitudinal study of ageing. J Am Geriatr Soc 66(9):1670–1675

Huang YT, Yang HI (2017) Causal mediation analysis of survival outcome with multiple mediators. Epidemiology (Cambridge, Mass) 28(3):370

Lange T, Vansteelandt S, Bekaert M (2012) A simple unified approach for estimating natural direct and indirect effects. Am J Epidemiol 176(3):190–195

Lange T, Rasmussen M, Thygesen LC (2013) Assessing natural direct and indirect effects through multiple pathways. Am J Epidemiol 179(4):513–518

Lin SH, Young JG, Logan R, VanderWeele TJ (2017) Mediation analysis for a survival outcome with time-varying exposures, mediators, and confounders. Stat Med 36(26):4153–4166

Martinussen T, Vansteelandt S, Andersen PK (2020) Subtleties in the interpretation of hazard contrasts. Lifetime Data Anal 26(4):833–855

Mittinty MN, Vansteelandt S (2020) Longitudinal mediation analysis using natural effect models. Am J Epidemiol 189(11):1427–1435

Pearl J (2001) Direct and indirect effects. In: Proceedings of the seventeenth conference on uncertainty in artificial intelligence, Morgan Kaufmann Publishers Inc., pp 411–420

Putter H, Fiocco M, Geskus RB (2007) Tutorial in biostatistics: competing risks and multi-state models. Stat Med 26(11):2389–2430

Rafnsson SB, Orrell M, d'Orsi E, Hogervorst E, Steptoe A (2020) Loneliness, social integration, and incident dementia over 6 years: prospective findings from the english longitudinal study of ageing. J Gerontol Se B 75(1):114–124

Robins JM, Greenland S (1992) Identifiability and exchangeability for direct and indirect effects. Epidemiology 3(2):143–155, http://www.jstor.org/stable/3702894

Robins JM, Hernan MA, Brumback B (2000) Marginal structural models and causal inference in epidemiology. Epidemiology 11(5):550–560

Steen J, Loeys T, Moerkerke B, Vansteelandt S (2017) Flexible mediation analysis with multiple mediators. Am J Epidemiol 186(2):184–193

Steen J, Loeys T, Moerkerke B, Vansteelandt S (2017) Medflex: an r package for flexible mediation analysis using natural effect models. J Stat Softw 76(11):1–46

Steptoe A, Breeze E, Banks J, Nazroo J (2013) Cohort profile: the english longitudinal study of ageing. Int J Epidemiol 42(6):1640–1648

Vandenberghe S, Duchateau L, Slaets L, Bogaerts J, Vansteelandt S (2018) Surrogate marker analysis in cancer clinical trials through time-to-event mediation techniques. Stat Methods Med Res 27(11):3367–3385

VanderWeele TJ (2016) Mediation analysis: A practitioner's guide. Ann Rev Public Health 37(1):17–32. https://doi.org/10.1146/annurev-publhealth-032315-021402, pMID: 26653405

VanderWeele T, Vansteelandt S (2014) Mediation analysis with multiple mediators. Epidemiol Methods 2(1):95–115

Vansteelandt S, Bekaert M, Lange T (2012) Imputation strategies for the estimation of natural direct and indirect effects. Epidemiol Methods 1(1):131–158

Vansteelandt S, Linder M, Vandenberghe S, Steen J, Madsen J (2019) Mediation analysis of time-to-event endpoints accounting for repeatedly measured mediators subject to time-varying confounding. Stat Med 38(24):4828–4840

Young JG, Stensrud MJ, Tchetgen Tchetgen EJ, Hernán MA (2020) A causal framework for classical statistical estimands in failure-time settings with competing events. Stat Med 39(8):1199–1236

Zheng W, van der Laan M (2017) Longitudinal mediation analysis with time-varying mediators and exposures, with application to survival outcomes. J Causal Inference 5(2):20160006