Studying “the effect of change on change”: a different viewpoint

Eyal Shahar

Professor, Division of Epidemiology and Biostatistics,
Mel and Enid Zuckerman College of Public Health, University of Arizona

Correspondence:

Eyal Shahar
Professor
Division of Epidemiology and Biostatistics
Mel and Enid Zuckerman College of Public Health
1295 N. Martin Ave.
Campus Mail P.O. Box 245211
Tucson, AZ 85724

Email: shahar@email.arizona.edu
Abstract

When a causal variable and its presumed effect are measured at two time points in a cohort study, most researchers prefer to fit some type of a change model. Many of them believe that such an analysis is superior to a cross-sectional analysis because change models estimate “the effect of change on change”, which sounds epistemologically stronger than “a cross-sectional association”. In this article I trace two regression models of change to their cross-sectional origin and explain that these models do not test anything conceptually different from cross-sectional models. A change model is superior to a cross-sectional model because it corresponds to a self-matched design. Therefore, the main advantage of regressing “change on change” is complete control of time-stable confounders.
Introduction

The most rudimentary example of longitudinal data analysis may be a cohort study where the putative cause and its putative effect are measured twice: at baseline and again some time during follow up. In such cases researchers welcome the opportunity to study “the effect of change on change”, intuitively assuming that longitudinal changes, rather than a cross-sectional association, is a different kind of effect. After all, causal effects are grasped as “the change in Y in a person, if X were changed”. In this article I suggest that intuition leads us astray. A model of changes between two time points is not testing anything conceptually different from a cross-sectional model, and its key advantage operates in the domain of confounding.

To provide a concrete example, rather than use generic notation, I will consider variables from a question of interest to sleep researchers: what is the effect of disordered breathing during sleep (of which sleep apnea is the most severe manifestation) on systolic or diastolic blood pressure (BP)? Although measuring sleep-disordered breathing is not simple, a commonly used variable is called the respiratory disturbance index (RDI)—the average number of abnormal breathing episodes per hour of sleep. The data structure is simple. In a typical cohort study the RDI and BP are measured twice, several years apart.

<table>
<thead>
<tr>
<th>Time 1</th>
<th>Time 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>RDI₁</td>
<td>RDI₂</td>
</tr>
<tr>
<td>BP₁</td>
<td>BP₂</td>
</tr>
</tbody>
</table>

In the interest of simplified notation, the subscript “ᵢ” (indicating the ᵢ-th person) is omitted throughout.
The first type of a change model

To study the effect of the RDI on BP, we may fit two cross-sectional linear regression models, namely, regress BP₁ on RDI₁ and regress BP₂ on RDI₂. Most researchers, however, will prefer to fit some change model, the simplest of which is

\[ BP₂ - BP₁ = β₀ + β (RDI₂ - RDI₁) + ε \]  \hspace{1cm} \text{(Model 1)}

This change model, however, could be traced to simple subtraction of two cross-sectional models—provided that the effect of the RDI on BP is forced to be identical at both times:[1, page 10]

\[ BP₂ = μ₂ + β RDI₂ + γZ + α + ε₂ \] \hspace{1cm} \text{(Model 2)}
\[ BP₁ = μ₁ + β RDI₁ + γZ + α + ε₁ \] \hspace{1cm} \text{(Model 3)}

\[ BP₂ - BP₁ = β₀ + β (RDI₂ - RDI₁) + ε \] \hspace{1cm} \text{(Model 1)}

Where,

- \( Z \) is a vector of confounders whose effect on BP is identical at the two time points.
- \( α = \) all differences between persons that remain constant over time (and are not captured by \( γZ \))
- \( β₀ = μ₂ - μ₁ \)
- \( ε = ε₂ - ε₁ \)
Comparing model 1 (the change model) to models 2 and 3, we realize that we haven’t really studied any unique causal idea of change that was missing from the two cross-sectional models. The coefficient of the change variable ($\beta$)—the so-called effect of change from model 1—is not different from the coefficient of the cross-sectional effect. Notice, again, that we had assumed identical cross-sectional effects of the RDI on BP, which implies no effect modification by “time” on an additive scale. (A little deeper thinking would suggest that “time” is a surrogate for time-dependent variables, rather than clock ticking.)

So why fit the “change model” instead of fitting two cross-sectional models?

As shown above, the key advantage is control of all time-stable confounders (both measured and unmeasured Z variables), which are eliminated by subtracting the cross-sectional models.[1, page 10] No cross-sectional model alone offers that important advantage!

Interestingly, we may reach the very same conclusion about confounding control from the viewpoint of study design, too. The change model is essentially derived from a self-matched cohort design, because we estimate the RDI effect on BP within each person (and then take the average over all people.) Of course, by matching on personal identity, the person-specific estimates of the RDI effect cannot be confounded by any characteristic that has not changed between the two time points (gender, genes, etc).
It is perhaps easier to recognize the matching property of the model if the continuous
causal variable, RDI, is replaced by a binary (0, 1) exposure variable, say, taking an anti-
hypertensive drug (DRUG). Then, the difference BP\(_2\)–BP\(_1\) for a person who is observed
at both DRUG=1 and at DRUG=0 is derived from two BP values that such a person has
 contributed: once as exposed (DRUG=1) and once as unexposed (DRUG=0). We also
see that people who took the drug at both time points, or at neither time point, do not
contribute anything to the estimate. Stated differently: the change model in a cohort
study may be viewed as the observational counterpart of the randomized crossover trial.
In both designs people alternate between at least two exposure values.

Finally, notice that the change model has no built-in conditioning. We do not condition
the association of RDI\(_2\) with BP on RDI\(_1\), because both coefficients are forced to have the
same absolute value (BP\(_2\) – BP\(_1\) = \(\beta_0 + \beta\) RDI\(_2\) – \(\beta\) RDI\(_1\) + \(\varepsilon\)). From the perspective of
causal diagrams,[ref] the model simply assumes that RDI\(_t\)\(\rightarrow\)BP\(_t\) (where \(t=1, 2\)).

**The second type of a change model**

So far we have forced homogeneity of the RDI effect on BP at both times. From a
theoretical standpoint, we should often accept that assumption: Why would the effect of
the RDI be different at two arbitrary time points, especially when no relevant causal
reasoning has dictated the choice of the time points at which these variables were
measured? Furthermore, each measured RDI presumably serves as a surrogate for a
summary of historical RDI values. Sleep researchers do not assume that the value on a single night determines a contemporaneous blood pressure value.

Nonetheless, it is easy to relax the homogeneity assumption and derive a different kind of a “change model”: \[1, page 15\]

\[
\begin{align*}
BP_2 &= \mu_2 + \beta_2 RDI_2 + \gamma Z + \alpha + \epsilon_2 \\
BP_1 &= \mu_1 + \beta_1 RDI_1 + \gamma Z + \alpha + \epsilon_1
\end{align*}
\]

After subtraction: \(BP_2 - BP_1 = \beta_0 + \beta_2 RDI_2 - \beta_1 RDI_1 + \epsilon\) (Model 4)

With reorganization of the last equation, we may model \(BP_2\) alone as the dependent variable and combine \(BP_1\) with \(\beta_0\) to form a person-specific intercept:

\[
BP_2 = [BP_1 + \beta_0] + \beta_2 RDI_2 - \beta_1 RDI_1 + \epsilon \quad \text{(Model 5)}
\]

Again, as we see above the key advantage of this model is control of all time-stable confounders, both measured and unmeasured. But there is more to be said: Since we allowed the effect of \(RDI_2\) to differ from that of \(RDI_1\), we have also conditioned the associations. This model has four key properties:

1) The association of \(RDI_2\) with \(BP_2\) is conditional on \(RDI_1\)

2) The association of \(RDI_1\) with \(BP_2\) is conditional on \(RDI_2\)

3) Both conditional associations control all time-stable confounders

4) The coefficient of \(RDI_1\) (namely, \(\beta_1\)) may be biased! In particular, if \(RDI_1\) affects \(RDI_2\)
A causal diagram perspective of model 5

To understand the fourth property of model 5, we should recall two principles of directed acyclic graphs (causal diagrams):[2, 3]

1) The marginal associations between two variables is the product of causal paths between them (the effect of interest) and back-door paths (confounding).

2) Conditioning on a common effect of two variables (a collider) creates or contributes to the association between the colliding variables.

Figure 1 shows a possible, realistic, causal diagram according to which RD1 affects RDI2 whereas RDI2 and BP2 share a common cause, U (a confounder of the effect of RDI2 on BP2.) A causal chain from RDI1 to BP2 via BP1 is also depicted.

Figure 1. A theoretical casual diagram for RDI and BP, when measured at two time points.

Since RDI2 is a collider on a path from RDI1 to BP2 (via U), we should *not* condition on RDI2 when we try to estimate the effect of RDI1 on BP2. That conditioning violates a key rule of directed acyclic graphs for estimating effects,[3] because it opens a non-causal associational path between RDI1 and BP2 (RDI1—U→BP2).
Thus, when heterogeneity of the two cross-sectional associations is allowed, we should ignore the coefficient of RDI\(_1\). Only the coefficient of RDI\(_2\) is valid! So what do we gain from allowing such heterogeneity? Or in other words: what do we gain from conditioning the association of RDI\(_2\) with BP\(_2\) on RDI\(_1\)? We may gain the blocking of confounding paths via RDI\(_1\), such as the two back-door paths (thick arrows) that are shown in Figure 2. When we regress BP\(_2\) on RDI\(_2\) alone (a cross-sectional association at time 2), these confounding paths remain open.

![Figure 2](image)

**Figure 2.** A theoretical causal diagram according to which the cross-sectional association of RDI\(_2\) and BP\(_2\) contains confounding by RDI\(_1\)

**Other views**

Most textbooks on longitudinal data analysis recognize, of course, the benefit of change models for control of time-stable confounders, but they apparently follow the assumption that “a longitudinal change” is a different causal idea from “a cross-sectional change”.

For example, the authors of one book [4] seem to suggest *conceptual* difference between what they call \(\beta_C\) (the cross-sectional effect; my \(\beta_1\)) and what they call \(\beta_L\) (the effect of
longitudinal change according to their parameterization; my $\beta_2$). The following citations illustrate their viewpoint:

“…the prime advantage of a longitudinal study is its effectiveness for studying change.” (4, page 17)

“…the major advantage of a longitudinal study is its ability to distinguish the cross-sectional and longitudinal relationships between the explanatory variables and the response.” (4, page 23)

Like others, these authors acknowledge the benefit of controlling time-stable confounders and recommend conditioning on the baseline causal variable:

“…one may simply view $X_{i1}$ [my RDI_1] as a confounding variable whose absence may bias our estimate of the true longitudinal effect [my $\beta_2$, the coefficient of RDI_2].” (4, page 24).

Evidently, the effect that is confounded in their mind is “the true longitudinal effect”. As shown here, however, that longitudinal effect is simply the cross-sectional effect at time 2, after controlling for all time stable covariates and the baseline causal variable.

In another textbook, the authors dismiss altogether the studying of change between two time points “because two-wave studies…confound true change with measurement error.”
(5, page 10.) Apparently, the key benefit of two wave studies—controlling of all time-stable confounders—is not deemed that important to these authors.

Conclusion

The two models of the effect of change in RDI on change in BP simply estimate coefficients of cross-sectional associations. They offer control of time-stable variables, which is not offered by any isolated cross-sectional association, and that’s the main reason for preferring these models. If heterogeneity of the cross-sectional associations is allowed, then only the coefficient of RDI$_2$ should be reported, and in that case, we are also controlling all confounding paths via RDI$_1$. Most important, and contrary to prevailing thought, a model of changes between two time points does not estimate any special causal idea of “longitudinal change”.
References


