Mediation analysis of time-to-event endpoints accounting for repeatedly measured mediators subject to time-varying confounding

Stijn Vansteelandt,
Ghent University, Belgium
and London School of Hygiene and Tropical Medicine, U.K.
stijn.vansteelandt@ugent.be
Martin Linder,
Novo Nordisk, Denmark
Sjouke Vandenberghe,
Ghent University, Belgium
Johan Steen,
Ghent University Hospital, Belgium
and Jesper Madsen
Novo Nordisk, Denmark

Abstract

In this article, we will present statistical methods to assess to what extent the effect of a randomised treatment (versus control) on a time-to-event endpoint might be explained by the effect of treatment on a mediator of interest, a variable which is measured longitudinally at planned visits throughout the trial. In particular, we will show how to identify and infer the path-specific effect of treatment on the event time via the repeatedly measured mediator levels. The considered proposal addresses complications due to patients dying before the mediator is assessed, due to the mediator being repeatedly measured, and due to post-treatment confounding of the effect of the mediator by other mediators. We illustrate the method by an application to data from the LEADER cardiovascular outcomes trial.
Key-words: g-formula; mediation; longitudinal data; path-specific effect; time-dependent confounding.

1 Introduction

There is a growing interest in statistical analyses that support insight into causal mechanisms whereby an exposure affects intermediate variables (or mediators), to then in turn produce an outcome. In many such studies, interest lies in one or more mediators that are measured multiple times during the course of the study. The LEADER trial (1), for instance, evaluated the effect of liraglutide (as opposed to placebo), over and above standard care, on time from randomisation to first major adverse cardiovascular event (MACE) in patients with Type II diabetes and high cardiovascular risk, in accordance with FDA guidelines. Liraglutide is a once-daily injectable drug for the treatment of Type II diabetes, commonly branded as Victoza. A total of 9340 patients were randomised to either of the two treatments with a median follow-up time of 3.8 years. The primary analysis showed a protective effect of randomised assignment to liraglutide on time to first MACE, amounting to a hazard ratio of 0.87 (95% confidence interval 0.78 to 0.97; this is based on an intention-to-treat analysis, thus not adjusting for time periods off drug, which constituted only a minor fraction of the total patient observation time). Effects were also seen on other endpoints, e.g. glycated haemoglobin levels (HbA1c), blood pressure, body weight and urinary albumin to creatinine ratio, which were all repeatedly assessed at fixed time points 3 to 12 months apart. The question that motivated this research was to what extent the treatment effect on cardiovascular events is mediated by these repeatedly measured intermediate variables.

The literature on structural equation models provides methods for mediation analysis that can handle multiple, repeatedly measured mediators (2). The assumptions are strong, however. Besides relying on linear models without interactions for all involved outcomes and mediators, untestable additivity assumptions are often made at the individual level.

These assumptions, which are frequently left implicit, state for instance that the effect of treatment while holding the mediators fixed is the same for all individuals, which is biologically implausible (3). The causal inference literature on mediation analysis, pioneered by Robins and Greenland (3) and Pearl (4), has explicated these weaknesses and tried to overcome them by providing a framework that is also suitable to non-linear modelling. However, this framework has mostly confined itself to applications involving single mediators assessed at a single time. The reason is that difficulties of identification tend to pop up as confounders of the mediator - outcome association are themselves affected by treatment. This happens in particular when multiple mediators are at play or repeatedly measured mediators are assessed, for then the association between (a given) mediator (at a given time) and outcome may be confounded by other or previously measured assessments of the mediator.

Mediation analyses are ideally based on repeated assessments of the mediator for each individual. This is because the scientific interest typically lies in the effect of an exposure mediated via an entire mediator 'process'. E.g. the interest in the LEADER trial lies in the effect mediated, regardless of the time at which the mediator of interest was assessed. The indirect effect via a single assessment of the mediator is likely to capture only part of the indirect effect via the mediator 'process', as it does not pick up the indirect effects through earlier or later instances of the mediator. The absence of repeated mediator assessments therefore likely results in attenuation of the indirect effect.

When repeatedly measured assessments of the mediator are available, it is tempting to consider simplifying the problem of mediation analysis by aggregating the longitudinal mediators to a single one (e.g., in terms of some area under the curve, or the last recorded glycated haemoglobin level). Such simplified mediation analyses may provide a useful starting point, but are difficult to justify as the final analysis for various reasons.

As previously suggested, a single summary cannot usually capture the full complexity of the mediator, and this may lead to a weakening of the indirect effect. Such simplified mediation analysis moreover prohibits adequate control for confounding. For instance, area under the curve measurements combine the repeated assessments of the mediator over time; this prohibits adequate control for confounding because some of the covariate measurements measured during the study, will then be both cause and effect of the obtained area under the curve which typically has undesirable consequences (5). Basing the analysis on the last recorded level of the mediator may appear to overcome this problem, but is equally problematic. The reason is that the association between the last recorded level of the mediator and the time-to-event endpoint is confounded by previously recorded levels of the mediator; adjusting for these would be undesirable as it would eliminate part of the indirect treatment effect. Furthermore, the last recorded level of the mediator may be influenced by the event time itself (whenever the mediator is subject to a period effect), thereby inducing problems of reverse causality. In view of this, we will focus on approaches that explicitly acknowledge the repeated measures nature of the mediator.

As previously suggested, the fact that the association between (a given) mediator (at a given time) and outcome may be confounded by previously measured assessments of the mediator typically complicates identification. The effect of exposure on outcome as transmitted along a single pathway (e.g. the effect via the first assessment of the mediator alone) is however generally not of interest when the mediator is repeatedly measured: the primary scientific interest then lies in the effect via the 'mediator process', which is defined by a combination of pathways that involve the different assessments of the mediator over time. Interestingly, closer examination of the identification results for mediation analysis clarifies that the effect of exposure on outcome as transmitted along specific combinations of pathways (e.g. involving different or repeatedly measured mediators) is sometimes

4

easier to identify (6). In this article we will therefore make use of a general theory of identification in nonparametric structural equations models with independent errors and possible latent variables (6), to infer the effect of exposure on outcome transmitted along a combination of pathways. In doing so, we will address complications of working with a time-to-event endpoint, such as that individuals may die before the mediator is assessed.

In the next section, we describe the setting under which we will work. In Section 3, we first discuss our proposal to infer the direct and indirect effect of interest accounting for repeatedly measured mediators subject to time-varying confounding. We start with explaining how one may calculate these effects and end with a discussion of how the proposed approach can be viewed as a generalisation of dynamic path analysis (7), as well as how it relates to alternative proposals (8; 9; 10). In Section 4 we present the method applied to the LEADER data and compare the results to less complex methods (all variants of the common Cox regression model) and in Web Appendix B we present a simulation study in order to evaluate the behavior of our proposal. We conclude with some final remarks and ideas about possible extensions in Section 5.

2   Setting

Consider a study design which randomises independent patients $i = 1, ..., n$ over 2 treatment arms A_i, coded 1 for treatment and 0 for control, and intends to subsequently record longitudinal measurements of the mediator $M_{i1}, ..., M_{ik}$ at visits $1, ..., k$, along with a time-to-event endpoint T_i. In actual fact, mediator measurements are only recorded until the end-of-study time k or until the event of interest happens, whichever comes first. Our results will also be applicable to non-randomised exposures, as they will accommodate adjustment for possible baseline confounding variables L_0. Furthermore, the time-to-event endpoint may be censored administratively or due to loss to follow-up. We assume, for no-

tational convenience, that all patients are seen at the same, equidistant, time points. The latter restriction is readily relaxed (by using time points $t_{i1}, ..., t_{ik}$), provided that these observation times are non-informative (and the no unmeasured confounding assumptions that we will assume, remain plausible). We moreover assume the absence of competing risks at this point, which we will address in the discussion of this paper.

Informally, in relation to the example from the LEADER trial, our proposed analysis will infer how different the risk of being event-free at a given time would be in the liraglutide arm if the mediator levels for each patient in that arm changed to the levels that we would have seen for that patient on the placebo arm; we will give a formal description in Section 3. Because such mediation analysis conceptualises modifying the mediator, it will be important to control for confounding of the association between mediator and outcome at each time. Our subsequent analysis, like nearly all mediation analyses in the literature, assumes that sufficient data is available on prognostic factors of the event of interest that are also associated with the mediator, to trust that sufficient control for confounding can be made. We will use $L_{i0}, L_{i1}, ..., L_{ik}$ to denote those confounders (e.g. concomitant medication, ...) measured for patient i at baseline and at visits $1, ..., k$, respectively; L_{it} in particular includes the at-risk indicator $I(T > t)$, which is 1 for subjects who are event-free at time t and 0 otherwise. Thus L_{i0} refers to baseline covariates (e.g. age, gender, baseline level of the mediator, ...), and $L_{i1}, ..., L_{ik}$ refer to potential confounders measured at visits $1, ..., k$, the same visits at which the mediator levels were assessed. Importantly, we will assume throughout that those confounders L_{it} measured at visit t are not influenced by the mediator level assessed at that time, although we will allow for those confounders to affect mediator levels at time t (as well as at later times) and for them to have been influenced by mediator levels at earlier time points as in Figure 1. When only previous confounder measurements are known to influence the mediator at

6

visit t and the mediator at visit t may influence confounders at visit t and later times, then one must redefine L_{it} to include only covariates measured at visit $t - 1$.

Figure 1 about here.

The causal diagram in Figure 1 visualises the data-generating mechanism which we will postulate, in the absence of censoring (the additional complications posed by censoring of the event time will be addressed later). We will assume throughout that it represents a nonparametric structural equation model with independent errors (11; 12). It shows a clear, pre-specified causal ordering of the confounder and mediators over time, as considered in (13) and (14) for cross-sectional multiple mediator settings. In spite of this, we will not adopt the proposal by these authors for two reasons. First, it infers the mediated effect via each of the mediators at each time separately, which gives a more refined decomposition than we are aiming for. Indeed, our interest lies in the effect mediated by $M_{i1}, ..., M_{ik}$, regardless of the specific time. In particular, we will infer the effect of randomised assignment to treatment as transmitted along the combination of pathways whereby treatment *directly* influences one of the mediators $M_{i1}, ..., M_{ik}$ (not via $L_{i1}, ..., L_{ik}$), which may in turn influence the risk of the event of interest through an arbitrary mechanism (possibly via $L_{i2}, ..., L_{ik}$). Second, in their sequential approach, VanderWeele and Vansteelandt (13) and Steen et al. (14) generally assume that the mediators share no unmeasured common causes and moreover assume that none of the intermediate confounders share unmeasured common causes with the outcome. We will relax these assumptions for two reasons. First, to render the required no unmeasured confounding assumptions plausible, $L_{i0}, L_{i1}, ..., L_{ik}$ will likely include a large number of covariates, some of which may be influenced by treatment. This makes it unlikely that all these components are only associated with the endpoint of interest by means of a causal

7

effect. In our development below, as in Figure 1, we will therefore allow for the association between the covariates $L_{i0}, L_{i1}, ..., L_{ik}$ and the outcome to be itself confounded by possibly unmeasured factors U_l. In fact, our proposal will also be valid when - unlike suggested by Figure 1 - these unmeasured factors are time-varying and influenced by the history, including treatment. Second, it would be unlikely that repeated assessments of the mediator for the same individual have no unmeasured causes in common. In our development below, as in Figure 1, we will therefore allow for the association between the mediators $M_{i1}, ..., M_{ik}$ to be affected by possibly unmeasured factors U_m. Traditional longitudinal mediation analyses (2), as well as dynamic path analysis which can be viewed as an extension thereof to survival endpoints (7; 15), implicitly assume the absence of such common causes (or frailties) U_l and U_m, because they attempt a more refined decomposition of the exposure effect. They moreover invoke Markov assumptions which assume the absence of long term effects of covariates and mediators on covariates and mediators measured later in time; such assumptions are easy to avoid in our proposed approach below.

Note that the mediated effect on which we will focus, excludes pathways whereby treatment initially influences time-dependent patient characteristics L, which then in turn influence the mediator and thereby the risk of the event. Those pathways will be attributed to the indirect effect via those patient characteristics. This seems logical from an interpretational point of view, but is also a more fundamental requirement: we will later see that the effect of treatment transmitted along the combination of *all* pathways that intercept one or multiple mediators $M_{i1}, ..., M_{ik}$ (regardless of where in the causal chain it intercepts these variables) cannot be identified without making overly stringent assumptions.

8

3 Proposal

3.1 Estimands

To define the direct and indirect effects of interest, we will make use of so-called path-specific effects (16), expressed as differences or ratios of survival probabilities. In particular, we will calculate how likely it would be to be event-free at a chosen time t in the experimental treatment arm if the mediator levels for each patient changed to the levels that we would have seen if that patient had been assigned to control, but the levels of the time-varying confounders had otherwise remained unchanged. That is, the probability for a randomly chosen patient to be event-free at time t on treatment if L_1 took on the value $L_1(1)$ that we would have seen for that patient on treatment, if M_1 were set to the level $M_1(0, L_1(1))$ that we would have seen on control for that patient if L_1 had been set to the previously chosen value, if L_2 were set to the level $L_2(1, M_1(0, L_1(1)))$ that we would have seen on treatment for that patient if L_1 and M_1 had been set to the previously chosen values, and so on (see the Appendix for more detail). Let us denote the corresponding probability $S_{1,0}(t)$. When repeating this for all times t, we obtain two survival curves which one may contrast to visualise the targeted path-specific effect via M. In particular, one may express it as the contrast $S_{1,1}(t)/S_{1,0}(t)$, $S_{1,1}(t) - S_{1,0}(t)$ or $\{1 - S_{1,1}(t)\}/\{1 - S_{1,0}(t)\}$ for each time t. One may additionally contrast $S_{1,0}(t)$ with how likely it is to be event-free at that time in the control arm. Let us denote this probability $S_{0,0}(t)$. If we repeat this for all times t, we will once more obtain two survival curves which we may then contrast to visualise the path-specific effect not via M. In particular, one may express this as the contrast $S_{1,0}(t)/S_{0,0}(t)$, $S_{1,0}(t) - S_{0,0}(t)$ or $\{1 - S_{1,0}(t)\}/\{1 - S_{0,0}(t)\}$ for each time t. The mediated proportion can then be visualised as the ratio of the path-specific effect via M to the total intention-to-treat effect over time:

$$\frac{S_{1,1}(t) - S_{1,0}(t)}{S_{1,1}(t) - S_{0,0}(t)}.$$

In the example from the LEADER trial, one may moreover have interest in contrasting $S_{1,1}(t)$ and $S_{0,0}(t)$ with $S_{0,1}(t)$, the survival probability at time t in the control arm if the mediator levels for each patient changed to the levels that we would have seen if that patient had been assigned to experimental treatment, but the levels of the time-varying confounders had otherwise remained unchanged. The methodology presented in this article can easily be applied to derive this contrast as well.

One subtlety in the interpretation of $S_{1,0}(t)$ is that some patients may die sooner when assigned to control than when assigned to experimental treatment, in which case their mediator values on control may appear ill defined until the considered time t (8). Note however that the variables $L_t, t = 1, ..., k$, include the at risk indicator $I(T > t)$. Setting M_1 to the level $M_1(0, L_1(1))$, M_2 to the level $M_2(0, L_1(1), M_1(0, L_1(1)), L_2(1, M_1(0, L_1(1))))$, ... thus amounts to setting the mediators to the level they would have taken in the control arm if the history of time-varying confounders, including the survival status at that time, were set to the level on the experimental arm. We will discuss implications in more detail in the discussion section (see also Web Appendix A).

In the forthcoming sections, we will explain how one may calculate the probabilities $S_{1,0}(t)$ for different times t. The probabilities $S_{1,1}(t)$ and $S_{0,0}(t)$ can likewise be obtained upon reversing the codings 0 and 1 in the proposal below. They could also be obtained directly from a standard nonparametric analysis of both treatment arms, although we recommend calculating them in a model-based way as suggested below, to give results that are better comparable with the calculation of $S_{1,0}(t)$. Our results in this section derive from the general identification results in Web Appendix A. These identification results can be obtained via application of the edge g-formula (17), which is a generalisation of the well-

known g-formula to the identification of the distribution of so-called nested counterfactuals (or counterfactual responses to so-called edge interventions). The identification results are nonparametric, suggesting that arbitrary models can be used in each of the stages. However, below we will suggest the use of Cox regression models as a special case.

3.2 General procedure

In Web Appendix A, we show formally that the chance $S_{a,a^*}(t)$, for given values a and a^*, can formally be identified as

$$
\begin{aligned}
S_{a,a^*}(t) \;=\; & \int f\left(T > t \mid T > \lfloor t \rfloor, \overline{m}_{\lfloor t \rfloor}, \bar{l}_{\lfloor t \rfloor}, A = a\right) \\
& \times \prod_{s=1}^{\lfloor t \rfloor} f\left(m_s \mid T > s, \bar{l}_s, \overline{m}_{s-1}, A = a^*\right) f\left(l_s \mid T > s-1, \bar{l}_{s-1}, \overline{m}_{s-1}, A = a\right) dm_s dl_s,
\end{aligned}
$$

where $\lfloor t \rfloor$ is the visit time prior to (and including) time t, and we define $\overline{m}_s \equiv (m_1, ..., m_s), \bar{l}_s \equiv (l_1, ..., l_s)$, and $m_0 = \emptyset$; here, where $f\left(m_s \mid T > s, \bar{l}_s, \overline{m}_{s-1}, A = a^*\right)$ is shorthand notation for $f\left(M_s = m_s \mid T > s, \overline{L}_s = \bar{l}_s, \overline{M}_{s-1} = \overline{m}_{s-1}, A = a^*\right)$. Specialising this to times t between the first and second mediator assessment and noting that L_1 is composed of the at-risk indicator $I(T > 1)$ and patient characteristics V_1, the chance $S_{1,0}(t)$ can formally be calculated as

$$
\begin{aligned}
\int & P(T > t \mid T > 1, A = 1, m_1, v_1, l_0) f(m_1 \mid T > 1, A = 0, v_1, l_0) \\
& \times f(v_1 \mid T > 1, A = 1, l_0) P(T > 1 \mid A = 1, l_0) f(l_0) dm_1 dv_1 dl_0.
\end{aligned} \tag{1}
$$

The above identification results show some similarity to the g-formula and the mediational g-formula (18), but do not follow from those theories, which apply to non-nested counterfactuals only.

Monte Carlo integration can be used for evaluating the above identity, and moreover gives it intuitive meaning. This involves first fixing L_0 for each individual $i = 1, ..., n$ at the

11

observed value l_{0i}. Next, for each individual, a (possibly counterfactual) event time $t_i(1)$ is drawn from the distribution $f(t|A = 1, l_{0i})$ and next, for individuals with $t_i(1) > 1$, the counterfactual level $V_1(1)$ is fixed at a random draw $v_{1i}(1)$ from the distribution $f(v_1|T > 1, A = 1, l_{0i})$. Subsequently, for each individual with $t_i(1) > 1$, $M_1(0, L_1(1))$ is fixed at a random draw $m_{1i}(0, l_{1i}(1))$ from the distribution $f(m_1|T > 1, A = 0, V_1 = v_{1i}(1), l_{0i})$, and $T(1, L_1(1), M_1(0, L_1(1)))$ is fixed at a random draw $t_i(1, l_{1i}(1), m_{1i}(0, l_{1i}(1)))$ from the distribution $f(t|T > 1, A = 1, M_1 = m_{1i}(0, l_{1i}(1)), V_1 = v_{1i}(1), l_{0i})$; for each individual with $t_i(1) \leq 1$, $t_i(1, l_{1i}(1), m_{1i}(0, l_{1i}(1)))$ is fixed at $t_i(1)$. The chance $S_{1,0}(t)$ can then be estimated as the proportion of individuals with $t_i(1, l_{1i}(1), m_{1i}(0, l_{1i}(1))) > t$.

The main drawback of the above Monte-Carlo strategy is that it involves modelling the joint distribution of all variables, which becomes especially cumbersome when L_1 is high-dimensional. Upon rewriting (1) in terms of a series of nested conditional expectations

$$E \left(E \left[E \left\{ P(T > t | T > 1, A = 1, M_1, L_1, L_0) | T > 1, A = 0, L_1, L_0 \right\} | A = 1, L_0 \right] \right.$$
$$\left. \times P(T > 1 | A = 1, L_0) \right)$$

(in line with a common representation of the g-formula), repeated regressions can be used instead (see Web Appendix A for more detail). We will illustrate this for the data structure in Table 1, which shows artificial data for 10 patients; here E_i is an indicator for the event (1 if the event occurred while enrolled in the trial, 0 otherwise).

Table 1 about here.

In particular, at each time t, the chance $S_{1,0}(t)$ can be calculated as follows:

1. Fit a Cox regression model among people who survived the previous visit in the experimental arm in function of the history of mediators and confounders up to that visit, accounting for censoring in the default way. This analysis accommodates

12

non-informative censoring, given the history of measured mediators and confounders up to the considered visit. Next, use the fitted model to predict the chance of surviving the given time t for each patient in the study who survived the previous visit, setting the mediators and confounders to their observed values. Denote the result $Q_i^{\lfloor t \rfloor}(t)$ for patient i. Besides maximum partial likelihood estimators of the regression coefficients, this requires an estimator of the cumulative baseline hazard, for which we used the Breslow estimator. For instance, with parameter estimates $\hat{\alpha}_1 = 0.02, \hat{\alpha}_2 = 0.05$ and $\hat{\alpha}_3 = -0.05$ for the log hazard ratio corresponding to M_1, L_0 and L_1, respectively, the chance of surviving the given time t (e.g. 5 months) for the first patient is $\exp[-\hat{\Lambda}_0(5) \times \exp\{(0.02 \times 63.36) + (0.05 \times 62.79) + (-0.05 \times 62.36)\}]$, with $\hat{\Lambda}_0(5)$ the estimated cumulative baseline hazard which equalled 0.06 after 5 months.

2. Next, repeat the following for each of the previous visits $k = \lfloor t \rfloor$ to 1:

   (a) Regress $Q^k(t)$ on the history of the mediators $\bar{M}_{k-1} = (M_1, ..., M_{k-1})$ and the history of the covariates $\bar{L}_k = (L_0, ..., L_k)$ among people who were event-free at visit k in the control arm. Since $Q^k(t)$ lies between 0 and 1, a quasi-binomial regression with logit link may be an appropriate choice. Next, we can use the model to calculate a prediction $Q_m^k(t)$ for all subjects who were event-free at visit k, based on their observed data on mediators and confounders. For instance, with parameter estimates $\hat{\beta}_0 = 1.48, \hat{\beta}_1 = 0.049$ and $\hat{\beta}_2 = -0.051$ for the intercept and the log odds ratios corresponding to L_0 and L_1, respectively, the prediction $Q_{mi}^k(t)$ for $k = 1$ for the first patient who was event-free at the first visit would be calculated as $\text{expit}\{1.48 + (0.049 \times 62.79) + (-0.051 \times 62.63)\}$ and would equal 0.79.

13

(b) Regress $Q_m^k(t)$ on the history of mediators $\bar{M}_{k-1}$ and confounders $\bar{L}_{k-1}$ among people who were event-free at visit k in the experimental treatment arm. Since $Q_m^k(t)$ lies between 0 and 1, a quasi-binomial regression with logit link may be an appropriate choice. Use the model to calculate a prediction $Q_l^{k-1}(t)$ for all patients who were event-free at visit $k-1$, based on their observed data on mediators and confounders. For instance, with parameter estimates $\hat{\delta}_0 = 1.71$ and $\hat{\delta}_1 = -0.003$ for the intercept and the log odds ratio corresponding to L_0, the prediction $Q_{li}^{k-1}(t)$ for $k=1$ for the first patient is $\text{expit}\{1.71 + (-0.003 \times 62.79)\}$ and equals 0.82. Note that $Q_l^{k-1}(t)$ is now also estimated for patients who were not event-free at visit k, using their data on $\bar{M}_{k-1}$ and $\bar{L}_{k-1}$.

(c) Fit a Cox model among people who were event-free at visit $k-1$ in the experimental treatment arm, in function of the history $\bar{M}_{k-1}$ and $\bar{L}_{k-1}$. Use the fitted model to estimate the chance of surviving visit k for each patient in the study who was event-free at visit $k-1$, setting $\bar{M}_{k-1}$ and $\bar{L}_{k-1}$ to the observed covariate values. Let $Q^{k-1}(t)$ denote the product of this predicted value and the value of $Q_l^{k-1}(t)$ obtained in the previous step. For instance, with the following parameter estimate for the log hazard ratio, $\hat{\theta}_1 = 0.003$, the probability of being event-free at the first visit on the experimental treatment is calculated as $\exp[-\hat{\Lambda}_0(3) \times \exp\{0.003 \times 62.79\}]$ for patient 1, with $\hat{\Lambda}_0(3)$ the estimated cumulative baseline hazard which equalled 0.02 after the first visit.

3. When the previous steps have been repeated for visits $k = \lfloor t \rfloor$ to 1, then average the value $Q^0(t)$ obtained in the final step across all patients. The resulting average is an estimate of $S_{1,0}(t)$. Averaging $Q^0(t)$ across all patients in the example would thus result in an estimate of $S_{1,0}(t)$ at 5 months, which equals 0.80.

14

Note that step 2(a) involves regressing predictions among patients who were event-free in the control arm (in order to integrate over the distribution of the mediator), while step 2(b) involves regressing predictions among patients who were event-free in the experimental treatment arm (in order to integrate over the distribution of time-varying confounders). This differential selection of patients in the respective steps is reflective of the 'construction' of nested counterfactuals.

In the above procedure, we have chosen to use separate Cox models at each time at which intermediate variables are assessed. Under correct specification of these models, the required censoring assumption is that censoring at each study visit k is non-informative, in the sense that the decision to discontinue the trial at a given time has no residual dependence on the remaining survival time amongst (alive and participating) patients in the same trial arm with the same history of the observed measurements $\bar{M}_k$ and confounders $\bar{L}_k$ at that time. Alternatively, one may choose to obtain predictions from a Cox model with time-varying covariates and avoid the need for separate Cox regression models at each visit time. This approach has the advantage of yielding potentially more precise predictions as this Cox regression model is fitted on the whole sample. The disadvantage however is a greater risk of misspecifying this Cox regression model. In the above procedure, we could alternatively have used parametric survival models using splines for the time effect (19), or binomial regression models for the chance of surviving time t, separately for each time t. We have not considered the latter option because the information may become sparse at the later time points, which does not pose complications when relying on the proportional hazards assumption in the Cox regression model.

For computational convenience and to limit the modelling efforts, we have chosen to use binomial regression models to model the predictions obtained from the Cox model. One concern about this strategy is that the considered binomial regression model may

15

fail to be congenial with the chosen Cox model. In view of this, we recommend the use of quasi-binomial regression with a logit (as opposed to probit) link. The use of such models ensures that when the above procedure is employed for the calculation of $S_{0,0}(t)$ or $S_{1,1}(t)$, misspecification (and in particular lack of congeniality) of the logistic regression models for $Q_m^k(t)$ and $Q_l^k(t)$ does not induce bias, because the average of the fitted values from a logistic regression model (in the treatment or control arm) equals the average of the outcome under that model (in the treatment or control arm), regardless of whether the model is correctly specified (see e.g. (20)). To additionally ensure unbiased estimation of $S_{1,0}(t)$ or $S_{0,1}(t)$, we further recommend that these logistic regression models obey the structure of the models that were used to obtain the predictions which they use as input. For instance, if the Cox model contains interactions between A and L_k, then these should also be included in the logistic regression model for $Q^s(t), s < k$; if the Cox model contains interactions between A and M_k, then these will likely give rise to interactions between A and predictors of M_k in the logistic regression model for $Q^s(t), s < k$.

3.3   Dynamic path analysis

The proposed approach can be viewed as a generalisation of dynamic path analysis (7; 15), which itself extends linear structural equation analysis to additive hazard models and normally distributed mediators which obey additive linear models. To see this, we first consider a setting with a single mediator and no confounders for pedagogic purposes. Assuming an additive hazard model (21) for the time-to-event outcome

$$\lambda(t|A, M_1) = \lambda_0(t) + \lambda_1(t)A + \lambda_2(t)M_1 I(t > 1)$$

and a linear regression model for the mediator

$$\mathrm{E}(M_1|T > 1, A) = \alpha_0 + \alpha_1 A,$$

with normal errors and constant variance σ^2, we can evaluate expression (1) as follows. First, since the survival probability at time t is $\exp\{-H(t)\}$, with $H(t)$ the cumulative hazard at time t, the survival probability $P(T > t | T > 1, A, M_1)$ can be evaluated as $\exp\{-\Delta\Lambda_0(t) - \Delta\Lambda_1(t)A - \Delta\Lambda_2(t)M_1\}$ with $\Delta\Lambda_j(t) = \Lambda_j(t) - \Lambda_j(1)$ and $\Lambda_j(t) = \int_0^t \lambda_j(s)ds$ for $j = 0, 1, 2$ the cumulative hazard. Second, using that the moment generating function $\mathbb{E}\{\exp(Zt)\}$ of a normal variate $Z \sim N(\mu, \sigma^2)$ equals $\exp(\mu t + \frac{\sigma^2 t^2}{2})$, integrating out M_1 in expression (1) yields

$$\exp\Big\{-\Delta\Lambda_0(t) - \Delta\Lambda_1(t) - \Delta\Lambda_2(t)E(M_1 | T > 1, A = 0) -$$
$$\frac{\Delta\Lambda_2(t)^2}{2}\mathrm{Var}(M_1 | T > 1, A = 0)\Big\},$$

which equals

$$\exp\Big\{-\Delta\Lambda_0(t) - \Delta\Lambda_1(t) - \Delta\Lambda_2(t)\alpha_0 - \frac{\Delta\Lambda_2(t)^2}{2}\sigma^2\Big\}.$$

Finally, multiplying this result with $P(T > 1|A) = \exp\{-\Lambda_0(1) - \Lambda_1(1)A\}$ yields

$$P(T_{1,0} > t) \;\; = \;\; \exp\Big\{-\Lambda_0(t) - \Lambda_1(t) - \Delta\Lambda_2(t)\alpha_0 - \frac{\Delta\Lambda_2(t)^2}{2}\sigma^2\Big\}.$$

Likewise,

$$P(T_{0,0} > t) \;\; = \;\; \exp\Big\{-\Lambda_0(t) - \Delta\Lambda_2(t)\alpha_0 - \frac{\Delta\Lambda_2(t)^2}{2}\sigma^2\Big\}$$
$$P(T_{1,1} > t) \;\; = \;\; \exp\Big\{-\Lambda_0(t) - \Lambda_1(t) - \Delta\Lambda_2(t)\alpha_0 - \Delta\Lambda_2(t)\alpha_1 - \frac{\Delta\Lambda_2(t)^2}{2}\sigma^2\Big\},$$

from which at each time t, the path-specific effect not via M and via M can be calculated in terms of survival probabilities as

$$PDE(t) = \frac{S_{1,0}(t)}{S_{0,0}(t)} = \exp\{-\Lambda_1(t)\}$$

and

$$PIE(t) = \frac{S_{1,1}(t)}{S_{1,0}(t)} = \exp\{-\Delta\Lambda_2(t)\alpha_1\} = \exp\{-(\Lambda_2(t) - \Lambda_2(1))\alpha_1\},$$

respectively, for $t > 1$ and 1 for $t \leq 1$. These expressions hold more generally for studies with multiple waves if the event time obeys an additive hazard model at each time (conditional upon the information observed until that time) and the mediator in each wave of the study obeys a normal linear regression model with additive effects, constant variance and normal errors. Similar expressions are also used in dynamic path analysis (7). The only difference lies in the indirect effect that has the additional term $\Lambda_2(1) \times \alpha_1$ in our proposal. This shows a first limitation of dynamic path analysis as currently considered in (7) in that this strategy ignores the potential for the event happening prior to the mediator assessment of the first wave. If all patients survive till the first mediator measurement, as in the examples in (7), $\Lambda_2(1)$ equals zero and the approaches will coincide. As such our results give formal justification for the expressions considered in dynamic path analysis as expressing path-specific effects under the above listed assumptions. They also signal the other limitations of dynamic path analysis in that it is limited to specific additive models for the event time and the mediators, and that it cannot (easily) accommodate time-varying confounders, nor lagged effects of the mediators.

4   Example from the LEADER trial

LEADER was a multi-centre, international, randomised, double-blind clinical trial evaluating liraglutide ($A = 1$) against placebo ($A = 0$), both added to standard of care. The trial was designed in accordance with guidance from FDA (22) regarding the evaluation of cardiovascular risk for new antidiabetic therapies. In total, 9340 patients with Type II diabetes at high risk for cardiovascular disease were randomised to one of the two treatment groups with a median follow-up time of 3.8 years (range 3.5-5.0 years). The subjects were attending planned visits 3 months after randomisation and subsequently every 6 months hereafter where at least HbA1c (M) was measured. The primary endpoint was the time T

from randomisation to first occurrence of a so called major cardiovascular event (MACE) defined as non-fatal myocardial infarction, non-fatal stroke or cardiovascular death. A first MACE occurred in significantly fewer patients in the liraglutide group (13.0%) than in the placebo group (14.9%) corresponding to an estimated hazard ratio of 0.87 (95% CI [0.78; 0.97]; P = 0.01) in the pre-specified primary analysis, which was a Cox regression model with treatment as the only fixed effect.

Significant positive effects of liraglutide vs. placebo were also found on cardiovascular risk factors such as glycated hemoglobin (HbA1c), body weight, urinary albumin to creatinine ratio and blood pressure, and the interest is to evaluate to what extent these potential pathways might explain liraglutide's protective effect on cardiovascular events. The main results from the study can be found in (1).

Figure 2 about here.

To illustrate the proposed approach for mediation analysis, we shall here restrict the attention to the potential mediation on the primary endpoint via the effect of liraglutide on HbA1c levels. Throughout, we will moreover assume that censoring is non-informative in the sense that the time to first MACE is equally distributed in patients who do versus do not discontinue the trial at a given time, but were assigned to the same arm and have the same history of HbA1c at that time. In Figure 2, the mean HbA1c levels over time are shown, as estimated by a mixed model for repeated measurements with adjustment for baseline covariates. As reported in (1), the estimated treatment difference was -0.40 percentage points (95% CI $[-0.45; -0.34]$) at the 36-month visit, which was the last scheduled visit with laboratory testing for the entire trial population. In Figure 3, the estimated survival curves $S_{1,0}(t)$ are shown along with $S_{1,1}(t)$ and $S_{0,0}(t)$. The survival curves were estimated in accordance to the method presented in Section 3 where HbA1c

measured after 3, 6, 12, 18, 24, 30 and 36 months, respectively, were included as mediators. Due to the novelty of the results, we could not present analyses that additionally adjusted for confounders other than supplementary HbA1c measurements, e.g. body weight, insulin use, ... Therefore, the results (which are also based on models with additive effects) are only to be considered hypothesis-generating and should be interpreted with caution. The proposed method provides a framework for further analyses that can adjust for various confounders and explore other potential pathways than glycaemic control represented by HbA1c. More exhaustive analyses are reserved for subsequent communication in a medical journal.

Figure 3 about here.

Figure 4 visualises the estimated mediated proportions at each visit. Inference was based on the nonparametric bootstrap with 1000 resamples. The results indicate that HbA1c is mediating parts of the effect of liraglutide on time to first MACE. Note however that the estimates of the mediated proportions are subject to uncertainty (reflected in the wide confidence intervals), and potentially residual confounding bias. Moreover, the mediated proportion appears to grow over time while the total effect as measured by the hazard ratio is constant. This may suggest that the way by which glycaemic control represented by HbA1c biologically influences the cardiovascular risk may be complex. For instance, it may be the case that the mediated proportion over time depends on the number of available HbA1c measurements that are included in the analysis. Nearly identical results were found when body weight, urinary albumin to creatinine ratio and systolic blood pressure were considered as additional potential mediators (not shown).

Figure 4 about here.

As a comparison to the method presented in Section 3, three relatively simple mediation analyses were also conducted:

1. Change in HbA1c after 6 months was used as mediator and included as a covariate in a Cox regression model with treatment as fixed factor, and baseline HbA1c as additional covariate.

2. Change in HbA1c over time was used as mediator and included as a time-dependent covariate in a Cox regression model with treatment as fixed factor and baseline HbA1c as additional covariate.

3. Change in the trapezoidal area under the HbA1c curve divided by time (updated mean) was used as mediator and included as a time-dependent covariate in a Cox regression model with treatment as fixed factor and baseline HbA1c as additional covariate.

For all three models, the mediated proportion was calculated as the difference in log hazard ratios between the models without and with the mediator, respectively, divided by the log hazard ratio from the model without the mediator. The results, which can be found in Table 2, are included because they represent rather standard analyses, even though they are known to be biased. Analyses 1 and 3 indicate some mediation. However, note that these mediated proportions have been calculated on different scales and are difficult to interpret because they reduce the mediator process to a single summary and, partly as a result, do not properly adjust for time-varying confounding by the mediator history. As pointed out in Section 1, mediation analyses based on a single summary measure of the mediator may result in an underestimation of the mediated proportion, and this could very likely be the explanation for these differences. Furthermore, analysis 2 does not indicate any clear mediation, which is in line with the expectation that an analysis based

on the last recorded value of the mediator will eliminate parts of an existing indirect effect. Also this analysis ignores confounding by the history of the mediators.

Table 2 about here.

5   Discussion

In this paper we have proposed a strategy to infer the effect of a randomised treatment on a time-to-event outcome as transmitted along the combination of pathways, whereby treatment directly - other than through a sufficient set of time-varying confounders - influences one of a sequence of repeatedly measured mediator measurements $M_{i1}, ..., M_{ik}$, which in turn influence the risk of the event of interest through an arbitrary mechanism. This proposal builds on the general theory on identification of path-specific effects in non-parametric structural equations models with independent errors (6). It can be considered as a generalisation of the structural equation models extension to time-to-event outcomes, the so-called dynamic path analysis (7), but can be used in more realistic settings as it can handle events happening prior to the first assessment of the mediator, is not limited to specific additive models for the event time and mediators, and easily accommodates time-varying confounders and long-term effects of mediators and covariates. As the general identification results in Web Appendix A are non-parametric, arbitrary models can be used in each of the steps of the procedure and our proposal is thus not limited to time-to-event outcomes. With a continuous outcome for instance, the Cox regression models at each wave would be replaced with a single model for the mean outcome at time t. Limited simulation studies in Web Appendix B demonstrate the adequate performance of the proposed methodology. SAS code for running this analysis is available in the online Supplementary Materials.

Our results shed light on the subtle interpretation of direct and mediated effects in studies with time-to-event endpoints. The notion of a direct effect conceptualises fixing the level of the mediator in the experimental treatment arm as it would have been on the control arm, but if a patient would have lived longer on the experimental treatment arm, then it becomes vague at what level the mediator ought to be controlled. For that reason, we have considered fixing the mediator at the level that would have been seen on the control arm if the patient had been kept alive for the same duration as in the experimental arm. In some cases, one may well hypothesize what would have happened if an event such as death had been prevented. For instance, if one of the study participants dies in a car crash, then we may well consider what that person's mediator level would have been at a given time had the car crash been prevented. In other cases, this is much harder to conceptualise. For instance, it is more difficult to imagine interventions that would prevent the event of interest in the considered patient populations. Strictly speaking one does not need to be precise about the kind of underlying interventions as inferences apply to all interventions that are non-invasive in the sense that if they had been applied to individuals who remained event-free, the same data for the mediator would have been observed. However, it does complicate interpretation in the same way as treatment effects can be difficult to interpret in the presence of drop-out due to death (23).

A simple fix to the above problem can be made in extreme cases where treatment is beneficial for all patients. In that case, $S_{0,1}(t)$ is always well and unambiguously defined, and thus one can use the contrast of $S_{0,1}(t)$ and $S_{0,0}(t)$ as a measure of indirect effect, and the contrast of $S_{1,1}(t)$ and $S_{0,1}(t)$ as a measure of direct effect. Alternatively, and more generally, note that expression (1) can also be interpreted as a so-called (randomised) interventional effect (24). In particular, as in (9), it can be interpreted as the chance

of being event-free at time t if all patients were randomised to liraglutide and at each time $s < t$, the mediator were randomly drawn from the distribution of the mediator in patients in the control arm who survived time s and have the same history of covariate data (as observed under the considered regime). One potential concern here is that patients who survive time s on the control arm may fail to be comparable with those who survive time s on the intervention arm, even after adjusting for the history of time-varying covariates. This may well happen as a result of survivor bias, due to which surviving patients may become more and more selective as time goes by. In our proposal, we have excluded this possibility by assuming the absence of unmeasured common causes of mediator (e.g. M_1) and time-varying confounders (e.g. L_1) in the causal diagram of Figure 1. Such assumptions tend not to be spelled out in the existing approaches for (randomised) interventional direct and indirect effects. Also Didelez (10) proposes related estimands. These estimands are well-defined, even when patients on the control arm tend to experience events sooner than those on the intervention arm. They are moreover identifiable under weaker assumptions than the considered path-specific effects, as they are not defined in terms of cross-world counterfactuals, but have not been formally extended to settings with confounding by time-varying covariates.

Our proposed approach is thus related to (9) and (10), but in contrast to these, focuses on the identification of path-specific effects. VanderWeele and Tchetgen Tchetgen (18) also adopted these interventional direct and indirect effects, but they consider random draws from the distribution of the mediator at a certain exposure level conditional on only baseline covariate data. Their indirect effect, unlike ours, thus includes pathways whereby the treatment influences time-varying confounders, then in turn influences the mediator and via that also the outcome. Their proposal also has the disadvantage that draws from the mediator distribution are ill defined when patients may die during the study. Lin et al.

(8) handled this problem by redefining nested counterfactuals to include a counterfactual survival status in a similar way as in our proposal; in doing so, they eliminate pathways whereby treatment influenced the mediator via survival from the indirect effect, as in our proposal. However, in contrast to (9) and our proposal, they do not consider random draws from the mediator distribution conditional on time-varying confounders. This raises questions whether the values drawn for a given patient will be sufficiently representative for what that patient might have 'naturally' experienced, making these estimands less suitable to develop insight into mechanism. For instance, the assessment of a direct effect demands fixing the mediator at subject-specific levels and it is unlikely that these can be 'predicted' well when only baseline confounders are used for prediction.

Zheng et al. (9) deemed the assumptions needed to infer the path-specific effects with multiple mediators and time-varying confounders too strong for the purpose of effect mediation in a survival study. In Web Appendix A, we argue that following arguments of Shpitser (6) and the recanting witness criterion (16), the path-specific effects represented by the causal diagram of Figure 1, are identified if a) the exposure A is randomly assigned, b) all common causes of the time-to-event outcome and the mediator at each time are measured and c) that this causal diagram represents a nonparametric structural equation model with independent error terms (12; 25). In particular, this means that the only variation in the variables (and their counterfactual values) on the causal diagram (not explained by previous variables in the diagram) is due to mutually independent error terms. Assumption b) further implies that unmeasured common causes of the mediators over time are allowed (as long as they do not directly influence the time-to-event outcome), as well as unmeasured common causes of baseline and time-varying confounders and the time-to-event endpoint. Note that common causes of time-varying confounders and mediators are not allowed to be unmeasured.

The above results ignored competing risks due to death from non-cardiovascular causes because the adjusted (Aalen-Johansen) cumulative incidence curves were almost identical to the Kaplan-Meier-curves, suggesting that no appreciable differences can be expected when accounting for competing risks. However, the proposed approach can relatively easily be extended to handle competing risks. In that case, in each step of the algorithm, we substitute the probability of being event-free by time t by the probability of being either event-free by time t or having experienced a competing event. The latter can be calculated either by combining the results from standard Cox regression models for the two cause specific hazards for both causes (26), or using binomial regression models for the cumulative incidence. Another obvious extension to the above proposal are settings where the mediator is multivariate at each time, as may be the case when examining the effect mediated via e.g. glycated hemoglobin and body weight. It then infers the effect of treatment mediated via at least one of those mediators. This does not render the procedure any more complicated, in the sense that it requires no additional modelling. The above procedure also readily extends to enable decomposition of the treatment effects via multiple mediators. For instance, consider two mediators $M_t^{(1)}$ and $M_t^{(2)}$, and a vector of covariates L_t^* at each time $t = 1, ..., k$, and suppose that $M_t^{(1)}$ may influence $M_t^{(2)}$, but not vice versa. Suppose furthermore that L_t^* may influence $M_t^{(1)}$ and $M_t^{(2)}$, but not vice versa. Then one may use the above procedure with $M_t \equiv M_t^{(2)}$ and $L_t \equiv (M_t^{(1)}, L_t^*)$ to infer the effect of treatment mediated via $M_t^{(2)}$ for $t = 1, ..., k$. One may likewise use the above procedure with $M_t \equiv M_t^{(1)}$ and $L_t \equiv (M_{t-1}^{(2)}, L_t^*)$ to infer the effect of treatment mediated via $M_t^{(1)}$ for $t = 1, ..., k$. When both effects are expressed on the risk difference scale, one may subtracting the sum of both effects from the total effect $S_{1,1}(t) - S_{0,0}(t)$ to obtain the direct treatment effect, which is not mediated by either $M_t^{(1)}$ or $M_t^{(2)}$, $t = 1, ..., k$.

References

[1] Marso SP, Daniels GH, Brown-Frandsen K, et al. Liraglutide and Cardiovascular Outcomes in Type 2 Diabetes.. *The New England Journal of Medicine* 2016; 375: 311-322.

[2] MacKinnon DP. *Introduction to Statistical Mediation Analysis*. Multivariate Applications SeriesLawrence Erlbaum Associates . 2008.

[3] Robins JM, Greenland S. Identifiability and Exchangeability for Direct and Indirect Effects.. *Epidemiology* 1992; 3(2): 143-155.

[4] Pearl J. Direct and indirect effects.. In: Morgan Kaufmann: San Francisco; 2001: 411–420.

[5] Robins JM, Hernan MA, Brumback B. Marginal Structural Models and Causal Inference in Epidemiology. *Epidemiology* 2000; 11(5): 550-560.

[6] Shpitser I. Counterfactual graphical models for longitudinal mediation analysis with unobserved confounding.. *Cognitive Science* 2013; 37(6): 1011-1035.

[7] Strohmaier S, Roysland K, Hoff R, Borgan O, Pedersen TR, Aalen OO. Dynamic path analysis - a useful tool to investigate mediation processes in clinical survival trials.. *Statistics in Medicine* 2015; 34: 3866-3887.

[8] Lin SH, Young JG, Logan R, VanderWeele TJ. Mediation analysis for a survival outcome with time-varying exposures, mediators, and confounders. *Statistics in medicine* 2017; 36(26): 4153–4166.

[9] Zheng W, Laan v. dM. Longitudinal mediation analysis with time-varying mediators

and exposures, with application to survival outcomes. *Journal of causal inference* 2017; 5(2).

[10] Didelez V. Defining causal meditation with a longitudinal mediator and a survival outcome. *Lifetime data analysis* 2018: 1–18.

[11] Pearl J. *Causality: Models, Reasoning and Inference.* Cambridge University Press . 2000.

[12] Robins JM, Richardson TS. Alternative Craphical Causal Models and the Identification of Direct Effects.. In: Shrout P. , ed. *Causality and Psychopathology: Finding the Determinants of Disorders and Their Cures.*Oxford, England: Oxford University Press. 2010 (pp. 103-158).

[13] VanderWeele TJ, Vansteelandt S. Mediation Analysis with Multiple Mediators. *Epidemiologic Methods* 2013; 2: 95-115.

[14] Steen J, Loeys T, Moerkerke B, Vansteelandt S. Flexible mediation analysis with multiple mediators. *American journal of epidemiology* 2017; 186(2): 184–193.

[15] Fosen J, Ferkingstad E, Borgan O, Aalen OO. Dynamic path analysis - a new approach to analyzing time-dependent covariates.. *Lifetime Data Analysis* 2006; 12: 143-167.

[16] Avin C, Shpitser I, Pearl J. Identifiability of Path-Specific Effects.. In: Morgan Kaufmann Publischers Inc; 2005; San Francisco, CA, USA: 357-363.

[17] Shpitser E. Causal inference with a graphical hierarchy of interventions. *Annals of Statistics* 2016; 6: 2433-2466.

[18] VanderWeele TJ, Tchetgen Tchetgen EJ. Mediation analysis with time varying exposures and mediators. *Journal of the Royal Statistical Society B* 2017; 79(Part 3): 917-938.

[19] Royston P PM. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Statistics In Medicine* 2002; 21: 2175–2197.

[20] Vandenberghe S, Vansteelandt S, Loeys T. Boosting the precision of mediation analyses of randomised experiments through covariate adjustment. *Statistics In Medicine* 2017; 36(6): 939-957.

[21] Aalen OO. A linear regression model for the analysis of life times. *Statistics in Medicine* 1989; 8(8): 907-925.

[22] CDER F. *Guidance for industry: diabetes mellitus — evaluating cardiovascular risk in new antidiabetic therapies to treat type 2 diabetes.* Department of Health and Human Services; Silver Spring, MD: 2008.

[23] Wen L, Terrera GM, Seaman SR. Methods for handling longitudinal outcome processes truncated by dropout and death. *Biostatistics* 2017; 19(4): 407–425.

[24] Vansteelandt S, Daniel RM. Interventional Effects for Mediation Analysis with Multiple Mediators.. *Epidemiology* 2017; 28: 258-265.

[25] VanderWeele TJ, Vansteelandt S, Robins JM. Effect decomposition in the presence of an exposure-induced mediator-outcome confounder.. *Epidemiology* 2014; 25(2): 300-306.

[26] Putter H, Fiocco M, Geskus RB. Tutorial in biostatistics: Competing risks and multi-state models.. *Statistics in Medicine* 2007; 26: 2389-2430.
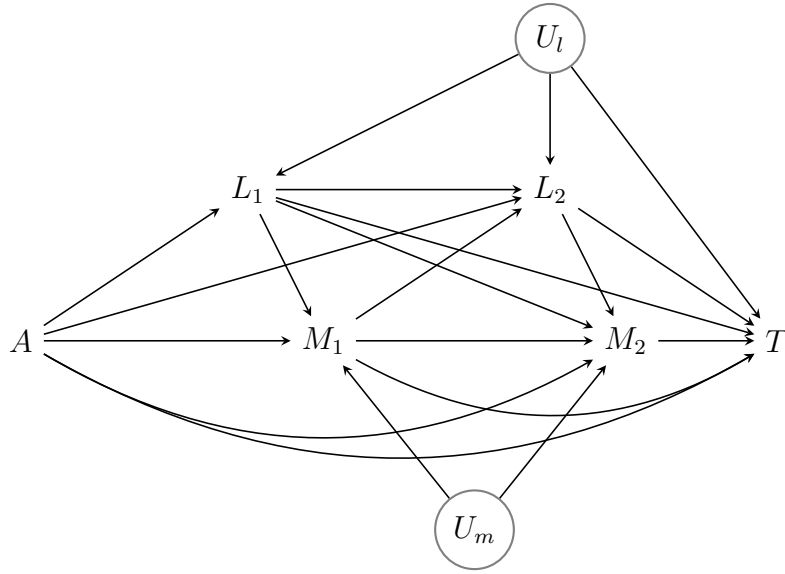
Figure 1: Causal diagram. U_m and U_l refer to unmeasured variables. The measured time-varying confounders L_1 and L_2 include survival at visit 1 and 2 (T may thus be viewed as survival beyond visit 2). Besides the assumptions embodied in this diagram, we assume that censoring at each time is non-informative in each trial arm, given the history of measured time-varying confounders and mediators at that time, in the sense defined in the main text.

Table 1: A toy example for a restricted set of patients.

| Patient	A_i	L_{0i}	L_{1i}	M_{1i}	E_i	T_i	$Q_i^1(t)$	$Q_{mi}^1(t)$	$Q_{li}^0(t)$	$Q_i^0(t)$
1	0	62.79	62.36	63.36	1	5.57	0.81	0.79	0.82	0.98
2	0	64.75	65.96	75.78	1	0.65	.	.	0.82	0.97
3	0	57.13	56.35	74.80	1	11.44	0.77	0.80	0.83	0.98
4	1	56.28	55.27	53.08	1	9.42	0.84	0.80	0.83	0.98
5	1	72.55	68.05	59.07	1	13.68	0.85	0.82	0.82	0.97
6	1	67.61	61.02	54.17	1	9.61	0.87	0.84	0.82	0.98
7	0	52.84	46.85	65.93	1	6.70	0.84	0.84	0.83	0.98
8	1	65.16	58.16	51.88	0	24.00	0.88	0.84	0.82	0.97
9	0	62.69	59.91	66.51	1	5.61	0.82	0.81	0.82	0.98
10	1	74.23	65.88	50.62	0	24.00	0.89	0.85	0.82	0.97
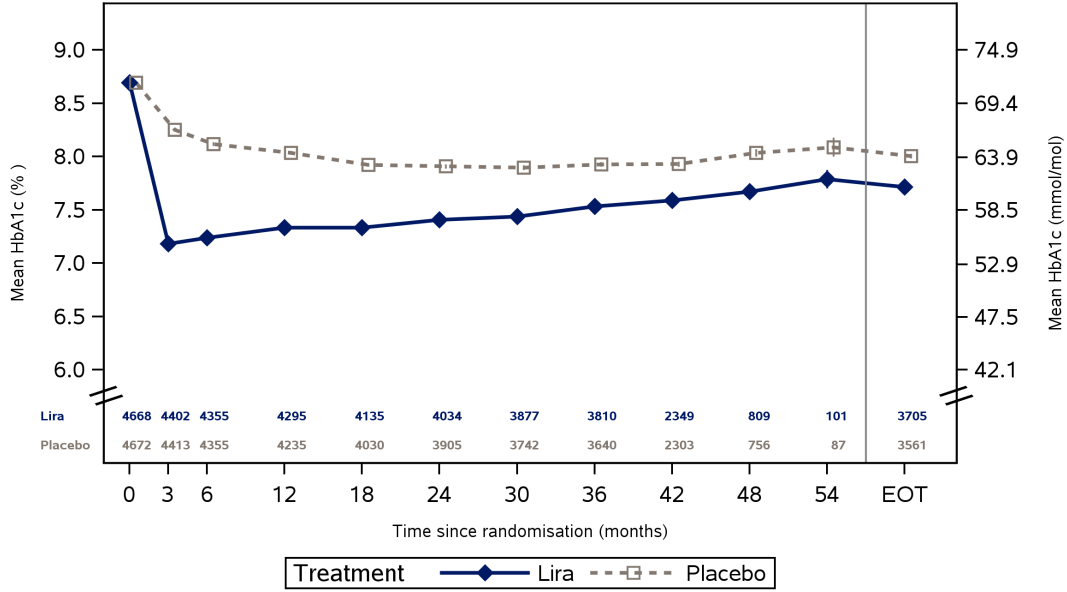
Figure 2: Estimated HbA1c levels over time by treatment group. EOT: End-Of-Trial visit (time varies by subject).

Table 2: Results from 3 simple mediation analyses

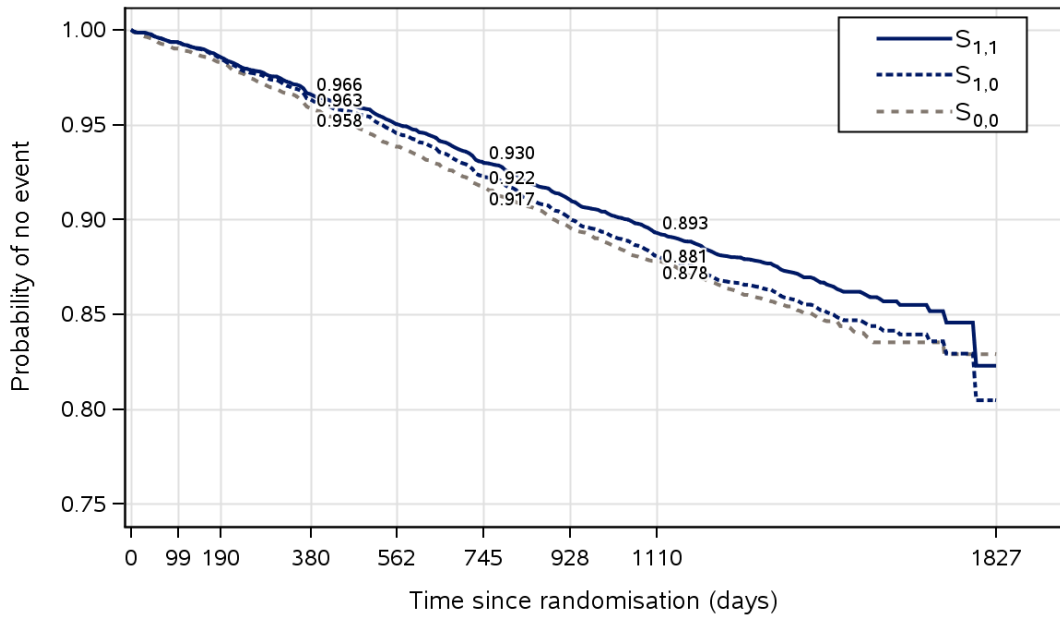| Model	Hazard ratio	95% CI	Mediated proportion
MACE primary analysis	0.87	(0.78; 0.97)	
6-month HbA1c change from baseline as time-fixed covariate	0.92	(0.81; 1.04)	0.40
HbA1c change from baseline as time-dependent covariate	0.88	(0.79; 0.99)	0.08
Updated mean of HbA1c as time-dependent covariate	0.92	(0.82; 1.03)	0.40

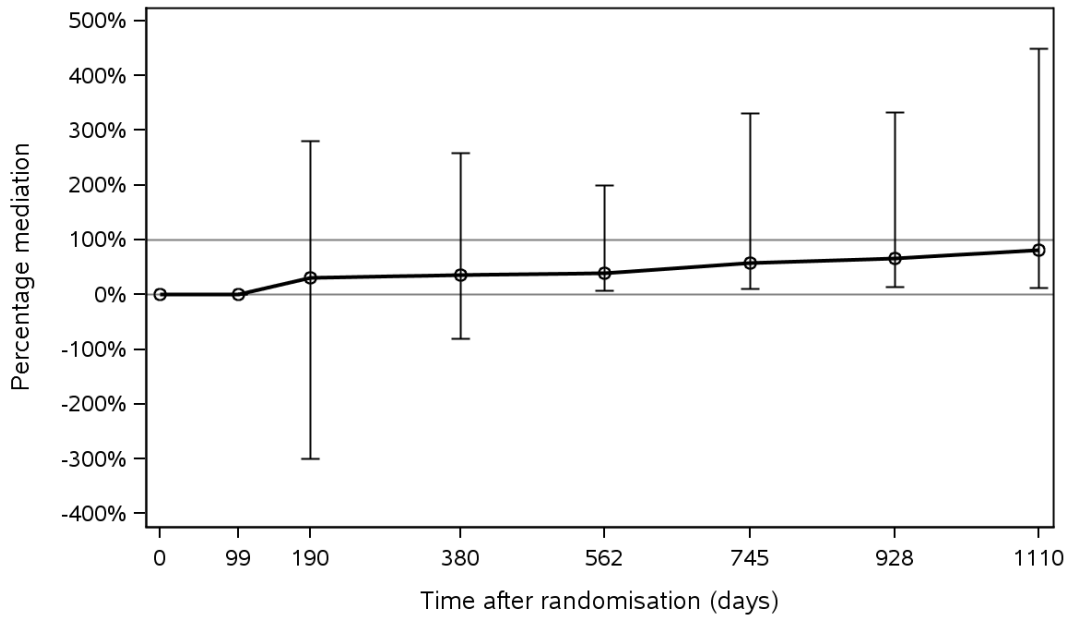Figure 3: Estimated probabilities $S_{1,0}(t), S_{1,1}(t)$ and $S_{0,0}(t)$.

Figure 4: Mediated proportions for time to first MACE with longitudinal HbA1c levels as mediators.

33