Forecasting Causal Effects of Interventions versus Predicting Future Outcomes

Christian Gische, Stephen G. West & Manuel C. Voelkle

View supplementary material 🗗

Published online: 08 Sep 2020.

Submit your article to this journal 🗗

Article views: 1684

View related articles 🗗

View Crossmark data 🗗

Routledge
Taylor & Francis Group

TEACHER'S CORNER

🔓 OPEN ACCESS | Check for updates

Forecasting Causal Effects of Interventions versus Predicting Future Outcomes

Christian Gische [1], Stephen G. West[2], and Manuel C. Voelkle [1,3]

[1]Humboldt-Universität zu Berlin, Germany; [2]Arizona State University, Arizona; [3]Max Planck Institute for Human Development, Germany

ABSTRACT

The present article provides a didactic presentation and extension of selected features of Pearl's DAG-based approach to causal inference for researchers familiar with structural equation modeling. We illustrate key concepts using a cross-lagged panel design. We distinguish between (a) forecasts of the value of an outcome variable after an intervention and (b) predictions of future values of an outcome variable. We consider the mean level and variance of the outcome variable as well as the probability that the outcome will fall within an acceptable range. We extend this basic approach to include additive random effects, allowing us to distinguish between average effects of interventions and person-specific effects of interventions. We derive optimal person-specific treatment levels and show that optimal treatment levels may differ across individuals. We present worked examples using simulated data based on the results of a prior empirical study of the relationship between blood insulin and glucose levels.

Introduction

Psychologists have long distinguished between experimental designs in which participants are randomly assigned to an active treatment or a control treatment and passive observational designs (i.e., non-experimental designs without a manipulation) in which the responses of participants are simply observed (e.g., Cronbach, 1957, 1975). Given randomization and manipulation of the treatment conditions, experimental designs permit strong causal inferences when their assumptions are met (Shadish et al., 2002; West & Thoemmes, 2010). A traditional view is that longitudinal passive observational designs only permit prediction of participants' scores on the criterion variable at future time points and do not allow for causal conclusions.

In this paper, we present an approach that shows that under certain conditions longitudinal passive observational designs contain information that allows researchers to forecast the causal effect of a hypothetical treatment, even if that treatment has not yet been carried out. We consider the commonly used cross-lagged panel design in which two variables X and Y are each measured at a set of equally spaced measurement waves. Throughout we make a clear distinction between (a) the forecast of the causal effect of a (hypothetical) intervention on an outcome at a future time point and (b) the prediction of the individual's outcome at a future time point. The machinery for making forecasts of causal effects of interventions is provided by Judea Pearl's (2009) graphical approach to causal inference (see also Elwert, 2013). We provide a brief didactic presentation and illustration of the relevant key ideas of Pearl's approach, showing the links to ideas from traditional path analysis familiar to readers of this journal.

Using this approach, we illustrate how to forecast the effect of an intervention on the level of an outcome variable at a future time point. We also show how to calculate the variance of the forecasted value and the probability that the outcome variable attains a value within a predefined acceptable range. We initially assume a population of homogeneous individuals and later extend the approach to consider potential individual differences in the mean level of each person's time series data (unobserved heterogeneity). To facilitate a clear illustration of these ideas, we use as a running example a simple system with two variables measured at $T = 4$ equally spaced time points. We assume that the data are generated in a dynamic system in which the value of X at time t, X_t, influences the value of Y at time $t + 1$, Y_{t+1}, and likewise Y_t influences X_{t+1}. The system also includes autoregressive effects in which X_t influences X_{t+1} and Y_t influences Y_{t+1}. In our illustration X_t represents blood insulin levels in non-diabetic adults at each of four equally spaced measurement waves measured in micro international units per milliliter (mcIU/ml). Blood glucose (Y_t) is also measured at each measurement wave in milligrams per deciliter (mg/dl). For ease of presentation and without loss of generality, we use variables centered at the individual mean for each person.

The causal relationships in the system postulated above can be plotted using causal diagrams (see Figure 1). Causal diagrams are well-defined mathematical objects that consist of nodes and edges. Pearl's approach focuses on causal diagrams without feedback or causal loops, that is, without directed paths that traverse the same variable more than once. These causal diagrams are termed *directed acyclic graphs* (DAG) and represent the researcher's theory and prior (empirical) knowledge about the causal relationships
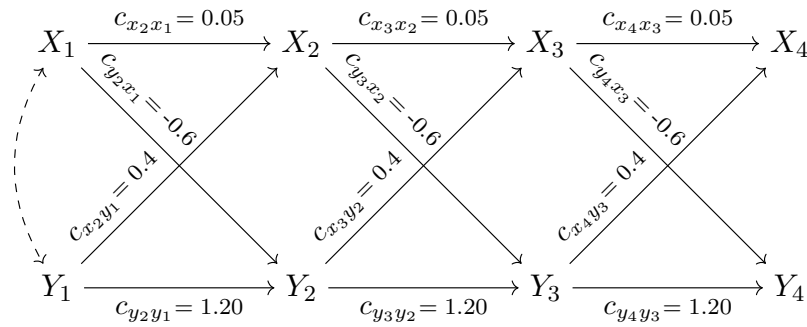
Figure 1. The causal diagram (DAG) of the causal model for the four-wave passive observational study is displayed. The dashed double-headed arrow (bidirected edge) represents a correlation between X_1 on Y_1 due to an unobserved common cause. Traditionally, in DAGs the disturbances denoted by ε_{x_t}, ε_{y_t}, $t = 1, \ldots, 4$ which represent other unmeasured influences on each variable are not represented. Solid single-headed arrows (directed edges) are labeled with path coefficients that quantify direct causal effects. For example, the cross-lagged coefficient $c_{x_2y_1}$ represents the direct effect of Y_1 on X_2. The autoregressive coefficient $c_{x_4x_3}$ represents the direct effect of X_3 on X_4. The coefficients $c_{y_{t+1}x_t}$ and $c_{y_{t+1}y_t}$, $t \geq 1$, are defined analogously.

between the variables.[1] A key advantage of representing the hypothesized causal process as a DAG is that it allows the researcher to analyze the causal system using mathematical tools from the fields of graph theory and graphical modeling (Cowell et al., 1999; Jordan, 1998; Koller & Friedman, 2009; Pearl, 2009). These mathematical tools allow researchers to establish whether specific causal effects can be theoretically estimated given observable data (causal identification). They also help researchers to recognize the key variables needed for covariate adjustment and to find the testable implications of the causal system. Some of these tools are implemented in the user-friendly program DAGitty (Textor et al., 2016). In linear additive models, the insights provided by these mathematical tools have substantial overlap with the insights available from the tracing rules (Loehlin & Beaujean, 2017; Wright, 1934) and covariance algebra (Bollen, 1989) traditionally used in structural equation modeling. Unlike the tracing rules, the mathematical tools from graph theory can be applied to nonlinear and nonparametric models.

We initially consider a dynamic system representing a single person from a population of *homogenous* individuals. In the section on "Between-Person Heterogeneity" we extend the approach to consider potential individual differences in the mean level (unobserved heterogeneity). To keep the illustration simple we further assume that the data generating mechanism is linear and can be represented by the following set of linear structural equations:[2] $t = 1, 2, \ldots, T = 4$

$$X_{t+1} = c_{x_{t+1}x_t}X_t + c_{x_{t+1}y_t}Y_t + \varepsilon_{x_{t+1}} \tag{1a}$$

$$Y_{t+1} = c_{y_{t+1}x_t}X_t + c_{y_{t+1}y_t}Y_t + \varepsilon_{y_{t+1}} \tag{1b}$$

Error terms (residuals, disturbances) are denoted by ε and a subscript indicating the corresponding variable. For example, $\varepsilon_{x_{t+1}}$ denotes the error term corresponding to variable X_{t+1}. The error term $\varepsilon_{x_{t+1}}$ captures those factors that determine the value of X_{t+1} that are not explicitly included in the model (omitted variables).

Consider the DAG in Figure 1 that represents the causal relationships for our simple two variable, four-wave illustration. Each solid single-headed arrow represents a direct causal relationship in this model. For simplicity of presentation, we will assume time-stable structural coefficients throughout the article. That is, we will assume that the autoregressive coefficients $c_{x_{t+1}x_t}$, $t \geq 1$, for the X-series and the autoregressive coefficients $c_{y_{t+1}y_t}$, $t \geq 1$, for the Y-series are constant over time. Likewise, we will assume that the cross-lagged coefficients from X at time t to Y at time $t + 1$, denoted by $c_{y_{t+1}x_t}$, $t \geq 1$, and the cross-lagged coefficients from Y_t to X_{t+1}, denoted by $c_{x_{t+1}y_t}$, $t \geq 1$, are constant over time. In addition, we will assume that the variances of the disturbances corresponding to the X-series, denoted by ψ_{x_t,x_t}, $t \geq 2$, and those corresponding to the Y-series, denoted by ψ_{y_t,y_t}, $t \geq 2$, are constant over time.

Traditionally, error terms are not explicitly drawn in a DAG and the following conventions are used: (a) A bidirected edge drawn between two variables in the DAG indicates that the corresponding *error terms* covary due to unobserved confounding. For example, we assume an unobserved common cause of X_1 and Y_1 that implies a covariance between the corresponding error terms ε_{x_1} and ε_{y_1}, denoted as $\psi_{x_1,y_1} = \text{COV}(\varepsilon_{x_1}, \varepsilon_{y_1})$ (for interested readers, a detailed discussion of the specification of the first measurement occasion is provided in the Appendix on Initial Variables). (b) The absence of a bidirected edge between two variables reflects the assumption that there is no unobserved common cause. For example, we assume that there is no unobserved confounder for the $X_2 \rightarrow Y_3$ relationship. Throughout this article, variances and covariances of disturbances are denoted by ψ and a subscript. For example, ψ_{x_t,x_t} denotes the variance of the disturbance ε_{x_t}, that is, $\psi_{x_t,x_t} = \text{V}(\varepsilon_{x_t})$.

For our illustration, we simulate data from a known data generating process. Numerical values have been assigned to each parameter in the DAG based on the values from a panel data study by Ito et al. (1998). These numerical values are displayed in Tables 1 and 5 and will be used to illustrate numerical calculations of causal effects later in the paper (see

---

[1]Sometimes the class of acyclical directed *mixed* graphs is used. Mixed refers to the fact that graphs in this class contain directed edges as well as bidirected edges (Kang & Tian, 2009; Richardson, 2003).
[2]The complete set of model equations and distributional assumptions are provided in the online supplementary material.

Table 1. Numeric values of parameters for the homogeneous population.

structural coefficients				variance-covariance parameters				
c_{xx}	c_{xy}	c_{yx}	c_{yy}	$\psi_{x_1 x_1}$	$\psi_{y_1 y_1}$	$\psi_{x_1 y_1}$	ψ_{xx}	ψ_{yy}
0.05	0.40	−0.60	1.20	131.76	632.94	254.12	20.00	40.00

Note. The numeric values of the structural coefficients correspond to a stationary and ergodic bivariate time series for the insulin-glucose dynamics displayed in Figures 1 and 2. The variances ($\psi_{x_1 x_1}, \psi_{y_1 y_1}$) and covariance ($\psi_{x_1 y_1}$) of the initial variables correspond to the long-term equilibrium values of the insulin-glucose process.

online supplementary material for a detailed account of the data generation). Note that these numerical values yield a stable and covariance stationary system (Hamilton, 1994).[3] Based on these values we obtain a standard deviation of blood insulin of $\text{SD}(X_t) = 11.48$ mcIU/ml and a standard deviation of blood glucose of $\text{SD}(Y_t) = 25.16$ mg/dl, $t \geq 1$.

Figure 2 modifies Figure 1 by introducing a specific form of manipulation introduced by Pearl. Historically in psychology, we have thought of a manipulation as adding (subtracting) a constant to the person's current level on X. Thus, if a dose of insulin were given, the person's insulin level would be expected to experience a fixed increase to a level that reflects both the dose received and the person's baseline insulin level prior to the manipulation. In contrast, Pearl (2009) proposed a manipulation that fixes a person's level on X to a constant value x. The latter type of manipulation is formalized via the do-operator and has proven to be very useful for theoretically defining and quantifying causal effects. For example, the do-operator applied to variable X_t is denoted by $do(X_t = x_t)$ or equivalently abbreviated as $do(x_t)$. By definition, the $do(x_t)$-operator has the following properties: It (a) fixes the level of X_t at x_t for each unit in the population, where x_t is referred to as the interventional level. The interventional level x_t is (b) not related to any causally prior variable. Finally, the $do(x_t)$-operator (c) leaves all other relationships in the dynamic system unchanged.

Part (a) reflects the manipulation in a hypothetical experiment in which the level of the treatment variable X_t is set by an experimenter. By fixing the level of the variable, each unit will have the identical value of X_t following the $do(x_t)$-manipulation and X_t will have 0 variance under the intervention. In other words, the $do(x_t)$-operation reflects a hypothetical experiment in which the treatment $X_t = x_t$ is applied uniformly over the whole population. Part (b) of the definition of the $do(x_t)$-operator reproduces the feature of randomized experiments that all prior covariates are independent of the treatment variable. Part (c) of the definition of the $do(x_t)$-operator, an assumption formally known as autonomy or modularity (Pearl, 2009; Peters et al., 2017; Spirtes et al., 2000), implies that the intervention $do(x_t)$ only alters the value of X_t and does not change the relationships between other measured variables not involved in the manipulation.

Figure 2 illustrates the do-operator applied to insulin levels at time $t = 2$, denoted as $do(x_2)$. Here, we imagine that each person's insulin level is set to the interventional value $X_2 = 11.48$ mcIU/ml, that is, one standard deviation above the person's mean level (which is 0 for each person after mean-centering in a homogeneous population). The causal links of X_2 to its direct predecessors in the causal chain (here: X_1 and Y_1) have been removed, but all other causal relationships in the model are maintained as before the $do(x_2)$-operation.

Up to this point, we have discussed the data generating mechanism but have not yet introduced the statistical model we will use for data analysis. Since we use simulated data the choice of a statistical model is straightforward: We chose a correctly specified statistical model, that is, one that exactly captures the data generating process described above. In our example, we will use a bivariate linear cross-lagged panel model with four measurement waves. This model can be equivalently understood as a linear structural equation model. Thus, it can be represented by a causal path diagram as commonly used to represent linear SEM (Bauer, 2003; Curran, 2003). These causal path diagrams are similar to DAGs, as long as we restrict our attention to linear models without feedback or causal loops. An important difference, however, is that a dashed double-headed arrow in a path diagram represents a covariance between variables (or disturbances) that is not further analyzed. In the
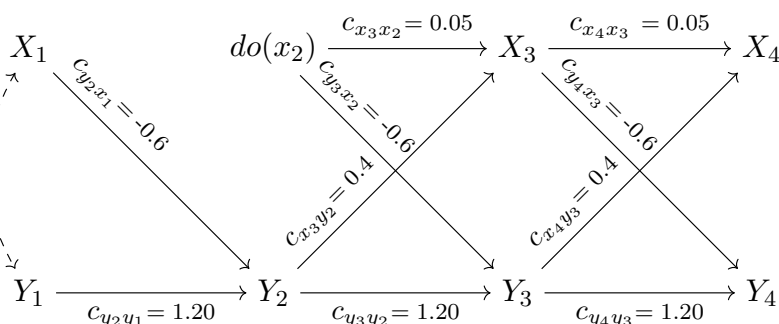

Figure 2. The causal diagram (DAG) displayed in Figure 1 is modified to account for the intervention $do(x_2)$. The variable X_2 has been replaced by $do(x_2)$ and the solid single-headed arrows entering X_2 have been removed, indicating that the interventional level $do(x_2)$ does not depend on X_1 or Y_1 but is set by an experimenter (in our illustrative example we choose $do(X_2 = 11.48)$). Note that all other variables and directed edges (causal arrows) remain unchanged as compared to the situation without the intervention $do(x_2)$ as depicted in Figure 1 reflecting the assumption of modularity.

---

[3]The two assumptions of homogeneous individuals and stationarity over time allow us to make statements about either a single individual over time or a population of homogeneous individuals. Given these two assumptions, the results of between subjects and within subjects analyses of the data are theoretically expected to be identical (Molenaar, 2004). Complexity is introduced if the parameters of the model are allowed to vary across individuals, that is, a person is drawn from a population of heterogeneous individuals (see section on "Between-Person Heterogeneity").

context of DAG-based causal inference, a dashed double-headed arrow represents a covariance that stems from an unobserved confounder. Throughout this paper, we draw DAGs and therefore double-headed arrows always indicate the existence of an unobserved confounder. From a statistical point of view, the linear bivariate cross-lagged panel model that captures the DAG depicted in Figure 1 is globally identified, that is, all parameters can theoretically be estimated uniquely from observational data (Bollen & Curran, 2004; Hsiao, 2014).

Throughout this paper, we know with certainty that the causal assumptions encoded in the DAG in Figure 1 are correct since we use simulated data. Note that all our data stems from the DAG depicted in Figure 1, that is, we have not actually performed an experiment. This *non-experimental* evidence is combined with our assumptions about the data generating mechanism to forecast the effect of the (hypothetical) intervention (see Figure 2). In practical applications, however, one usually does not know the true data generating mechanism and causal assumptions postulated by researchers might be wrong. We discuss statistical tests of model assumptions and methods to analyze the sensitivity of causal conclusions with respect to violations of assumptions in the section on "Current Limitations and Future Directions."

Given the important distinction between the system following the $do(x_2)$-operation depicted in Figure 2 and the passive observational system in the absence of any intervention depicted in Figure 1, we now explore the implications of these different situations for (a) forecasting the treatment effect of X_2 on Y_3 and (b) predicting the value of Y_3 based on X_2 in the absence of interventions.

Distinguishing between forecasting effects of interventions and predicting future outcomes

The effect of a treatment is always defined relative to another active or control treatment (Holland, 1986). The *do*-operator allows us to represent a specific type of treatment effect by comparing the distribution of an outcome variable under the intervention $do(x_t)$ (representing the treatment) to that of a second intervention $do(x'_t)$ (representing the control). The upper panel of Figure 3 displays the distribution of Y_3 given the interventions $do(X_2 = 11.48)$ (treatment; solid line) and $do(X_2 = 0)$ (control; dotted line).

We define the distribution of the outcome value Y_{t+k} in the population given the treatment $do(x_t)$ as the interventional distribution, denoted by $P(Y_{t+k}|do(x_t))$, where k is the number of waves after the intervention when the outcome is measured.[4] In this paper, we restrict our attention to the interventional distribution $k = 1$ wave after the treatment has been applied. To compare the outcome variable Y_{t+1} given two distinct
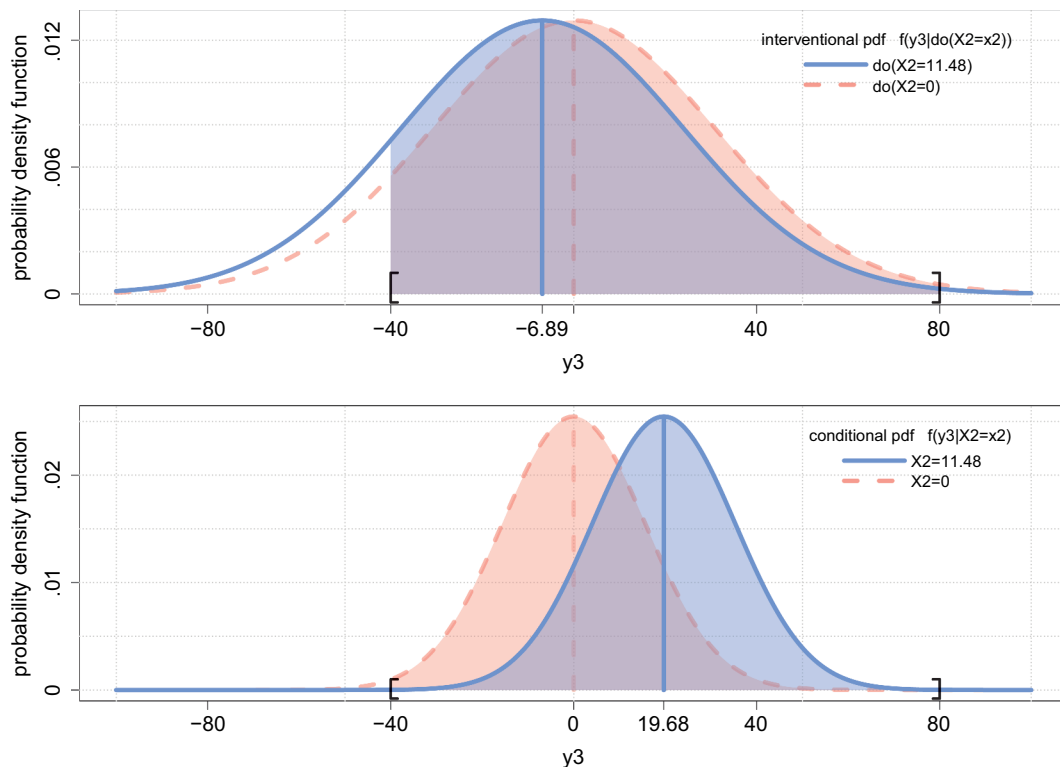

Figure 3. The upper panel displays the probability density functions (pdf) of the interventional distribution $P(Y_3|do(x_2))$ for two distinct treatment levels, namely $do(X_2 = 11.48)$ (solid line) and $do(X_2 = 0)$ (dashed line). Means are depicted as solid and dashed vertical line segments. The lower panel displays the pdf of the conditional distribution $P(Y_3|X_2 = x_2)$ for two distinct observed levels of blood insulin at time $t = 2$, namely $X_2 = 11.48$ (solid line) and $X_2 = 0$ (dashed line). The marked interval on the horizontal axes ($[-40, 80]$, person mean-centered metric) represents the range of acceptable mean-centered blood glucose levels. The shaded areas under the curves (designated by square brackets) represent the probability that blood glucose level at time $t = 3$ falls in the acceptable range.

---

[4]In its most general form, the interventional distribution is the multivariate joint distribution of *all* variables in the system given the intervention $do(x_t)$ is applied. For example, the joint interventional distribution for the system depicted in Figure 2 is given by $P(X_1, Y_1, X_2, Y_2, X_3, Y_3, X_4, Y_4|do(x_2))$. In this paper, we restrict our attention to the *marginal* interventional distribution of a single outcome variable, for example, $P(Y_3|do(x_2))$.

interventions $do(x_t)$ and $do(x'_t)$, one has to compare the two random variables $Y_{t+1}|do(x_t)$ and $Y_{t+1}|do(x'_t)$. This comparison can, for example, be done visually by plotting the probability density functions (pdf) of the two interventional distributions $P(Y_{t+1}|do(x_t))$ and $P(Y_{t+1}|do(x'_t))$, as displayed in the upper panel of Figure 3. In our illustration, we compare the values of blood glucose at time $t = 3$ given two distinct interventions on blood insulin at $t = 2$, namely $do(X_2 = 11.48)$ and $do(X_2 = 0)$. The latter intervention sets the value of blood insulin to the mean value ($= 0$ in mean-centered metric), the former intervention to one standard deviation above the mean. The upper panel of Figure 3 depicts the probability density functions of the corresponding interventional distributions $P(Y_3|do(X_2 = 0)$ (dashed line) and $P(Y_3|do(X_2 = 11.48)$ (solid line).

Another common way to compare two distributions is to examine differences in their means. The difference between the interventional means $E(Y_3|do(x_t)) - E(Y_3|do(x'_t))$ is called the average treatment effect (ATE) of treatment $do(x_t)$ relative to treatment $do(x'_t)$ on Y_{t+1}, abbreviated as the ATE of X_t on Y_{t+1} (Pearl, 2009). In our illustration, we analyze the ATE of treatment $do(X_2 = 11.48)$ relative to treatment $do(X_2 = 0)$ on Y_{t+1}, denoted as $E(Y_3|do(X_2 = 11.48)) - E(Y_3|do(X_2 = 0))$. The interventional means are depicted in the upper panel of Figure 3 as the solid and the dashed vertical line segments.

The comparison of two distributions is not restricted to a comparison of their location (means). Another important feature of the interventional distribution is its variance. The variances of the outcome distributions under the distinct treatments $do(x_t)$ and $do(x'_t)$ correspond to the *widths* of the bell-shaped curves in the upper panel of Figure 3. The two curves in the upper panel have the same widths, that is, the error when forecasting the effect of an intervention on X_2 on the level of Y_3 is the same for $do(X_2 = 11.48)$ and $do(X_2 = 0)$.

Comparisons that go beyond first- and second-order moments (means and variances) may be of importance in a number of research situations involving dynamic systems. For example, Monnier et al. (2017) propose the frequency of hypoglycemia (adversely low levels of blood glucose) as a target quantity that should be minimized during treatment. In the section titled "Interventional Probabilities versus Conditional Probabilities" we will consider the probability that a *do*-operation will lead to glucose levels that fall in an acceptable range of values that is specified a priori.

Throughout this article, we will make a clear distinction between (a) the forecast of the outcome value Y_{t+1} given an intervention $do(x_t)$, and (b) the prediction of the value of Y_{t+1} given the passive observation that variable X takes the value x at time t, denoted by $X_t = x_t$. Returning to our example, in case (b) we would like to predict the value of blood glucose at $t = 3$ based on the passive observation (e.g., self-measured monitoring of blood insulin) that the insulin level is one standard deviation above the mean level at $t = 2$. We will, therefore, use the conditional distribution of Y_3, given $X_2 = 11.48$, denoted by $P(Y_3|X_2 = 11.48)$, to predict the value of blood glucose at $t = 3$. We use the term *predictions* (of future values)

to refer to probabilistic statements that are based on the conditional distribution.

Comparing the two panels of Figure 3 provides a visual representation of the differences between (a) forecasts of the effects of interventions and (b) predictions of future values. Observing $X_2 = 11.48$ yields the conditional distribution $P(Y_3|X_2 = 11.48)$ represented by the solid line in the lower panel, whereas administering the treatment $do(X_2 = 11.48)$ yields the interventional distribution $P(Y_3|do(X_2 = 11.48))$ represented by the solid line in the upper panel. The conditional distribution of the predicted values of blood glucose levels is shifted to the right relative to the interventional distribution of the $do(x_2)$-forecast of blood glucose levels.

Given this initial impression of the differences between (a) the interventional and (b) the conditional distribution, we now systematically analyze the differences in mean levels and variances, followed by the probabilities of blood glucose levels falling in an acceptable range. As part of this systematic comparison, we present formulae for interventional distributions and the moments thereof (see also Gische & Voelkle, 2020). Formulae for conditional distributions and moments thereof are established results for the multivariate normal distribution (e.g., Rao, 1973).

For ease of presentation we will drop the time indices of the structural coefficients in the equations. For example, we will write c_{xx} instead of $c_{x_2x_1}$, $c_{x_3x_2}$, and $c_{x_4x_3}$ for the autoregressive coefficients of the X-series, because these coefficients are assumed to be stable over time. Analogously, we will write c_{yy} for the autoregressive coefficients of the Y-series as well as c_{xy} and c_{yx} for the cross-lagged coefficients. Likewise, we will drop the time indices of the variance parameters for $t \geq 2$. We will write ψ_{xx} instead of $\psi_{x_2x_2}$, $\psi_{x_3x_3}$, and $\psi_{x_4x_4}$ and ψ_{yy} instead of $\psi_{y_2y_2}$, $\psi_{y_3y_3}$, and $\psi_{y_4y_4}$. All numeric quantities are calculated based on the parameter values displayed in Tables 1 and 5.

A complete presentation of all relevant formulae and details regarding the derivation thereof is provided in the online supplementary material together with a justification of the numerical values based on prior empirical results reported by Ito et al. (1998).

Interventional mean versus conditional mean

In this section we systematically compare the mean level of blood glucose at $t = 3$ given (a) a $do(x_2)$-intervention has been applied on insulin levels at $t = 2$ and (b) prior observations on the levels of X_2 are available.

We start with a comparison of the baseline scenarios in which (a) no *do*-treatment is applied (denoted by $do(\emptyset)$), and (b) no prior observations are available (denoted by the empty set $\emptyset$). In both situations, the best guess of the value of Y_3 is zero since the data are mean-centered within each person (see row 1 of Table 2). Put more formally, in the baseline scenarios the interventional mean $E(Y_3|do(\emptyset))$ and the conditional mean $E(Y_3|\emptyset)$ are both equal to the unconditional mean, that is, $E(Y_3|do(\emptyset)) = E(Y_3|\emptyset) = E(Y_3) = 0$.

We now turn to a comparison of the scenarios where in case (a) the level of blood insulin has been set to $do(X_2 = 11.48)$ and in case (b) the value $X_2 = 11.48$ has been observed (see row 2 of Table 2).

Table 2. Mean forecast of effects of interventions vs. mean prediction of future values.

(a) intervention			(b) observation		
intervention	interventional mean	numeric	observation	conditional mean	numeric
$do(\emptyset)$	$E(Y_3\|do(\emptyset))$	0	$\emptyset$	$E(Y_3\|\emptyset)$	0
$do(x_2)$	$E(Y_3\|do(x_2))$	$-0.6x_2$	$X_2 = x_2$	$E(Y_3\|X_2 = x_2)$	$1.71x_2$
$do(x_2), do(y_2)$	$E(Y_3\|do(x_2), do(y_2))$	$-0.6x_2 + 1.2y_2$	$X_2 = x_2, Y_2 = y_2$	$E(Y_3\|X_2 = x_2, Y_2 = y_2)$	$-0.6x_2 + 1.2y_2$

Note. The left part of Table 2 contains three different interventional means. In row one, no intervention is applied ($do(\emptyset)$), in row two an intervention on blood insulin at $t = 2$ is applied ($do(x_2)$), and in row three interventions on both blood insulin and blood glucose at $t = 2$ are applied ($do(x_2), do(y_2)$). The right part of Table 2 contains three different conditional means that are used for predictions of future values. In row one, we condition on the empty set $\emptyset$ (no prior observations available), in row two on the observation $X_2 = x_2$, and in row three on the observations $X_2 = x_2$ and $Y_2 = y_2$.

In case (a), the intervention $do(X_2 = 11.48)$ is performed, which sets the value of blood insulin to one SD above the mean level at time $t = 2$. Due to the intervention both paths entering into X_2 are removed (see Figure 2). The formula for the interventional mean and its numeric evaluation are given by:

$$E(Y_3|do(x_2)) = c_{yx}x_2$$
$$E(Y_3|do(X_2 = 11.48)) = -0.6 \cdot 11.48 = -6.89 \quad (2)$$

Thus, the mean forecast of Y_3 given the intervention $do(X_2 = 11.48)$ is -6.89 mg/dl (see also row 2, column 3 of Table 2). Recall that the ATE of a treatment is always defined relatively to another treatment. Here we analyze the ATE of treatment $do(X_2 = 11.48)$ relative to the treatment $do(X_2 = 0)$. Thus, the resulting ATE is equal to $E(Y_3|do(X_2 = 11.48)) - E(Y_3|do(X_2 = 0)) = -6.89 - 0 = -6.89$. In other words, actively changing the level of blood insulin at $t = 2$ from 0 to 11.48 ($=$ one SD above the mean) results in a difference of -6.89 mg/dL in the forecasted blood glucose level at time $t = 3$.

In case (b) observing a blood insulin level of one SD above the person's mean at time $t = 2$ leads to a predicted value of $1.71 \cdot 11.48 = 19.63$ mg/dl of blood glucose at time $t = 3$ (see row 2, column 6 of Table 2). The coefficient 1.71 is the population regression coefficient in a simple linear regression of Y_3 on X_2 and is calculated according to the following formula:

$$E(Y_3|X_2 = x_2) = \frac{COV(X_2, Y_3)}{V(X_2)}x_2 \quad (3)$$

$$= \left(c_{yx} + \frac{c_{xx}c_{yx}c_{yy}\psi_{x_1x_1} + c_{xy}c_{yy}^2\psi_{y_1y_1} + (c_{xy}c_{yx}c_{yy} + c_{xx}c_{yy}^2)\psi_{x_1y_1}}{c_{xx}^2\psi_{x_1x_1} + c_{xy}^2\psi_{y_1y_1} + 2c_{xx}c_{xy}\psi_{x_1y_1} + \psi_{xx}}\right)x_2$$

$$= (-0.6 + 2.31)x_2 = 1.71x_2$$

The last line contains an numeric evaluation based on the parameter values displayed in Table 1. The population regression coefficient represents the sum of the direct causal effect and the spurious effects of X_2 on Y_3. These spurious effects correspond to backdoor paths from X_2 to Y_3 in Pearl's terminology. One such backdoor path is $X_2 \leftarrow X_1 \rightarrow Y_2 \rightarrow Y_3$; another backdoor path is from $X_2 \leftarrow Y_1 \rightarrow Y_2 \rightarrow Y_3$ (see Figure 1). Observing a mean blood insulin level ($=0$) at time $t = 2$ leads to a predicted value of $1.71 \cdot 0 = 0$ mg/dl of blood glucose at time $t = 3$. Thus, based on traditional methods for prediction (e.g., linear regression), we would obtain

a difference of 19.63 mg/dL between the predicted values of Y_3 given we observe a change in X_2 from 0 to 11.48.

In summary, in case (a) changing the $do(x_2)$ treatment level by plus one SD yields a -6.89 mg/dL decrease in the forecasted value of blood glucose at time $t = 3$. In case (b) an observed one SD increase in blood insulin at wave 2 is predicted to produce a 19.63 mg/dL increase in blood glucose at time $t = 3$. Both findings provide useful information that need to be used for distinct purposes. In case (a), a physician would correctly forecast a negative value ($-0.6 \cdot 11.48 = -6.89$ mg/dl) of blood glucose at time $t = 3$ after administering the dose $do(X_2 = 11.48)$ of insulin at time $t = 2$. In case (b), a patient obtaining the insulin reading $X_2 = 11.48$ at time $t = 2$ would correctly predict a positive level ($1.71 \cdot 11.48 = 19.63$ mg/dl) of blood glucose at time $t = 3$.

Incorrect conclusions only arise if the algebraic machinery of conditional expectations (case (b)) is used to forecast the effects of interventions (case (a)) – or the other way around – the interventional mean based on the do-operator (case (a)) is used to predict future values of Y in the absence of interventions (case (b)).

We have seen that interventional means (forecasts) and conditional expectations (predictions) are different quantities. The latter is a purely statistical quantity that can always be estimated based on observational data in a correctly specified model. The former is a causal quantity that might *not* be identified based on observational data even if the model is correctly specified. Pearl (2009) provides DAG-based criteria for the identification of specific causal effects, some of which are implemented in the computer program DAGitty (Textor et al., 2016). For example, to identify the ATE of X_2 on Y_3 in the present example, a graphical rule called the backdoor criterion can be used. The backdoor criterion states that observing a set of variables which blocks all backdoor paths from X_2 to Y_3 provides sufficient information to correctly calculate the ATE based on non-experimental data. In the present example, the minimal sufficient backdoor adjustment set for the ATE of X_2 on Y_3 is $\{Y_2\}$.[5] Thus, if Y_2 is held constant or statistically adjusted for, all backdoor paths are blocked.

Backdoor adjustment in linear models can for example, be obtained by linear regression. Thus, the ATE X_2 on Y_3 can be estimated using a linear regression of Y_3 on X_2 *and* Y_2 (Pearl, 2009, Theorem 5.3.2). The resulting coefficient of X_2 in the proposed regression of Y_3 on X_2 *and* Y_2 is equal to -0.6 (see row 3, column 6 of Table 2) which equals the coefficient in the interventional mean expression (see row 2, column 3 of Table 2)

---

[5]An introduction to DAG-based causal identification written in a didactic style of presentation is provided by Pearl et al. (2016).

Table 2). Thus, if the backdoor criterion is satisfied in a linear model, the interventional mean (causal quantity) can be estimated from observational data using linear regression.

Interventional variance versus conditional variance

This section studies the interventional variance of the outcome variable following a *do*-operation, a feature of the interventional distribution that has received less attention than the interventional mean. Using our running example, we focus on the variance of blood glucose at $t = 3$.

In the baseline scenarios in which (a) no *do*-treatment is applied (denoted by $do(\emptyset)$), and (b) no prior observations are available (denoted by the empty set $\emptyset$), the interventional variance $V(Y_3|do(\emptyset))$ and the conditional variance $V(Y_3|\emptyset)$ are both equal to the unconditional variance $V(Y_3)$. All quantities take on the value 632.93, as displayed in row 1 of Table 3.

In the left part of Table 3 (columns below (a) intervention) the interventional variances are displayed. For example, the interventional variance $V(Y_3|do(x_2))$ can be calculated according to the following formula (see online supplementary material for derivation):

$$V(Y_3|do(x_2)) = c_{yx}^2 c_{yy}^2 \psi_{x_1 x_1} + c_{yy}^4 \psi_{y_1 y_1} + 2 c_{yx} c_{yy}^3 \psi_{x_1 y_1} + (1 + c_{yy}^2) \psi_{yy} \qquad (4)$$

Note that the interventional level x_2 does *not* appear on the right-hand side of Equation (4). Thus, the interventional variance (in linear models with normally distributed error terms) does not depend on the exact level x_2 of the *do*-treatment that is applied. This can also be seen in the upper panel of Figure 3: The *shape* of the interventional distributions is the same for both treatment levels $do(X_2 = 0)$ (dashed line) and $do(X_2 = 11.48)$ (solid line). If a patient is treated at time $t = 2$ and we change the treatment dose (e.g., set the insulin level to 11.48 instead of 0), only the location of the interventional distribution changes and the forecast variance (*width* of the bell-shaped curves) is not affected by the specific level to which blood insulin is set by the experimenter.

In contrast, the interventional variance does change as we introduce additional interventions on different variables in the model. This can be verified by comparing the value in row 1 with the value of row 2 in column 3 of Table 3. These values illustrate the case where we administer the treatment $do(x_2)$ to a formerly untreated ($do(\emptyset)$) person: the forecast variance *increases* from 632.93 to 951.43. An increase in variance might appear counterintuitive at first glance. However, our illustration considers the dynamic interplay of blood insulin and blood glucose in a population of *healthy* (non-diabetic) persons. For such a population the glucose and insulin levels are governed by a self-regulating dynamic system. In our example the external intervention temporarily eliminates the ability of the system to self-regulate. Consequently, the system becomes temporarily destabilized and the variance increases.

In case (b), the conditional variances used for prediction decrease $(632.93 > 245.71 > 40.00)$ as we successively move from row 1 to row 3 in column 6 of Table 3. This result is intuitive: As we move from one row to the row below in the right part of Table 3, more observational information becomes available. Increasing the amount of available observational information (that stems from the same data generating mechanism) increases the precision of prediction as reflected in the sequence of decreasing conditional variances. The conditional variance $V(Y_3|X_2 = x_2)$ is calculated according to the following general formula (Rao, 1973):

$$V(Y_3|X_2 = x_2) = V(Y_3) - \frac{COV(X_2, Y_3)^2}{V(X_2)} \qquad (5)$$

Note that x_2 does *not* appear on the right-hand side of Equation (5). In other words, unlike the conditional expectations in Table 2, the conditional variance (in linear models with normally distributed error terms) does not depend on the exact value x_2 of the variable on which we condition. This result is also depicted in panel (b) of Figure 3, where the *shape* of the conditional distributions is the same for both observations $X_2 = 0$ (dashed line) and $X_2 = 11.48$ (solid line). The two curves only differ in location but have the same *width*.

So far, we have seen that the interventional variance and the conditional variance are different quantities. The conditional variance is a purely statistical quantity used to represent prediction error, whereas the interventional variance is a causal quantity used to represent forecast error when assessing the effect of a *do*-type intervention. The interventional variance (causal quantity) is identified since the backdoor criterion introduced in the section titled "Interventional Mean versus Conditional Mean" is not restricted to interventional means but applies to the entire interventional distribution (Pearl, 2009).

Given identification is ensured, the question arises whether there is a simple way to estimate the interventional variance of Y_3 from observational data. In our example, the question is: Can we compute the numerical value $V(Y_3|do(x_2)) = 951.43$ (*causal* quantity) from the numerical values of the *statistical* quantities in column 6 of Table 3? The answer is yes, but the procedure is complex and also requires information from

Table 3. Variance of forecasts (intervention) vs. variance of predictions (future values).

(a) intervention			(b) observation		
intervention	interventional variance	numeric	observation	conditional variance	numeric
$do(\emptyset)$	$V(Y_3|do(\emptyset))$	632.93	$\emptyset$	$V(Y_3|\emptyset)$	632.93
$do(x_2)$	$V(Y_3|do(x_2))$	951.43	$X_2 = x_2$	$V(Y_3|X_2 = x_2)$	245.71
$do(x_2), do(y_2)$	$V(Y_3|do(x_2), do(y_2))$	40.00	$X_2 = x_2, Y_2 = y_2$	$V(Y_3|X_2 = x_2, Y_2 = y_2)$	40.00

Note. The left part of Table 3 contains three different interventional variances used when forecasting the value of an outcome variable. In row one no intervention is applied ($do(\emptyset)$), in row two an intervention on blood insulin at $t = 2$ is applied ($do(x_2)$), and in row three interventions are applied on both blood insulin and blood glucose at $t = 2$ ($do(x_2), do(y_2)$). The right part of Table 3 contains three different conditional variances that are used for predictions of future values. In row one, we condition on the empty set $\emptyset$ (no prior information available), in row two on the observation $X_2 = x_2$, and in row three on the observations $X_2 = x_2$ and $Y_2 = y_2$.

column 6 of Table 2 (Kuroki & Cai, 2007). Unfortunately, the standard backdoor adjustment procedure for linear models (i.e., linear regression) that is used to estimate the interventional mean cannot be used to estimate interventional variances.

This problem can be resolved by using the technique for causal identification and estimation in linearly parameterized causal DAG-models suggested by Gische and Voelkle (2020). The central idea is that a causal quantity defined via the *do*-operator can be expressed as a function of the parameters of the statistical model. This idea can be illustrated for the interventional variance as stated in Equation (4). The expression on the left-hand side of Equation (4) is $V(Y_3|do(x_2))$ which is a causal quantity and contains the *do*-operator. The expression on the right-hand side is a polynomial function of the model parameters c_{yx}, c_{yy}, $\psi_{x_1x_1}, \psi_{y_1y_1}$, $\psi_{x_1y_1}$, ψ_{yy} and does *not* contain the *do*-operator. In our working example all parameters that appear on the right-hand-side of Equation (4) (i.e., c_{yx}, c_{yy}, $\psi_{x_1x_1}, \psi_{y_1y_1}$, $\psi_{x_1y_1}$, ψ_{yy}) are identified in the statistical model and can be estimated from observational data. Therefore, the interventional variance $V(Y_3|do(x_2))$ on the left-hand side of Equation (4) is also identified and estimates can be obtained from observational data.

Interventional probabilities versus conditional probabilities

For some phenomena, it is important that a variable in a dynamic system falls within a predefined acceptable range. In our running example, the physician's goal is that the patient's level of blood glucose will be maintained in an acceptable range (so-called euglycemic range) after a treatment. A patient's glucose level should not exceed a critical level (y^{upper}, upper threshold) where she will experience hyperglycemia. Nor should a patient's glucose level fall below a critical level (y^{lower}, lower threshold) where she will experience hypoglycemia (Bode et al., 2005). The thresholds y^{upper} and y^{lower} are defined a priori based on medical standards. Each condition outside the acceptable range poses a threat to a patient's health and should, therefore, be avoided.

We use the values $y^{upper} = 80$ mg/dl and $y^{lower} = -40$ mg/dl in our example. The original upper (180 mg/dl) and lower (60 mg/dl) thresholds were centered around a global mean of 100 mg/dl (Cryer, 2003; Danaei et al., 2011; Giugliano et al., 1997). Thus, a physician is interested in the event that Y_3 falls in the acceptable range $[-40, 80]$ in the mean-centered metric, given the treatment. We refer to the latter event as *successful treatment* and introduce an indicator variable Y_3^* that is equal to 1 in case of a successful treatment and 0 otherwise. The probability of a successful treatment is denoted by $P(-40 < Y_3 < 80|do(x_2))$ or in brief $P(Y_3^* = 1|do(x_2))$. The shaded area under the solid curve in the upper panel of Figure 3 represents the probability of a successful treatment given the intervention $do(x_2 = 11.48)$.

We now turn to a systematic analysis of interventional probabilities and contrast them with conditional probabilities. In the baseline scenarios in which (a) no *do*-treatment is applied (denoted by $do(\emptyset)$), and (b) no prior observations are available (denoted by the empty set $\emptyset$), the probability that blood glucose levels fall in the acceptable range at $t = 3$ is .94 as displayed in the first row of Table 4.

In case (a), administering $do(X_2 = 11.48)$ at time $t = 2$ yields a probability of blood glucose values within the acceptable range at time $t = 3$ equal to .86 (see row 2 column 3 in Table 4). The interventional probability of treatment success depends on the level of blood insulin that has been administered at $t = 2$ in a nonlinear way that exhibits a unique maximum as depicted by the solid line in Figure 4.[6]

In case (b), where a value of blood insulin equal to one standard deviation above the person's mean level is observed at time $t = 2$, the probability of blood glucose values being within the acceptable range at time $t = 3$ is equal to .9999 (see row 2 column 6 of Table 4). Again, the conditional probability $P(Y_3^* = 1|X_2 = x_2)$ has a nonlinear relationship to the level of blood insulin observed at $t = 2$ which is depicted as the dashed line in Figure 4.

Both curves displayed in Figure 4 have a single unique maximum. In case (a) a physician who must choose an interventional level should apply the treatment $do(X_2 = -33.34)$ to maximize the probability of treatment success (solid vertical line). Recall that we use mean-centered variables with $SD(X_2) = 11.48$. Thus, the treatment $do(X_2 = -33.34)$

Table 4. Interventional probabilities versus conditional probabilities.

	(a) intervention			(b) observation		
intervention	interventional probability	numeric	observation	conditional probability	numeric
$do(\emptyset)$	$P(Y_3^* = 1|do(\emptyset))$	.94	$\emptyset$	$P(Y_3^* = 1|\emptyset)$	.94
$do(x_2)$	$P(Y_3^* = 1|do(x_2))$	.86	$X_2 = x_2$	$P(Y_3^* = 1|X_2 = x_2)$	.9999
$do(x_2), do(y_2)$	$P(Y_3^* = 1|do(x_2), do(y_2))$	≈ 1	$X_2 = x_2, Y_2 = y_2$	$P(Y_3^* = 1|X_2 = x_2, Y_2 = y_2)$	≈ 1

Note. The left part of the Table 4 contains three different interventional probabilities that are used to forecast outcome values after an intervention has been applied. In row one, no intervention is applied ($do(\emptyset)$), in row two an intervention on blood insulin at $t = 2$ is applied ($do(x_2)$) and in row three an intervention on both blood insulin and blood glucose at $t = 2$ is applied ($do(x_2), do(y_2)$). The right part of Table 4 contains three different conditional probabilities that are used for predictions. In row one, we condition on the empty set $\emptyset$ (no prior information available), in row two on the observation $X_2 = x_2$, and in row three on the observations $X_2 = x_2$ and $Y_2 = y_2$. We use the numerical values $x_2 = 11.48$ and $y_2 = 0$ for both interventional levels and observed values. Y_3^* is an indicator variables that is equal to 1 if blood glucose levels fall within the acceptable range $[-40, 80]$ at $t = 3$, and is equal to 0 otherwise.

---

[6]Note that we did *not* set up an additional model for the probability of treatment success. Instead, we analytically derived the functional form of the population quantity $P(Y_3^* = 1|do(x_2))$ for the model depicted in Figure 1 and Figure 2. The complete set of model equations for a homogeneous population and the explicit formula for the interventional probability are provided in the online supplementary material.

corresponds to setting a patient's insulin level to 2.9 standard deviations below the mean level. In case (b) a patient who is passively measuring blood insulin levels at $t = 2$ would hope to measure a value of $x_2 = 11.67$, since observing this value maximizes the conditional probability of blood glucose levels within the acceptable range at $t = 3$ (dashed vertical line).

If a physician were to erroneously use the conditional distribution to optimize an intervention and applied the treatment $do(X_2 = 11.67)$, this non-optimal intervention would result in a 85.5% probability of treatment success. A physician who correctly used the interventional distribution to choose the optimal treatment would apply $do(X_2 = -33.34)$, resulting in a 94.8% probability of treatment success. Thus, the consequence of an incorrect decision yields a 9.3% percentage drop in the probability of treatment success.

In summary, we have seen that conditional probabilities (statistical quantity) and the interventional probabilities (causal quantity) differ. The statistical quantity can always be estimated from observational data in a correctly specified model, whereas the causal quantity needs to be identified. The backdoor criterion ensures causal identification of interventional probabilities in our example model. The exact formula of the interventional probability $P(Y_3^* = 1|do(x_2))$ is complex and the regression-based backdoor adjustment formula for linear models can not be applied to calculate interventional probabilities. Instead, techniques for causal identification and estimation in linearly parameterized causal DAG-models suggested by Gische and Voelkle (2020) can be applied as explicated in more detail at the end of the section titled "Interventional Variance versus Conditional Variance."

Between-person heterogeneity

At the beginning of this article, we assumed that individuals come from a homogeneous population. This assumption will often be unrealistic in practice. A treatment that would be optimal (e.g., the administered dose of insulin maximizes the probability that blood glucose realizes within the acceptable range) for an "average" person is not necessarily optimal for a given patient if the population from which the average was calculated consists of heterogeneous individuals (Holland, 1986; Molenaar, 2004; Sidman, 1960; Voelkle et al., 2018). In this section we incorporate person-specific differences in mean levels into our model and focus on person-specific effects of interventions, that is, we forecast the effect of an intervention for a single individual drawn from a heterogeneous population.

Modeling between-person heterogeneity via random intercepts

We use a panel data design with a fixed number of time points, $T = 4$, and sample size N as presented in the Introduction. Authors from different disciplines have proposed a variety of approaches for including person-specific differences into this framework (e.g., Hsiao, 2014; Usami et al., 2019; Wooldridge, 2010; Zyphur et al., 2019).

Here, we adopt the approach of including the latent variables η_x and η_y as additive random intercepts into the model (see Figure 5). The random intercept η_x is a term that includes all additive time-invariant factors that affect the level of blood insulin of a person; η_y is a term that includes all additive
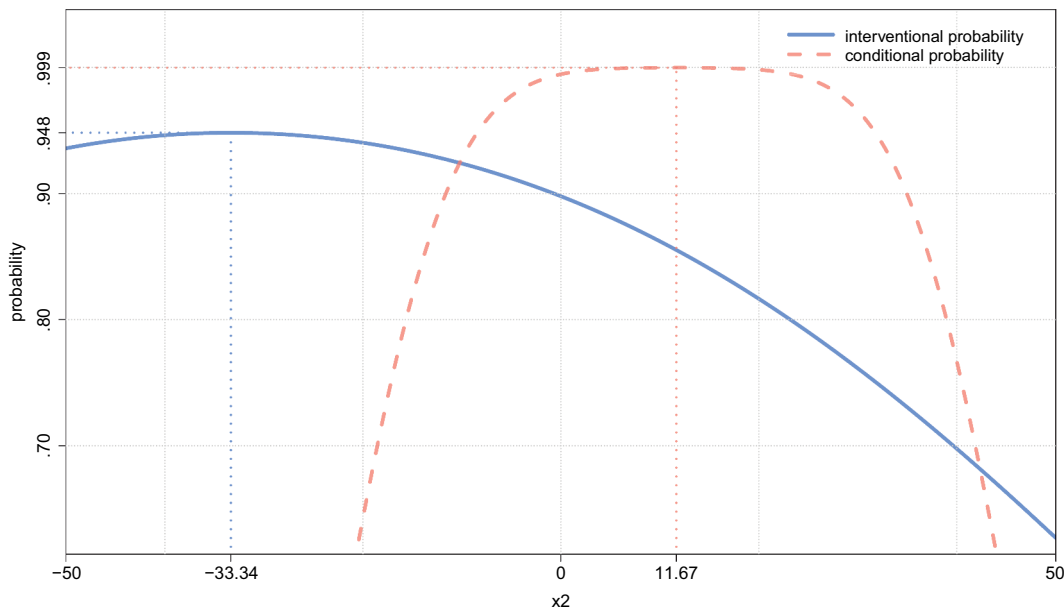

Figure 4. The interventional probability $P(Y_3^* = 1|do(x_2))$ on the vertical axis is depicted as a function of the (mean-centered) interventional level on the horizontal axis (solid line). In addition, the conditional probability $P(Y_3^* = 1|X_2 = x_2)$ on the vertical axis is depicted as a function of the (mean-centered) observed level $X_2 = x_2$ on the horizontal axis (dashed line). $Y_3^* = 1$ whenever blood glucose level at time $t = 3$ falls within the acceptable range. A vertical dotted line is drawn at the interventional level $do(X_2 = -33.34)$ that maximizes the probability of treatment success. Another vertical dotted line is drawn at the observed measurement level $X_2 = 11.67$ that maximizes the prediction probability that blood glucose levels will be within the acceptable range at time $t = 3$ (in the absence of an intervention).

time-invariant factors that affect the level of blood glucose of a person. Thus, random intercepts capture those factors that vary across persons but do not change over time. For example, the random intercept could capture such factors as sex, marital status, and genetic endowment if these factors are constant during the study. Without loss of generality, we assume that random intercepts have zero means, that is, $E(\eta_x) = E(\eta_y) = 0$.

Note that these time-invariant variables (e.g., sex, marital status, and genetic endowment) are neither explicitly modeled nor measured in the study. Instead, the time-invariant variables are captured in the random intercepts (i.e., they are statistically accounted for).[7] This approach gives random intercepts the status of latent variables as indicated by the dashed circles around η_x and η_y in the DAG representation of the model displayed in Figure 5. Since the random intercepts are latent we speak of *unobserved* heterogeneity.

The time-invariant variables collected in the random intercept η_x are assumed to have direct causal effects on the levels of blood insulin as indicated by the directed edges from η_x to X_1, X_2, X_3, and X_4 in Figure 5. For analogous reasons, directed edges from η_y to Y_1, Y_2, Y_3, and Y_4 are included in the DAG. The values of the structural coefficients corresponding to the paths from η_x to X_t, $t \geq 2$ are restricted to be equal to 1. These restrictions (a) assign a scale to the latent random intercept η_x and (b) reflect the assumption that the structural coefficients from the random intercepts to blood insulin levels X_t, $t \geq 2$ do not change over time. The same reasoning applies to the paths

from η_y to Y_t, $t \geq 2$. The bidirected dashed edge between the random intercepts η_x and η_y indicates that time-invariant variables that determine blood insulin levels might covary with time-invariant variables that determine the level of blood glucose due to unobserved confounding. The random intercepts are assumed to be uncorrelated with the error terms ε_{x_t}, ε_{y_t}, $t \geq 1$.

Special attention needs to be paid to the initial variables X_1 and Y_1 since they differ from the subsequent variables X_2, X_3, and X_4 and Y_2, Y_3, and Y_4 in an important way. The levels of X_t and Y_t are explained in our model by the values of X_{t-1} and Y_{t-1} when $t \geq 2$. In contrast, the initial variables X_1 and Y_1 are exogenous in our model, that is, there are no incoming directed edges into X_1 and Y_1. The dynamics of the insulin–glucose relationship, however, were going on in the individuals prior to the first measurement wave. Thus, the initial variables somehow need to account for the past of the process that is not explicitly modeled (Figure A1 in the Appendix on Initial Variables).

Different approaches for incorporating initial variables into the statistical model have been proposed (Browne & Zhang, 2007; Hamaker et al., 2005; Hsiao, 2014; Molenaar, 1985). From a causal inference perspective, we believe the most straightforward and interpretable approach is to draw directed edges from $\eta_x \overset{c_{x_1\eta_x}}{\to} X_1$ and $\eta_x \overset{c_{y_1\eta_x}}{\to} Y_1$ as well as from $\eta_y \overset{c_{x_1\eta_y}}{\to} X_1$ and $\eta_y \overset{c_{y_1\eta_y}}{\to} Y_1$ (see Figure 5). A detailed justification of this modeling strategy is presented in the Appendix on Initial
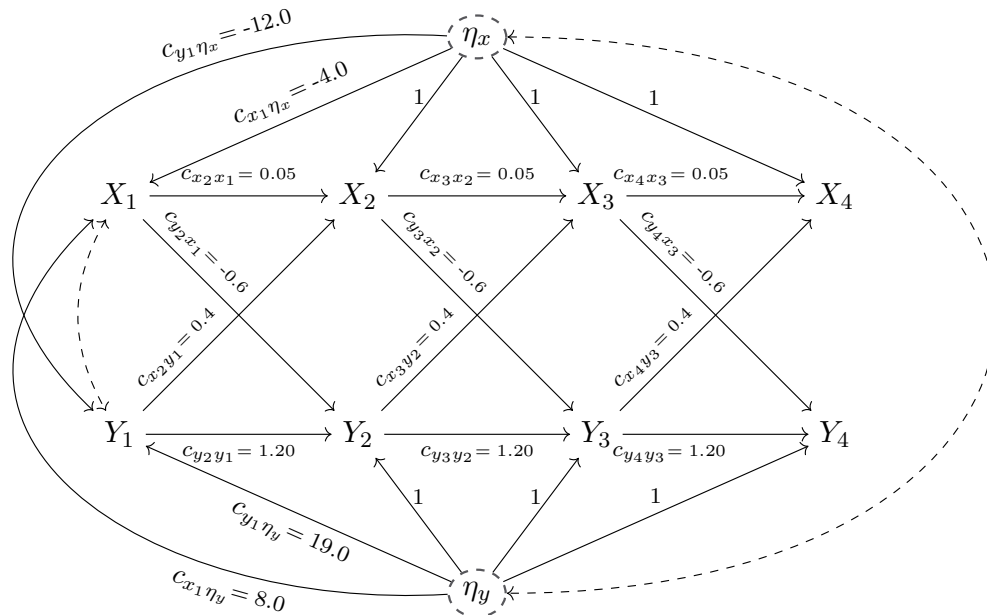

Figure 5. The causal diagram (DAG) displayed in Figure 1 is extended to include random intercepts (i.e., differences in the individual mean levels) that are represented in dashed circles, indicating that these variables are latent and therefore not observed. The directed edges from η_x to $X_1, ..., X_4$ and Y_1 indicate that time-invariant factors contained in the latent variables η_x have direct causal effects on blood insulin levels and the initial values of blood glucose levels. The directed edges from η_y to $Y_1, ..., Y_4$ and X_1 indicate that time-invariant factors contained in the latent variables η_y have direct causal effects on blood glucose levels and the initial values of blood insulin levels. The effects of the random intercepts on the initial variables are labeled with $c_{x_1\eta_x}$, $c_{x_1\eta_y}$, $c_{y_1\eta_x}$ and $c_{y_1\eta_y}$. The direct effects of η_x on X_2, X_2, and X_4 and of η_y on Y_2, Y_2, and Y_4 are assumed to be time-stable and equal to 1. The bidirected dashed edge between the random intercepts η_x and η_y indicate the existence of an unobserved confounder.

---

[7]In case one or more time-invariant variables (e.g., a person's sex) are explicitly included as predictors of, say, X_t, $t = 1, ..., T$, then η_x collects all *other* additive time-invariant factors that are not explicitly modeled (the omitted factors; e.g., marital status, genetic endowment) that determine the value of X_t.

Table 5. Numeric values of parameters related to random intercepts.

structural coefficients				variance-covariance parameters		
$c_{x_1\eta_x}$	$c_{y_1\eta_y}$	$c_{x_1\eta_y}$	$c_{y_1\eta_x}$	$\psi_{\eta_x\eta_x}$	$\psi_{\eta_y\eta_y}$	$\psi_{\eta_x\eta_y}$
−4.00	19.00	8	−12	5.00	10.00	2.50

Note. Parameters introduced to model unobserved person-specific differences in the mean levels (see Figures 5 and 6). The structural coefficients $c_{x_1\eta_x}$, $c_{y_1\eta_y}$, $c_{x_1\eta_y}$ and $c_{y_1\eta_x}$ correspond to the accumulated long-term effects of the random intercepts on insulin and glucose levels, respectively. The parameters $\psi_{\eta_x\eta_x}$, $\psi_{\eta_y\eta_y}$, and $\psi_{\eta_x\eta_y}$ represent the variance-covariance structure of the random intercepts.

Variables where we also refer to viable alternatives. As noted in the previous paragraph, the initial variables differ in an important way from subsequent measurements. This special status is accounted for by not restricting the structural coefficients $c_{x_1\eta_x}$, $c_{y_1\eta_x}$, $c_{x_1\eta_y}$ and $c_{y_1\eta_y}$ to be equal to one.

A measure of the *degree* of unobserved heterogeneity in the population is the variance of the random intercepts. In the following, we assume the numerical values $V(\eta_x) = 5$ and $V(\eta_y) = 10$, where the time-invariant factors covary with $COV(\eta_x, \eta_y) = 2.5$. All parameters introduced to model person-specific differences in mean levels via random intercepts are collected in Table 5.

A consequence of modeling unobserved heterogeneity via random intercepts is an increase in total variance as compared to the model for the homogeneous population. For example, for $t \geq 2$, the standard deviation of blood insulin levels increases from $SD(X_t) = 11.48$ mcIU/ml in the homogeneous model to $SD(X_t) = 26.30$ mcIU/ml in the model with random intercepts. Similarly, the standard deviation of blood glucose levels increases from $SD(Y_t) = 25.16$ mg/dl to $SD(Y_t) = 61.83$

mg/dl. A complete presentation of all relevant formulae and details regarding the derivation thereof is provided in the online supplementary material.

Figure 5 displays the causal diagram of the data generating mechanism. We use the bivariate linear cross-lagged panel model ($T = 4$) with random intercepts for data analysis. Since we use simulated data, we know with certainty that the causal model is correctly specified and that our statistical model captures the data generating process. In practical applications, however, one usually does not know the true data generating mechanism and causal assumptions postulated by researchers might be wrong. We discuss statistical tests of model assumptions and methods to analyze the sensitivity of causal conclusions with respect to violations of assumptions in the section on "Current Limitations and Future Directions." From a statistical point of view, we believe that the model is identified (Hisao, 2014; Oud and Delsing, 2014, Zyphur et al., 2019), that is, all model parameters can be estimated from observational data and person-specific values of the random intercepts can be calculated.

Average effects versus person-specific effects of interventions

As in the homogeneous model, we are interested in the effects of an intervention represented by the *do*-operator. Since the causal diagram of the model with random intercepts still belongs to the class of DAGs (see Figure 5), the *do*-operator is well-defined and can be directly applied to the model including random intercepts. Figure 6 displays the situation where the intervention $do(x_2)$ is applied.
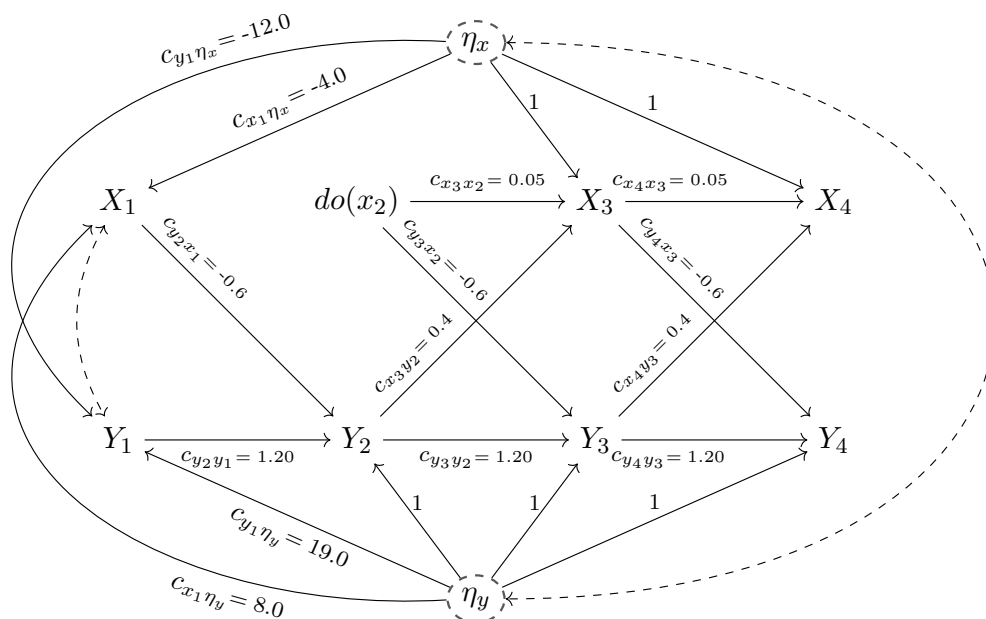

Figure 6. The causal diagram (DAG) from Figure 5 is modified to account for the intervention $do(x_2)$. The variable X_2 has been replaced by $do(x_2)$ and all directed edges that formerly entered X_2 have been removed, indicating that the interventional level $do(x_2)$ does not depend on X_1, Y_1 or η_x but is set by an experimenter (in our illustrative example we choose $do(X_2 = 26.30)$). Note that all other variables and directed edges (causal arrows) remain unchanged as compared to the situation without the intervention $do(x_2)$ as depicted in Figure 5 reflecting the assumption of modularity.

	forecast of effect of intervention	prediction of future value
average across entire population	average forecast of effect of intervention	average prediction of future value
person-specific	person-specific forecast of effect of intervention	person-specific prediciton of future value

Figure 7. The four cells of the depicted (2×2) table display all possible combinations that result from the binary distinctions 'forecast versus prediction' (columns) and 'average versus person-specific' (rows).

The interventional distribution $P(Y_3|do(x_2))$ describes blood glucose levels in the entire population of individuals, given that the treatment $do(x_2)$ has been administered. The inclusion of random intercepts allows us to define the person-specific effects of interventions.

Each individual in the population can be characterized by the values z_x and z_y of his or her time-invariant characteristics η_x and η_y. We define the person-specific interventional distribution of Y_3 for an individual characterized by $\eta_x = z_x$ and $\eta_y = z_y$ as $P(Y_3|do(x_2), \eta_x = z_x, \eta_y = z_y)$. The person-specific interventional distribution describes blood glucose levels at time $t = 3$ for an individual who is characterized by the time-invariant characteristics $\eta_x = z_x$ and $\eta_y = z_y$ after the treatment $do(x_2)$ has been applied. If multiple individuals have the same time-invariant characteristics the person-specific interventional distribution is also the subgroup-specific interventional distribution of this homogeneous group of individuals.

For both the interventional distribution $P(Y_3|do(x_2))$ and the person-specific interventional distribution $P(Y_3|do(x_2), \eta_x = z_x, \eta_y = z_y)$, the random intercepts η_x and η_y belong to every adjustment set for backdoor identification (see Figure 5; Pearl (2009); Shpitser and Pearl (2006)). Thus, to apply backdoor adjustment in the model with random intercepts, the values of the latter need to be known. Recall that random intercepts are latent variables and thus unobserved (i.e., not part of the available data). As a consequence, the backdoor adjustment formulae cannot be applied straightforwardly.[8] Causal identification can be established, however, by following the criteria for *linear* causal DAG-models suggested by Gische and Voelkle (2020) as explicated in more detail at the end of the section titled "Interventional Variance versus Conditional Variance."

Table 6. Person-specific characteristics for Amy, Joe, and Sam.

Person	z_x	z_y
Amy	$z_x^{Amy} = -SD(\eta_x) = -2.24$	$z_y^{Amy} = -SD(\eta_y) = -3.16$
Joe	$z_x^{Joe} = E(\eta_x) = 0$	$z_y^{Joe} = E(\eta_y) = 0$
Sam	$z_x^{Sam} = SD(\eta_x) = 2.24$	$z_y^{Sam} = SD(\eta_y) = 3.16$

Note. Table 6 displays the person-specific characteristics for three hypothetical individuals Amy, Joe, and Sam. Amy's and Sam's values are one standard deviation below and above the mean, respectively. Joe's values are equal to the mean of the distribution of person-specific characteristics in the population.

In the first half of the paper, we analyzed a homogeneous population and focused on the difference between forecasts of effects of interventions and predictions of future values (cells in the first row in Figure 7). The introduction of between-person heterogeneity in the second half of the paper suggests an additional distinction: The difference between average values across the entire population and person-specific values. In the following, we focus on the difference between the average effects of interventions versus person-specific effects of interventions (cells in the first column in Figure 7). Space limitations preclude a discussion of predictions of future values in the presence of unobserved heterogeneity in the main text. A formal treatment of the differences between forecasts of effects of interventions and predictions of future values in the presence of unobserved heterogeneity is presented in the online supplementary material.

In the following, we consider three hypothetical individuals from the population, namely Amy, Joe, and Sam, who are characterized by person-specific characteristics displayed in Table 6. We are interested in the blood glucose levels at time $t = 3$ after the intervention $do(X_2 = 26.30)$ has been applied, where 26.30 mcIU/ml is the standard deviation of blood insulin levels in the model with unobserved heterogeneity (recall that due to the inclusion of random intercepts into the model the standard deviation of blood insulin increased from 11.48 to 26.30.) In the following sections, we compute the person-specific effects of interventions for Amy, Joe, and Sam and compare these quantities to the average (unconditional) effect of an intervention in the entire population.

Person-specific interventional mean and variance

The unconditional mean of blood glucose is the best prediction of blood glucose level at time $t = 3$ if no further information is available. Since all observable variables are mean-centered and the random intercepts η_x and η_y have zero means, the

Table 7. Interventional means and variances in models with unobserved heterogeneity.

	Mean			Variance	
$E(Y_3)$	$E(Y_3|do(x_2))$	$E(Y_3|do(x_2), \eta_x = z_x, \eta_y = z_y)$	$V(Y_3)$	$V(Y_3|do(x_2))$	$V(Y_3|do(x_2), \eta_x = z_x, \eta_y = z_y)$
0	$-0.6x_2$	$-0.6x_2 - 14.4z_x + 23.8z_y$	3822.93	5939.03	951.43

Note. The left part of Table 7 contains the unconditional mean (used for prediction of future values) and interventional means (used to forecast outcome values after an intervention has been applied). The right part of Table 7 contains the unconditional variance (prediction error) and interventional variances (forecast errors).

---

[8]The causal quantities $P(Y_3|do(x_2))$ and $P(Y_3|do(x_2), \eta_x = z_x, \eta_y = z_y)$ are not identified based on the DAG depicted in Figure 5 in the most general non-parametric DAG setup. One way to ensure identification is to impose further assumptions on the functional form of the structural equations (e.g., linearity).

unconditional mean is equal to zero (see column one of Table 7). The unconditional variance in the presence of unobserved heterogeneity is equal to 3822.93 (see column 4 of Table 7) which is the variance of the prediction error when predicting Y_3 in the absence of additional information.

The interventional mean and variance of Y_3 describe the distribution of blood glucose levels at $t = 3$ in the entire population after the intervention $do(x_2)$ has been applied. They are given by the following formulae:

$$E(Y_3|do(x_2)) = c_{yx}x_2 \tag{6a}$$

$$V(Y_3|do(x_2)) = (c_{y_1\eta_x}c_{yy}^2 + c_{x_1\eta_x}c_{yx}c_{yy})^2\psi_{\eta_x\eta_x} \tag{6b}$$

$$+ (c_{y_1\eta_y}c_{yy}^2 + c_{x_1\eta_y}c_{yx}c_{yy} + c_{yy} + 1)^2\psi_{\eta_y\eta_y}$$

$$+ 2(c_{y_1\eta_x}c_{yy}^2 + c_{x_1\eta_x}c_{yx}c_{yy})(c_{y_1\eta_y}c_{yy}^2 + c_{x_1\eta_y}c_{yx}c_{yy} + c_{yy} + 1)\psi_{\eta_x\eta_y}$$

$$+ c_{yx}^2c_{yy}^2\psi_{x_1x_1} + c_{yy}^4\psi_{y_1y_1} + 2c_{yx}c_{yy}^3\psi_{x_1y_1} + (c_{yy}^2 + 1)\psi_{yy}$$

Equation (6a) reveals that the interventional mean in the presence of unobserved heterogeneity has the same linear functional form as in the homogeneous model (see Equation [2]). The first three lines of Equation (6b) are related to the random intercepts, whereas the last line is equal to the interventional variance in the homogeneous case (see Equation [4]). The value of $V(Y_3|do(x_2))$ in the presence of unobserved heterogeneity is equal to 5939.03 (see column 5 of Table 7), which is the variance of the forecast error when forecasting blood glucose levels at time $t = 3$ after the intervention $do(x_2)$ has been applied.

The person-specific interventional moments describe the distribution of blood glucose levels at $t = 3$ for an individual who is characterized by the time-invariant features $\eta_x = z_x$ and $\eta_y = z_y$, after the treatment $do(x_2)$ has been applied. The person-specific interventional mean is given by the following formula:

$$E(Y_3|do(x_2), \eta_x = z_x, \eta_y = z_y) \tag{7}$$

$$= c_{yx}x_2 + (c_{y_1\eta_x}c_{yy}^2 + c_{x_1\eta_x}c_{yx}c_{yy})z_x$$
$$+ (c_{y_1\eta_y}c_{yy}^2 + c_{x_1\eta_y}c_{yx}c_{yy} + c_{yy} + 1)z_y$$

Equation (7) reveals that the person-specific interventional mean is a function of the interventional level x_2 and the time-invariant characteristics z_x and z_y of that person. For example, the person-specific interventional mean for Sam is computed by plugging in Sam's numeric values $z_x^{Sam} = 2.24$ and $z_y^{Sam} = 3.16$ (see row 3 of Table 6) into the equation for the person-specific mean as stated in row 3 of Table 7:

$$E(Y_3|do(X_2 = 26.30), \eta_x = z_x^{Sam}, \eta_y = z_y^{Sam}) = \tag{8}$$

$$-0.6 \cdot 26.30 - 14.4 \cdot 2.24 + 23.8 \cdot 3.16 = 27.17$$

Likewise the person-specific interventional means for Amy (-58.73) and Joe (-15.78) can be computed. Thus, given the same treatment $do(X_2 = 26.30)$, different forecasts of blood glucose levels at time $t = 3$ are obtained for different individuals, depending on their respective time-invariant characteristics.

Note that the linear coefficient of the treatment level x_2 is constant and equal to $c_{yx} = -0.6$ in the equation for the person-specific mean (see Equation [7] and column 3 of Table 7). As a consequence, the person-specific ATE is independent of the time-invariant characteristics of a person and constant across individuals. For example, the person-specific ATE of X_2 on Y_3 for Sam is given by:

$$E(Y_3|do(X_2), \eta_X = z_X^{Sam}, \eta_y = z_y^{Sam})$$
$$- E(Y_3|do(X_2'), \eta_X = z_X^{Sam}, \eta_y = z_y^{Sam})$$
$$= -0.6x_2 - 14.4z_X^{Sam} + 23.8z_y^{Sam}$$
$$- (-0.6x_2' - 14.4z_X^{Sam} + 23.8z_y^{Sam})$$
$$= -0.6(x_2 - x_2')$$

$$\tag{9}$$

Equation (9) reveals that the ATE is independent of Sam's time-invariant characteristics (i.e., z_x^{Sam} and z_y^{Sam} cancel out), a result that also holds for every other individual in the population.[9] In other words, a *change* in the treatment level, say from $do(X_2 = x_2)$ to $do(X_2 = x_2')$, will cause the same *change* in forecasted values for each individual.

The difference between the (unconditional) interventional moments and the person-specific interventional moments can be illustrated as follows. A physician initially forecasts the value $E(Y_3|do(X_2 = 26.30)) = -0.6 \cdot 26.30 =$

Table 8. Probabilities of treatment success in models with unobserved heterogeneity.

(unconditional) probability of treatment success	person-specific probabilities of treatment success
$P(Y_3^* = 1|do(X_2 = 26.30)) = .52$	$P(Y_3^* = 1|do(X_2 = 26.30), \eta_x = z_x^{Amy}, \eta_y = z_y^{Amy}) = .27$ $P(Y_3^* = 1|do(X_2 = 26.30), \eta_x = z_x^{Joe}, \eta_y = z_y^{Joe}) = .78$ $P(Y_3^* = 1|do(X_2 = 26.30), \eta_x = z_x^{Sam}, \eta_y = z_y^{Sam}) = .94$

Note. Y_3^* is an indicator variable that is equal to 1 if blood glucose levels at $t = 3$ fall inside the acceptable range $[-40, 80]$ (treatment success), and is equal to 0 otherwise. The left part of the Table 8 contains the (unconditional) probability of treatment success after the treatment $do(X_2 = 26.30)$ has been applied. The right part of Table 8 contains person-specific probabilities of treatment success for three hypothetical individuals. Amy's time-invariant values are one standard deviation below the mean, Joe's values are exactly at the mean, and Sam's values are one standard deviation above the mean.

---

[9]This result is a consequence of modeling unobserved heterogeneity via additive random intercepts. To capture heterogeneity in the person-specific ATE we need more complex models (e.g., models including product terms).

−15.78 for a *new* patient under treatment of whom the person-specific characteristics are yet unknown. After learning that a patient is Amy, Joe, or Sam (and their respective person-specific characteristics) the physician updates his or her forecast to -58.73, -15.78, or 27.17, respectively. Note that by learning the person-specific characteristics of a new patient, the variance of the forecast error drops from 5939.03 to 951.43 (see columns 5 and 6 of Table 7).

Person-specific probabilities of treatment success

As in the section titled "Interventional Probabilities versus Conditional Probabilities" we use the binary variable Y_3^* to indicate ($Y_3^* = 1$) that blood glucose levels at time $t = 3$ are within an acceptable range $[-40, 80]$ (mean-centered metric), an event we refer to as treatment success. Recall that $Y_3^* = 0$ indicates the presence of hypo- or hypoglycemia, that is, blood glucose levels outside the acceptable range.

In the presence of unobserved heterogeneity, the (unconditional) probability of treatment success is equal to .52 given the treatment $do(X_2 = 26.30)$ (see column 1 of Table 8). That is, a physician would estimate a 52% probability of treatment success after the intervention $do(X_2 = 26.30)$ has been applied to a new patient from the heterogeneous population (a patient whose person-specific characteristics are yet unknown).

The right column of Table 8 contains the person-specific probabilities of treatment success for the three hypothetical individuals Amy, Joe, and Sam. After learning that a patient is Amy, Joe, or Sam (and their respective person-specific characteristics) the physician would update his or her initial beliefs about the probability of treatment success from 52% to 27%, 78%, or 94%, respectively. The reasons for this update are twofold. First, the person-specific means differ across the three individuals (see Equations [7] and [8]). Second, the amount of uncertainty when forecasting Y_3 drops from

5939.03 to 951.43 after learning the person-specific characteristics (see columns 5 and 6 of Table 7).

Figure 8 displays the person-specific probabilities of treatment success for Amy (dashed line), Joe (solid line) and Sam (dotted line) as a function of the interventional level (population mean-centered metric). Each curve has a unique maximum. Due to the simple nature of the model the three curves share the same maximal value of 94.8%.[10] Note that the person-specific optimal treatment levels differ across persons: $do(X_2 = -104.92)$ for Amy, $do(X_2 = -33.33)$ for Joe, and $do(X_2 = 38.25)$ for Sam. In other words, if a given person is administered his or her individually optimal treatment level, the probability of treatment success for that person is 94.8%.

Joe is the average patient, that is, his person-specific characteristics coincide with the respective population means (see row 2 of Table 6). The person-specific optimal treatment level for Joe is -33.33 which is also the average optimal treatment level in the entire population. If a physician treats a *new* patient from the population for whom the person-specific characteristics are yet unknown, the optimal treatment decision would be $do(X_2 = -33.33)$. If the physician learns that the new patient is Joe (i.e., the average patient) the initial treatment $do(X_2 = -33.33)$ coincides with Joe's person-specific optimal treatment resulting in a 94.8% probability of treatment success. On the other hand, if the new patient turns out to be Amy or Sam, then the treatment $do(X_2 = -33.33)$ is no longer optimal (given this new person-specific information) and would result in a 71% probability of treatment success in either case. The optimal treatment decision after learning that the new patient is Amy or Sam would be $do(X_2 = -104.92)$ or $do(X_2 = 38.25)$, respectively, resulting in a 94.8% probability of treatment success in either case. Thus, updating an initial population-based optimal treatment decision to include Amy's or Sam's person-specific characteristics, results in an absolute increase of 23.8% in the probability of treatment success.
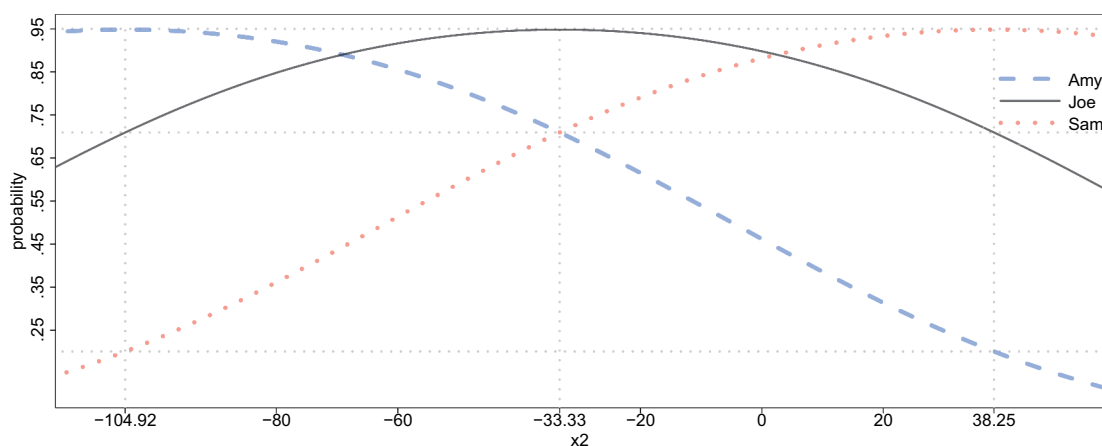

Figure 8. The person-specific interventional probabilities (vertical axis) are depicted for the three hypothetical individuals Amy (dashed line), Joe (solid line) and Sam (dotted line) as a function of the (mean-centered) interventional level (horizontal axis). Vertical-dotted lines are drawn at the interventional levels that maximize the respective probabilities of treatment success. For example, the interventional level $do(X_2 = 38.25)$ maximizes the probability of treatment success for Sam, which is 94.8%. A dotted horizontal line is drawn at .948 which is the probability of treatment success under optimal treatment. Two more dotted horizontal lines are drawn at .71 and .20 which correspond to person-specific probabilities of treatment success under different types of non-optimal treatment (see Table 6).

---

[10]More complex models (e.g., models allowing for person-specific differences in error term variances) would be needed to model individual differences in the probabilities of treatment success under optimal treatment. We will return to this limitation in the Discussion.

Discussion

In this article, we focused on the example of a four-wave linear cross-lagged panel design to provide a didactic presentation of Pearl's (2009) approach to causal inference for researchers familiar with structural equation modeling.[11] Throughout we emphasized the important distinction between (a) forecasting the future value of a variable following an intervention at a prior wave and (b) predicting the future value of a variable following measurement of variables at a prior wave. In task (a) a causal effect is computed based on the interventional distribution in a DAG-based framework. We focus on $do(x_t)$-type interventions that set each individual's value on the variable X at time t to a fixed constant value. In task (b) the future value of a variable is predicted using the conditional distribution (statistical quantity). Task (a) requires identification of the interventional distribution and its moments (causal quantities). Causal identification in the case of a homogeneous population can be established using the machinery of Pearl's (2009) approach to causal inference which includes graphical criteria such as the backdoor criterion. Causal identification in case of a heterogeneous population can be established using a parametric procedure suggested by Gische and Voelkle (2020).

Our initial development of the approach assumed that all individuals were homogeneous. Later, we relaxed this assumption to account for unobserved heterogeneity (person-specific differences in the mean levels of the two variables) by including additive random intercepts into the linear cross-lagged panel framework. We showed that – even in this simple model – a treatment decision that is optimal on average can be improved substantially if person-specific characteristics can be incorporated into the model. In other words, treating a new, unknown patient with a treatment dose that has proven to be optimal on average (e.g., based on the results of randomized control trials) will only be optimal for a specific person as long as additional person-specific information is *not* available (Voelkle et al., 2018). As (a) relevant person-specific information becomes available or (b) repeated measures of the individual become available that allow us to control for time-invariant person characteristics, this information can be incorporated to develop a treatment level that more adequately matches the optimal level of treatment for the individual.

Pearl's (2009) approach to DAG-based causal inference was developed using a general nonparametric framework. We used a linear parameterization of the cross-lagged panel framework to simplify the results and show the connection to linear structural equation modeling. The linear parameterization has weaknesses and strengths. On the one hand, if linear relationships represent a serious misspecification of the unknown true nonlinear functional form, the models will be misspecified and the results will be biased. On the other hand, the linear specification permits the inclusion of random intercepts, that is, person-specific differences in the mean levels of the time series, which reduces the risk of *causal* misspecification. In linear models with random intercepts additive time-

invariant unobserved confounders are being statistically controlled. This, in turn, allows for the identification of causal effects that would *not* be identified in the most general nonparametric setup.[12] In the case of linear models with random intercepts, we can to use the approach developed by Gische and Voelkle (2020) to establish identification of causal effects. The latter approach could also be used to obtain estimates of causal quantities in the presence of random intercepts.

Current limitations and future directions

The methods presented in this paper permit causal conclusions to be reached based on a combination of observational data and assumptions about the data generating mechanism. The effects of a hypothetical *do*-intervention were forecasted without performing an experiment or otherwise perturbing the dynamic system. For causal conclusions based on observational data to be valid, the underlying assumptions need to hold. The methods discussed in this paper rely on the correct a priori specification of the causal ordering of the variables and the assumption of a linear dynamic process that is both stable over time and that is not altered by the intervention (modularity, autonomy). In the present article, we only allowed for between-person differences in the individual mean levels of the variables. Computing causal forecasts based on models that account for between-person differences in the autoregressive and cross-lagged effects remains a task for future research.

Statistical tests are available that can detect certain types of misspecification. In the present paper the set of possible causal orders is limited in part by the cross-lagged panel design (e.g., time ordering of variables). Further tests of the causal ordering of the variables (e.g., maximum lag of autoregressive effects) can be performed using procedures proposed by Chen et al. (2014), Shipley (2003), Thoemmes et al. (2018), and Tian and Pearl (2002). The assumption of autonomous mechanisms is in principle testable, but requires that both observational and experimental data on the process are available. The consequences of violations of non-testable assumptions can sometimes be gauged via sensitivity analyses and robustness checks (Ding & VanderWeele, 2016; Imai et al., 2010; Rosenbaum, 2011). Given that our developments were in the context of linear causal models, the linearity of the functional relations should be examined. Graphical checks utilizing lowess fits or splines commonly used in regression can be performed (e.g., Cohen et al., 2003; Suk et al., 2019). Formal statistical tests of linearity are available both for nested and non-nested models (Amemiya, 1985; Lee, 2007; Schumacker & Marcoulides, 1998).

In the development of our approach, we did not consider measurement error. Cole and Preacher (2014) have highlighted the serious bias that can arise from the failure to address measurement error in structural equation models. In our insulin-glucose example we would expect this problem to be minimal since instantaneous measures of blood insulin and blood glucose both have high precision. In cases in which only

---

[11]Technical details, formal derivations, and R-code for numerical computations are provided in the online supplementary material.
[12]All models discussed in the paper belong to the class of DAGs. Random intercepts are by definition latent variables and are therefore unobserved. Therefore, many standard nonparametric adjustment formulae used to compute causal quantities can no longer be applied in the presence of random intercepts.

a single measure of each variable is available at each measurement wave, the bias can be reduced by using methods to correct the observed variables for measurement error (e.g., Fuller, 1987) or structural equation models with latent variables can be constructed that yield measurement error-free estimates. Our proposed approach to cross-lagged panel designs can potentially be extended to define, identify, and estimate causal effects between latent variables. Again, this is a task for future research.

Finally, our example was based on simulated data based on empirical findings reported by Ito et al. (1998). To simplify our didactic presentation of the approach, we ignored both possible contemporaneous correlations among the residuals of the insulin and glucose time-series as well as possible autocorrelations in the residuals for the blood insulin and blood glucose series. From a causal perspective, this means that we generated the data to reflect the absence of time-varying unobserved confounders in both the autoregressive and the cross-lagged relationships. With a real data example, these assumptions need to be checked, for example, by using the methods to detect certain types of misspecifications referred to in the second paragraph of this section.

In conclusion, the proposed method combines the desirable features of cross-lagged panel designs, a sound definition of causal quantities based on DAGs, and the flexibility of SEM for data analysis. We hope that our presentation facilitates the integration of linear SEM and DAG-based causal inference and enables researchers to better distinguish between tools for predictive and causal research questions.

Acknowledgments

Disclosure statement

The authors do not have any conflicts of interest to disclose.

Funding

ORCID

Christian Gische 🆔 http://orcid.org/0000-0002-7372-2894
Manuel C. Voelkle 🆔 http://orcid.org/0000-0001-5576-8103

References

Amemiya, T. (1985). *Advanced econometrics*. Harvard University Press.

Bauer, D. J. (2003). Estimating multilevel linear models as structural equation models. *Journal of Educational and Behavioral Statistics*, 28, 135–167. https://doi.org/10.3102/10769986028002135

Bode, B. W., Schwartz, S., Stubbs, H. A., & Block, J. E. (2005). Glycemic characteristics in continuously monitored patients with type 1 and type 2 diabetes: Normative values. *Diabetes Care*, 28, 2361–2366. https://doi.org/10.2337/diacare.28.10.2361

Bollen, K. A. (1989). *Structural equations with latent variables*. John Wiley & Sons.

Bollen, K. A., & Curran, P. J. (2004). Autoregressive latent trajectory (ALT) models: A synthesis of two traditions. *Sociological Methods & Research*, 32, 336–383. https://doi.org/10.1177/0049124103260222

Browne, M. W., & Zhang, G. (2007). Developments in the factor analysis of individual time series. In R. Cudeck & R. C. MacCallum (Eds.), *Factor analysis at 100: Historical developments and future directions* (pp. 265–292). Lawrence Erlbaum Associates.

Chen, B., Tian, J., & Pearl, J. (2014). Testable implications of linear structural equation models. In *Proceedings of the 28th AAAI conference on artificial intelligence* (pp. 2424–2430). AAAI Press.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Lawrence Erlbaum Associates.

Cole, D. A., & Preacher, K. J. (2014). Manifest variable path analysis: Potentially serious and misleading consequences due to uncorrected measurement error. *Psychological Methods*, 19, 300–315. https://doi.org/10.1037/a0033805

Cowell, R. G., Dawid, A. P., Lauritzen, S. L., & Spiegelhalter, D. J. (1999). *Probabilistic networks and expert systems*. Springer.

Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist*, 12, 671–684. https://doi.org/10.1037/h0043943

Cronbach, L. J. (1975). Beyond the two disciplines of scientific psychology. *American Psychologist*, 30, 116–127. https://doi.org/10.1037/h0076829

Cryer, P. E. (2003). Glucose homestasis and hypoglycemia. In P. R. Larsen (Ed.), *Williams textbook of endocrinology* (10th ed.). W.B. Saunders.

Curran, P. J. (2003). Have multilevel models been structural equation models all along? *Multivariate Behavioral Research*, 38, 529–569. https://doi.org/10.1207/s15327906mbr3804_5

Danaei, G., Finucane, M. M., Lu, Y., Singh, G. M., Cowan, M. J., Paciorek, C. J., Lin, J. K., Farzadfar, F., Khang, Y.-H., Stevens, G. A., Rao, M., Ali, M. K., Riley, L. M., Robinson, C. A., & Ezzati, M. (2011). National, regional, and global trends in fasting plasma glucose and diabetes prevalence since 1980: Systematic analysis of health examination surveys and epidemiological studies with 370 country-years and 2·7 million participants. *The Lancet*, 378, 31–40. https://doi.org/10.1016/S0140-6736(11)60679-X

Ding, P., & VanderWeele, T. J. (2016). Sensitivity analysis without assumptions. *Epidemiology*, 27, 368–377. https://doi.org/10.1097/EDE.0000000000000457

Elwert, F. (2013). Graphical causal models. In S. Morgan (Ed.), *Handbook of causal analysis for social research* (pp. 245–273). Springer.

Fuller, W. A. (1987). *Measurement error models*. John Wiley & Sons.

Gische, C., & Voelkle, M. C. (2020). *Beyond the mean: A flexible framework for studying causal effects using linear models*. Retrieved from https://www.researchgate.net/publication/335030449_Gische_Voelkle_Causal_Inference_in_Linear_Models

Giugliano, D., Marfella, R., Coppola, L., Verrazzo, G., Acampora, R., Giunta, R., Nappo, F., Lucarelli, C., & D'Onofrio, F. (1997). Vascular effects of acute hyperglycemia in humans are reversed by l-arginine. *Circulation*, 7, 1783–1790. https://doi.org/10.1161/01.CIR.95.7.1783

Hamaker, E. L., Dolan, C. V., & Molenaar, P. C. M. (2005). Statistical modeling of the individual: Rationale and application of multivariate stationary time series analysis. *Multivariate Behavioral Research*, 40, 207–233. https://doi.org/10.1207/s15327906mbr4002_3

Hamilton, J. (1994). *Time series analysis*. Princeton, NJ: Princeton University Press.

Holland, P. (1986). Statistics and Causal Inference. *Journal of the American Statistical Association*, 81, 945–960. https://doi.org/10.2307/2289064

Hsiao, C. (2014). *Analysis of panel data* (3rd ed.). Cambridge University Press.

Imai, K., Keele, L., & Yamamoto, T. (2010). Identification, inference and sensitivity analysis for causal mediation effects. *Statistical Science*, 25, 51–71. https://doi.org/10.1214/10-STS321

Ito, K., Wada, T., Makimura, H., Matsuoka, A., Maruyama, H., & Saruta, T. (1998). Vector autoregressive modeling analysis of frequently sampled oral glucose tolerance test results. *The Keio Journal of Medicine*, 47, 28–36. https://doi.org/10.2302/kjm.47.28

Jordan, M. I. (Ed.). (1998). *Learning in graphical models*. Kluwer Academic Press.

Kang, C., & Tian, J. (2009). Markov properties for linear causal models with correlated errors. *Journal of Machine Learning Research*, 10, 41–70.

Koller, D., & Friedman, N. (2009). *Probabilistic graphical models: Principles and techniques - adaptive computation and machine learning*. The MIT Press.

Kuroki, M., & Cai, Z. (2007). Evaluation of the causal effect of control plans in nonrecursive structural equation models. In *Proceedings of the twenty-third conference on uncertainty in artificial intelligence* (pp. 227–234). AUAI Press.

Lee, S.-Y. (2007). *Structural equation modeling: A Bayesian approach*. John Wiley & Sons.

Loehlin, J. C., & Beaujean, A. A. (2017). *Latent variable models: An introduction to factor, path, and structural equation models* (5th ed.). Routledge.

Molenaar, P. C. M. (1985). A dynamic factor model for the analysis of multivariate time series. *Psychometrika*, 50, 181–202. https://doi.org/10.1007/BF02294246

Molenaar, P. C. M. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement*, 2, 201–218. https://doi.org/10.1207/s15366359mea0204_1

Monnier, L., Colette, C., Wojtusciszyn, A., Dejager, S., Renard, E., Molinari, N., & Owens, D. R. (2017). Toward defining the threshold between low and high glucose variability in diabetes. *Diabetes Care*, 40, 832–838. https://doi.org/10.2337/dc16-1769

Oud, J., & Delsing, M (2010). Continuous Time Modeling of Panel Data by Means of SEM. In van Monfort, K., Oud, J., and Satorra, A. (Eds.), Longitudinal Research with Latent Variables (Ch. 7). New York, NY: Springer.

Pearl, J. (2009). *Causality* (2nd ed.). Cambridge University Press.

Pearl, J., Glymour, M., & Jewell, N. P. (2016). *Causal inference in statistics: A primer*. John Wiley & Sons.

Peters, J., Janzing, D., & Schölkopf, B. (2017). *Elements of causal inference*. The MIT Press.

Rao, C. (1973). *Linear statistical inference and its applications* (2nd ed.). Wiley.

Richardson, T. (2003). Markov properties for acyclic directed mixed graphs. *Scandinavian Journal of Statistics*, 30, 145–157. https://doi.org/10.1111/1467-9469.00323

Rosenbaum, P. (2011). *Observational studies* (2nd ed.). Springer.

Schumacker, R., & Marcoulides, G. (1998). *Interaction and nonlinear effects in structural equation modeling*. Lawrence Erlbaum Associates.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.

Shipley, B. (2003). Testing recursive path models with correlated errors using d-separation. *Structural Equation Modeling*, 10, 214–221. https://doi.org/10.1207/S15328007SEM1002_3

Shpitser, I., & Pearl, J. (2006). Identification of conditional interventional distributions. In R. Dechter & S. T. Richardson (Eds.), *Proceedings of the 22nd conference on uncertainty in artificial intelligence* (pp. 437–444). AUAI Press.

Sidman, M. (1960). *Tactics of scientific research*. Basic Books.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford University Press.

Spirtes, P., Glymour, C., & Scheines, R. (2000). *Causation, prediction, and search* (2nd ed.). The MIT Press.

Suk, H. W., West, S. G., Fine, K. L., & Grimm, K. J. (2019). Nonlinear growth curve modeling using penalized spline models: A gentle introduction. *Psychological Methods*, 24, 269–290. https://doi.org/10.1037/met0000193

Textor, J., van der Zander, B., Gilthorpe, M. S., Liśkiewicz, M., & Ellison, G. T. H. (2016). Robust causal inference using directed acyclic graphs: The R package 'dagitty'. *International Journal of Epidemiology*, 45, 1887–1894. https://doi.org/10.1093/ije/dyw341

Thoemmes, F., Rosseel, Y., & Textor, J. (2018). Local fit evaluation of structural equation models using graphical criteria. *Psychological Methods*, 23, 27–41. https://doi.org/10.1037/met0000147

Tian, J., & Pearl, J. (2002). On the testable implications of causal models with hidden variables. In A. Darwiche & N. Friedman (Eds.), *Proceedings of the 18th conference on uncertainty in artificial intelligence* (pp. 519–527). Morgan Kaufmann.

Usami, S., Murayama, K., & Hamaker, E. L. (2019). A unified framework of longitudinal models to examine reciprocal relations. *Psychological Methods*, 24, 637–657. https://doi.org/10.1037/met0000210

Voelkle, M. C., Gische, C., Driver, C. C., & Lindenberger, U. (2018). The role of time in the quest for understanding psychological mechanisms. *Multivariate Behavioral Research*, 53, 782–805. https://doi.org/10.1080/00273171.2018.1496813

West, S. G., & Thoemmes, F. (2010). Campbell's and Rubin's perspectives on causal inference. *Psychological Methods*, 15, 18–37. https://doi.org/10.1037/a0015917

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data* (2nd ed.). MIT Press.

Wright, S. (1934). The method of path coefficients. *The Annals of Mathematical Statistics*, 5, 161–215. https://doi.org/10.1214/aoms/1177732676

Zyphur, M. J., Allison, P. D., Tay, L., Voelkle, M. C., Preacher, K. J., Zhang, Z., … Diener, E. (2019). From data to causes I: Building a general cross-lagged panel model (GCLM). *Organizational Research Methods*. https://doi.org/10.1177/1094428119847278

Appendix on Initial Variables

In this Appendix, we provide a rationale for our treatment of initial variables and briefly discuss alternative approaches. We focus on the model depicted in Figure 5 that includes individual heterogeneity. The model depicted in Figure 1 can be viewed as a special case.

The initial variables X_1 and Y_1 in Figure 5 are (a) connected by a bidirected dashed edge, (b) have incoming directed edges from the random intercepts η_x and η_y, and (c) the values of the structural coefficients corresponding to the directed edges are denoted by $c_{x_1\eta_x}$, $c_{x_1\eta_y}$, $c_{y_1\eta_x}$ and $c_{y_1\eta_y}$. All three choices (a), (b), and (c) rest on the key assumptions depicted in Figure A1 that address the dynamics of the insulin–glucose relationship *prior* to the first measurement wave. The DAG depicted in Figure A1 encodes the assumption, that the *causal structure* of the insulin glucose dynamics is time-stable. In other words, the dynamics prior to the first measurement (gray part) contains only autoregressive and cross-lagged effects of order one. Note that we do *not* assume that the *values* of the structural coefficients are necessarily time-stable prior to the first measurement wave as indicated by leaving out labels for the directed edges in the gray part of the graph.

Drawing (a) a bidirected edge among the initial variables in Figure 5 indicates the existence of unobserved confounders. Assuming that the dynamics have already been going on before the first measurement wave (as displayed in Figure A1) implies that the X_1-Y_1 relationship is confounded by previous insulin and glucose levels. Examples for such confounding paths are $X_1 \leftarrow X_0 \rightarrow Y_1$ or $X_1 \leftarrow Y_0 \rightarrow Y_1$.

Drawing (b) directed edges from the random intercepts to the initial variables can be justified as follows: Directed edges drawn from η_x to the insulin levels X_t reflect the assumption that the omitted person-specific time-invariant variables have direct causal effects on the insulin levels at time t. Since the timing of the first measurement wave was determined by factors outside of the system (e.g., grant money deadlines, availability of the lab) there is no reason to believe that η_x should not have a direct causal effect on the initial variable X_1 in Figure 5. Including the past of the dynamic process prior to the first measurement (gray part in Figure A1) reveals that insulin level at time $t = 1$ does not have a special status in the ongoing *underlying real-world* process. The fact that we treat

it differently in our *model* (e.g., we assign the structural coefficient $c_{x_1\eta_x}$ to the path $\eta_x \to X_1$ instead of restricting its value to be equal to one) only arises due to the timing of our measurements which is determined by factors independent of the insulin-glucose dynamics (see Singer & Willett, 2003). The same argument holds true for the directed edge from η_y to the glucose level Y_1.

The directed edge from η_y to X_1 in Figure 5 reflects the assumption that η_y has a direct causal effect on X_1 *in the depicted model.*[13] Note that the model in Figure 5 omits those parts of the insulin glucose dynamics that were going on prior to the initial measurement wave. The DAG in depicted in Figure A1 explicitly contains the entire past of the insulin-glucose dynamics (gray part). As a consequence, η_y has indirect effects on X_1 (e.g., $\eta_y \to Y_0 \to X_1$,

$\eta_y \to Y_{-1} \to X_0 \to X_1$) in the model depicted in Figure A1. These indirect paths from Figure A1 translate to direct paths in the linear cross-lagged panel model with $T = 4$ measurement waves as depicted in Figure 5. The same argument holds true for the directed edge from η_x to the insulin levels Y_1.

Statement (c) says that the values of the structural coefficients corresponding to the directed edges are denoted by $c_{x_1\eta_x}$, $c_{x_1\eta_y}$, $c_{y_1\eta_x}$ and $c_{y_1\eta_y}$ and are therefore not restricted to be equal to one. In practice, we recommend estimating the values of these coefficients from the data without imposing any restrictions. For our data simulation we used a particular set of values, namely $c_{x_1\eta_x} = -4.00$, $c_{x_1\eta_y} = 8.00$, $c_{y_1\eta_x} = -12.00$ and $c_{y_1\eta_y} = 19.00$, as depicted in Figures 5 and 6 (see online supplementary material for a derivation of these values).
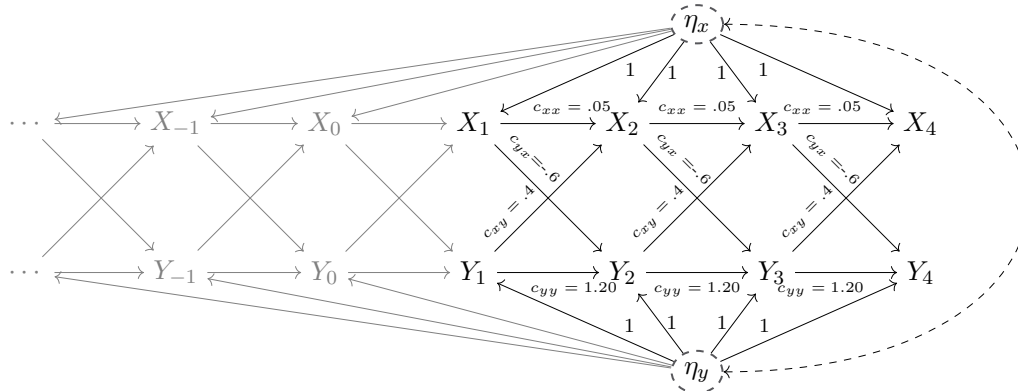

Figure A1. The depicted causal diagram (DAG) extends Figure 5 to account for the fact that the insulin-glucose dynamics were going on in the individuals prior to the first measurement wave. The four measurement waves that are observed in our study are drawn in black whereas the unobserved past of the process is drawn in gray. The graph in Figure A1 encodes the assumption, that the causal structure of the insulin-glucose dynamics is time-stable. In other words, the dynamics prior to the first measurement contain only autoregressive and cross-lagged effects of order one as well as direct effects from η_x to the insulin measures and from η_y to the glucose measures. We do not necessarily assume that the values of the structural coefficients are time-stable prior to the first measurement wave as indicated by leaving the directed edges in the gray part of Figure A1 unlabeled.

---

[13]Note that the term direct effect is well-defined only with respect to a certain model. Saying that changes in insulin levels X have a direct effect on glucose levels Y might be a useful model for practicing physicians and patients suffering from diabetes. At the same time, this model excludes several chains of chemical reactions (i.e., molecular processes) that are involved in the transmission of changes in X to changes in the levels of Y. These intermediate chemical processes might be of interest to biochemists or pharmacologists who develop a drug for patients suffering from diabetes. Thus, a direct effect of X on Y *in the physician's model* will then be an indirect effect (i.e., a mediated effect) *in the biochemists' model* that explicitly includes chemical reactions as intermediate variables.