Longitudinal Studies in Organizational Stress Research: A Review of the Literature With Reference to Methodological Issues

Dieter Zapf and Christian Dormann
University of Konstanz

Michael Frese
University of Amsterdam

Demonstrating causal relationships has been of particular importance in organizational stress research. Longitudinal studies are typically suggested to overcome problems of reversed causation and third variables (e.g., social desirability and negative affectivity). This article reviews the empirical longitudinal literature and discusses designs and statistical methods used in these studies. Forty-three longitudinal field reports on organizational stress were identified. Most of the investigations used a 2-wave panel design and a hierarchical multiple regression approach. Six studies with 3 and more waves were found. About 50% of the studies analyzed potential strain-stressor (reversed causation) relationships. In about 33% of the studies there was some evidence of reverse causation. The power of longitudinal studies to rule out third variable explanations was not realized in many studies. Procedures of how to analyze longitudinal data are suggested.

Occupational stress research has been quickly developing during the last two decades. This has led to many excellent studies—many of them longitudinal—that produced a wide range of knowledge in this field. However, even though this development is welcomed, we argue that the power of longitudinal designs has not yet been fully realized. This is because designs and analyses of the stress investigations are far from optimal. Thus, the review of the literature in this article on longitudinal stress research from a methodological angle is supposed to show the pitfalls and difficulties in design and analysis strategies but should also point out good examples of longitudinal research. Obviously, we are not the only ones who present a critique of stress research (e.g., Brief, Burke, George, Robinson, & Webster, 1988; M. J. Burke, Brief, & George, 1993; Contrada & Kranz, 1987; Frese & Zapf, 1988; Kasl, 1978, 1986, 1987; Kessler, 1987; Leventhal & Tomarken, 1987; Spector, 1992; Zapf, 1989). However, this critique differs from other articles because it is oriented toward that part of the literature that renders the best data and the most knowledge: the longitudinal studies.

Dieter Zapf and Christian Dormann, Faculty of Social Sciences, University of Konstanz, Konstanz, Germany; Michael Frese, Department of Psychology, University of Amsterdam, Amsterdam, The Netherlands.

We thank Harry Garst for a critical reading of a prior version of this article.

Correspondence concerning this article should be addressed to Dieter Zapf, Sozialwissenschaftliche Fakultät, Universität Konstanz, Postfach 5560, D-78444 Konstanz, Germany. Electronic mail may be sent via Internet to dieter.zapf@uni-konstanz.de.

In this article we first argue that because stress research is a multicausal field, causal effects of specific stressors on strain cannot be very high. Second, we summarize those methodological aspects that make it difficult to detect causal effects. Third, we summarize the problems of causal inference that are associated with typical research designs and applied statistical procedures. Fourth, we review existing longitudinal studies to assess how they addressed the question of causal inferences.

On the Size of Stressor-Strain Relationships

Some authors have argued (e.g., Cohen & Edwards, 1989; Kasl, 1978; Nelson & Sutton, 1990; Rabkin & Struening, 1976; Vossel, 1987) that investigators should be disappointed by the small amount of explained variance in stressor-strain correlations. In contrast to these authors, we argue that a small correlation should be expected both from a content and from a methodological point of view (cf. also Leventhal & Tomarken, 1987; Semmer, Zapf, & Greif, in press). Many factors influence physical and mental health. Among the factors discussed in the literature are physical constitution, past diseases or accidents (Kasl, 1983; Lipowski, 1975), personality traits (Depue & Monroe, 1986; Semmer, in press), health behaviors such as smoking or drinking (Sieptime & Wardle, 1995), leisure time stressors (Bamberg, 1992), family stressors (Gutek, Repetti, & Silver, 1988), environmental factors such as air pollution (Seaber & Irgren, 1992), age (Birdi, Warr, & Oswald, 1995; Warr, 1992), sex (Frankenhaeuser, 1991), social class related factors (such as
of explained variance in certain subsamples but attenuate the effects in the total sample (cf. Wall et al., in press).

Another reason for attenuated stressor-strain relationships is the healthy worker effect (Waldron, 1982). Because it is likely that workers who are seriously ill will stop working, there is a restriction of range in the dependent variable (health). A similar effect occurs when a worker quits the job because he or she feels that the job is too stressful (cf. Kessler, 1987).

Another group of reasons that leads to low empirical causal stressor-strain relationships is the specific time relationships of stressors and strains (Frese & Zapf, 1988). In that article the authors presented different models of how a stressor may affect ill health in the course of time. The authors discussed five types of the so-called exposure time model, which assume that a stressor has some impact on psychological and psychosomatic dysfunctioning, however, in various ways:

1. According to the stress reaction model the impact of a stressor increases psychologically dysfunctioning with exposure time. Once the stressor is removed, there is an improvement in psychological functioning. An assumption of this model is that the increase of a stressor would have the same impact on psychological dysfunctioning as its decrease, simply the sign of the impact should be reversed.

2. According to the accumulation model, strain comes about as a result of the accumulation and it does not go away even after the stressors have been reduced. Such effects might appear in shift workers after a certain "breaking point" has been achieved (Frese & Okonek, 1984). The mechanism of the accumulation model must be considered because longitudinal designs rely on changes of the respective variables. According to the accumulation model, however, increases of stressors should have to be treated differently from decreases (see below).

3. The dynamic accumulation model, in contrast to the accumulation model, assumes that there is an inner dynamic that leads to a further increase even after the stressor has been removed, although this increase is probably decelerated. This may be so because the original stressor had a general weakening effect on the psychophysical system so that new stressors have a higher impact than normal. In such a case an individual may also be more vulnerable to stressors not related to the work situation at all (e.g., in his or her leisure time; Bamberg, 1991).

4. The adjustment model is related to the stress reaction model because there is at first a linear increase of dysfunctional stressor. However, after process sets in and although the stressor itself can be described theory (Lazarus & Folkman, 1984), coping strategies (coping help seeking), which are control for coping strategies) one multiple regression that underestimates. For example, Schon item in a sample of 80 did not yet accommodate.

The sleeper effect is a long time lagging as for example, in (Leymann & Gustafsson, 1978). It is possible that the stress when dysfunctioning an effect is higher with less intensity of the stressor.

Most of the longitudinal item typically used (regression or analysis imply a parallelism between struts. That is, if a stressor goes up, the units, if a strain goes similarly as above-mentioned time, it not necessarily so. A decline of strain if the models suggest to try stressors differentially would then expect causal stressors, and, for example in case of decreasing stress between developing a Thus, linear data are underestimated the true association. We are not considering the assumptions. In all, there are many organizational stress is effects are not high in the empirical design variables, measuremental procedures can easily detect causal relation mind when we discuss designs and statistical.
increase of dysfunctioning with the duration of the stressor. However, after a certain point, an adjustment process sets in and the dysfunctioning decreases although the stressor is still present. The adjustment model can be described quite well within Lazarus’ theory (Lazarus & Folkman, 1984): One develops coping strategies toward the stressors (e.g., denial or help seeking), which reduce ill health. If one does not control for coping strategies (or uses job tenure as an indirect control) one may find stressor-strain relationships that underestimate the true causal association. For example, Schonfeld (1992) controlled for job tenure in a sample of newly employed teachers who did not yet accommodate to the stressors.

5. The sleeper effect model implies that dysfunctioning appears a long time after exposure to the stressor, as for example, in post-traumatic stress disorders (Leymann & Gustafsson, in press; Theorell, Leymann, Jodko, Konarski, & Norbeck, 1994). It is possible that the stressor is not present any more when dysfunctioning appears. It also assumes that the effect is higher with longer exposure to and stronger intensity of the stressor.

Most of the longitudinal studies reviewed for this article typically used some sort of linear models (regression or analysis of variance). These methods imply a parallelism between changes in stressors and strains. That is, if a stressor goes up for x units, strain goes up for y units. If a stressor goes down for x units, strain goes similarly down for y units. As some of the above-mentioned time course models suggest, this is not necessarily so. Some models do not assume a decline of strain if the stressors are reduced. Other models suggest to treat the time frame of increasing stressors differently from decreasing stressors. One would then expect causal relationships for increasing stressors, and, for example, zero relationships in the case of decreasing stressors of different time frames between developing and recovering from ill health. Thus, linear data analysis methods would usually underestimate the true strength of the stressor-strain association. We are not aware of articles taking such considerations into account.

In all, there are many reasons why causal effects in organizational stress research cannot be very high. If effects are not high to begin with, then problems of the empirical design such as a wrong time lag, third variables, measurement issues, or insensitive statistical procedures can easily contribute to the failure to detect causal relationships. This has to be kept in mind when we discuss strengths and weaknesses of designs and statistical methods below.

Problems of Testing Causal Effects in Longitudinal Studies

The bulk of empirical research (clearly more than 90%) on stressors and health is cross-sectional. The weaknesses of such a design are widely acknowledged and researchers are well aware that it is usually impossible to demonstrate causal relationships in such designs. It has been suggested that longitudinal studies can reduce the problems associated with cross-sectional studies, in particular, the problem of reverse causation and the treatment of third variables.

In line with Cook and Campbell (1979), we speak of a causal effect as existing (a) if there is covariation of the stressor with ill health, (b) if the stressor appeared before ill health developed, and (c) if other plausible explanations can be ruled out (e.g., a weakening of the body’s constitution that led first to more stress and later to an increase in the manifest ill health). Thus, causal inferences cannot be proven (H. M. Blalock, 1961; Dywer, 1983; Holland, 1986) but can be made plausible by ruling out alternative explanations.

Longitudinal studies have one important restriction, namely, that it makes no sense to assume that a variable at Time 2 has an impact on a variable at Time 1. This restriction can be used to argue for a causal inference under certain circumstances. If there is a relationship between two variables, there are basically only two alternative explanations to a causal inference of x on y: First, there is a reverse causation of y on x, and second, a third variable z influences both x and y and thus produces the relationship between x and y. Typically, longitudinal research is seen to be able to exclude reversed causal relationships hypotheses. However, as argued below, interpretation of longitudinal data are also not immune against third variable problems.

Reverse Causal Hypothesis

For the reversed causation hypothesis, there are at least two plausible groups of explanations possible: the so-called drift-hypothesis and the true strain-stressor hypotheses. First, according to the drift-hypothesis (Fresen, 1982; Kohn & Schooler, 1983) individuals with bad health drift to worse jobs, for example, first by becoming unemployed and then by getting a worse job because of their personal record of frequent absenteeism. Or, people with high absenteeism are transferred to positions with less responsibilities, which go along with higher work stressors. Moreover, in selection, preferred employees are those with higher levels of social competence, self-esteem.
and stress tolerance for skilled jobs. Thus, healthier people get the better jobs. In a longitudinal study this would lead to a causal impact of strain on stressors. Second, stressors may sometimes be affected by strain, as exemplified in the relationship between social stressors and depression. One could argue that the level of depression is related to the quality of important personal relationships. Depressed people tend to assess their environment more negatively, thus contributing to a more negative group climate (Beck, 1972). This may cause an increase in conflicts between coworkers in the group, leading to higher social stressors at work.

Thus, reversed causal influences of strains on stressors are plausible (Leventhal & Tomakren, 1987). There are also cases where reciprocal causal relationships seem plausible (positive feedback loop). For example, an increase of social stressors caused by the worker's depression can contribute in turn to an increase of depression.

**Third Variables**

The other problem for the interpretation of causal relationships is third variables. Third variables may affect stressors and strains by using the same methods, thus producing common method variance, for example, through social desirability, acquiescence, or negative affectivity. Other third variables affect stressors and strains independently from the method used, for example social status (bad housing conditions, higher environmental pollution, financial problems), education, sex, and age. For longitudinal research the stability of third variables over time is crucial. Occasion factors, background variables, and nonconstant variables can be differentiated.

**Occasion factors.** Occasion factors are hypothetical and usually unmeasured variables that have an impact on stressor (independent) and strain (dependent) variables. As Dwyer (1983) emphasized, such "occasion factors are the Achilles' heel of cross-sectional designs, where they must be measured and controlled explicitly to avoid bias in the estimation of structural coefficients" (p. 360). Examples of occasion factors are weather, time of the day, or mood variables. If participants of a study are in a good mood, they might, for example, see themselves as less depressed or report less psychosomatic complaints and simultaneously assess the work stressors as less pronounced in contrast to participants in a bad mood who may exaggerate strain and stressors. This creates an artificial correlation between stressors and strain. Because it is unlikely that participants are in a similar mood again when they answer the same questionnaire several months later, there are no true occasion factor correlations over time.

While influences of mood can exaggerate cross-sectional stressor-strain relationships, there is a twofold effect for longitudinal studies: On the one hand, mood works like error variance, thus, attenuating the observed effects. On the other hand, a part of the effect is carried by the stability of the dependent variable. Assume for example a true zero correlation between stressor $x_1$ and strain $y_1$, and no causal effect of $x_1$ on $y_2$, but an effect of mood at Time 1, which leads to an observed correlation of .30 between $x_1$ and $y_1$. If the stability of the dependent variable ($y_1$, $y_2$) is .50, then one would still find an observed correlation of .15 (.30 times .50) between $x_1$ and $y_2$, although the true effect is zero. This effect, however, disappears when $x_1$ is partialled out from the correlation between $x_1$ and $y_2$.

**Background variables.** Another type of third variables is assumed to be completely stable over time. They are called background variables (Dwyer, 1983), particularly sociodemographic variables such as age, sex, education, social status (Frese, 1985), or personality traits such as negative affectivity (Brief et al., 1988). The effects of background variables are carried over time; that is, the correlation between stressors Time 1 and strain Time 2 is as exaggerated as the correlation of stressors Time 1 and strain Time 1. However, by partialling out strain Time 1, as is done in hierarchical multiple regressions, such third variable effects are controlled for.

**Nonconstant variables.** The most problematic type of variables are those that have some stability over time and that influence both the independent and dependent variable (Dwyer, 1983). One can argue that social desirability does not have to be constant over time (although it will always have a certain amount of stability). We assume that social desirability is influenced by a person's sense of insecurity. Sense of insecurity may vary over time because of reasons in and outside the person, and, therefore, a similar variation in social desirability will occur. This means that varying social desirability may affect stressors and strains differentially over time.

We refer to background variables and to nonconstant variables that have not been explicitly measured as common factors (Dwyer, 1983).

**Strategies to Test Causal Effects**

In the following discussion we summarize the reasons why longitudinal designs do not automatically prove causality and do not automatically reject third variable effects. Campbell (1979; 1987; Kessler & typical strategies can concretize what the analysis strategies (1985). This study, psychological job complaints based lag of 1-2 month workers. Both p measure of time accidents, organi tal stressors; on the form, which is a an 1985, is analyzed in details, see Frese.

**Stressor Time 1**

Some authors (1989) have analyzed measured the stress and strain at Time 1 or multiple regression data (cf. F). It does not have a cross-sectional design measured later if the causal relationship social status and social status w income, nobody had a causal imp would like, for a hypothesis in strategy would fit if of .30 between c. 2. This would hypothesis was
aggregate cross-pressures, there is also, thus, attenuation; a part of the dependent variable (y₁, y₂) is not causal effect (Time 1), or correlation zero between x₁ and y₂, although the observed correlation between y₁ and y₂, although the result of questionnaires is a true measure of time pressure, uncertainty, danger of accidents, organizational problems and environmental stressors; the results of the questionnaires form, which is one of the three measures by Frese, 1985, is analyzed here and psychosomatic complaints were measured with a questionnaire (for details, see Frese, 1985).

**Stressor Time 1 and Strain Time 2 Designs**

Some authors (e.g., Richter, Schirmer, & Dettmar, 1989) have analyzed causal effects with designs that measured the stressors at Time 1 (referred to as x₁) and strain at Time 2 (referred to as y₂). Correlational or multiple regression analyses were used to analyze such data (cf. Figure 1). Such an analysis strategy does not have advantages compared with a cross-sectional design. The fact that a variable was measured later than another variable is little proof of a causal relationship. Consider, for example, a father's social status and his son's income. Even if the father's social status was measured years after the son's income, nobody would assume that the son's income had a causal impact on the father's social status. If one would like, for example, to find support for the drift hypothesis in Frese's (1985) data, this analysis strategy would find a significant (p < .01) correlation of .30 between complaints Time 1 and stressors Time 2. This would be taken as evidence that the drift hypothesis was supported. However, as we shall see,

![Figure 1. Effects of x₁ on y₂.](image)

![Figure 2. The hierarchical multiple regression approach: Regression of y₂ on y₁ and x₁.](image)
pressure on anxiety with a time lag of 1 year, a model comprising a synchronous path of time pressure on anxiety at Time 2 would approximate the true effect better than a model with a 10-year lagged effect of time pressure Time 1 on anxiety Time 2. In studies that use a time lag of 1 year, a true 3-months time lag should be better represented by synchronous effects than by lagged effects. Therefore, not being able to test these types of synchronous stressors-strain effects is a serious weakness of this approach.

4. In addition to the objections just mentioned, there are further problems such as assumptions of uncorrelated measurement errors (cf. Dwyer, 1983).

Applying hierarchical multiple regression to Frese's (1985) data leads to the results in Table 1: There is a significant effect of stressors on complaints but no effect of complaints on stressors. The regression analyses show that the correlation between complaints Time 1 and stressors Time 2 can be explained by the cross-sectional correlations and the stabilities of the variables.

**Full Two-Wave Panel Design**

Several strategies are used to analyze data of a two-wave panel design: (a) cross-lagged panel correlation analysis; (b) hierarchical regression with lagged effects, hierarchical regression with synchronous effects, and comparing the regressions of y₂ and x₂; (c) simultaneous estimation procedures such as structural equation approaches.

**Cross-lagged panel correlation (CLPC) technique.** The CLPC technique (R. Blalock, 1962; Campbell, 1963; Campbell & Stanley, 1963) comprises six correlations: the cross-sectional correlations at Time 1 and Time 2, the stabilities or autocorrelations \(r(x_1, x_2)\) and \(r(y_1, y_2)\); and the cross-lagged correlations \(r(x_1, y_2)\) and \(r(y_1, x_2)\). The core element of the CLPC technique is the statistical comparison of the two cross-lagged correlations of \(r(x_1, y_2)\) and \(r(x_2, y_1)\) in Figure 3, for example, tested with Steiger's (1980) formula 15. Researchers who use this technique are searching for the causal predominance of either of the variables.

Although the logic of CLPC is intuitively appealing, several articles have argued against using the CLPC technique (e.g., Dwyer, 1983; Feldman, 1975; Link & Shroot, 1992; Locascio, 1982; Rogosa, 1980; Williams & Podskakov, 1989).

It can be algebraically shown that the difference of the cross-lagged correlations is directly dependent on the stabilities of \(x\) and \(y\). Thus, a difference in cross-lagged correlation opposite to the true direction can appear if the difference in stability path coefficients more than offsets the difference in cross-lagged causal parameters (Locascio, 1982, p. 1031). Other problems are differences in variances and differences in cross-sectional correlations indicating that assumptions of the CLPC approach are not met (see Rogosa, 1980, or Williams & Podskakov, 1989, for algebraic details).

It has also been shown that CLPC is only able to reject certain types of third variable models. A significant difference of the cross-lagged correlations leads to the conclusion that the correlation between \(x\) and \(y\) is not only due to a synchronous common factor. Critical assumptions that have to be made for this conclusion is that the causal influence does not change over time and the common factor exerts its effects on both variables with identical time lags. The CLPC technique has occasion-factor relationship of although differ sometimes in assessment of time so because the less intuitive in comparison more difficult to detect the cross-laggin technique does lag, although it is interpreted as lags. Rather, the effects.

Suggestions for stability (e.g., opinion, implicit) logic that could regression, which

Applying (a) has already been done be repeated h exclusivity causal structure; variances over were partial CLPC proceed on to lagged correlatio; 1 – complaint Time 1 – stress effect of psy complaints.

**Hierarchical structure of analysis of log data.** By including the more profound explanation. However, the causal effects Comparisons sions with s suggested to ships. In such are put into the following ord (c) stressor Ti is prone to oc not neutraliz

---

**Table 1**

*Application of the Hierarchical Multiple Regression to the Frese (1985) Data (Standardized Regression Coefficients)*

<table>
<thead>
<tr>
<th>Independent variable</th>
<th>Dependent variable: Complaints at Time 2</th>
<th>β</th>
<th>Independent variable</th>
<th>Dependent variable: Stressors at Time 2</th>
<th>β</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complaints at Time 1</td>
<td>.66**</td>
<td></td>
<td>Stressors at Time 1</td>
<td>.64**</td>
<td></td>
</tr>
<tr>
<td>Stressors at Time 1</td>
<td>.19**</td>
<td></td>
<td>Complaints at Time 1</td>
<td>.08b</td>
<td></td>
</tr>
</tbody>
</table>

* Adjusted \( R^2 = .56 \).
* Adjusted \( R^2 = .41 \).
* \( p < .05 \).
* \( p < .01 \).
The CLPC technique is, however, not able to reject occasion-factor models that lead to an artifactual relationship of the variables at Time 2. Moreover, although differences in cross-lagged correlation can sometimes indicate a causal interdependence, the assessment of the direction of the causal influence is exclusively based on theoretical assumptions. This is so because the logic behind this technique is much less intuitive as it seems to be from the simple comparison procedure (see Kenny, 1975). Also it is difficult to determine an effect size from the results of the cross-lagged panel technique. Finally, the CLPC technique does not provide information on the time lag, although in practice, significant effects are mostly interpreted as lagged effects by most of the researchers. Rather, they can also stem from synchronous effects.

Suggestions have been made to partial out the stabilities (e.g., Polz & Andrews, 1964). This, in our opinion, implies a shift from the CLPC logic toward the logic that is inherent in hierarchical multiple regression, which also tests the reversed hypothesis.

Applying CLPC to the Frese (1985) data has already been done in the original paper and needs not be repeated here. Because stationarity (i.e., a stable causal structure over time) and equality of the variances over time existed (in addition, stabilities were partialled out), it was legitimate to apply a CLPC procedure. There were significant effects of stessors on complaints, shown by comparing the lagged correlations of $r = .47$ (stessor Time 1 - complaints Time 2) with $r = .30$ (complaints Time 1 - stessors Time 2). Thus, there is a causal effect of psychological stessors on psychosomatic complaints.

Hierarchical regression analyses. The general structure of multiple regression equations for the analysis of longitudinal data has already been briefly described. The effects of third variables are controlled by including them into the regression in the first step. The more potential third variables are included, the more explanations of spuriousness can be ruled out. However, there is still the problem that synchronous causal effects cannot be identified by this approach. Comparisons between reversed hierarchical regression with synchronous effects have also been suggested to identify synchronous causal relationships. In such regressions, the independent variables are put into the hierarchical regression in the following order: (a) third variables, (b) strain Time 1, (c) stessor Time 1, (d) stessor Time 2. This approach is prone to occasion factors because their influence is not neutralized as in the case of lagged effects. However, it is advantageous that changes in the independent variables are related to changes in the dependent variable. Although such a design allows for comparisons of the stessor-strain hypothesis with the reversed hypothesis, this is often not done (e.g., Billings & Moos, 1982) because the opposite hypothesis is a priori rejected by the authors.

Structural equation approaches. Structural equations approaches (most often referred to as LISREL analyses; Jöreskog & Sörbom, 1989, 1993) have several advantages with respect to the other analytical approaches discussed so far, although they are sometimes criticized (e.g., Brannick, 1995, for application problems and wrong interpretations).

Four advantages of LISREL models should be mentioned:

1. Measurement errors can be accounted for by the introduction of measurement models. Causal relationships between variables are modeled on the basis of latent constructs that are considered to be error free.

2. Structural equation models allow simultaneous estimates of causal relationships for all latent variables. Thus, multivariable-multiwave models can be analyzed.

3. Reciprocal relationships can be introduced into the models.

4. Various method and third variable problems can be modeled such as occasion factors and common factor models that account for effects of unmeasured third variables.

In short, everything that can be done with cross-lagged panel correlations and regression analyses can also be done with structural equation models (see introductions by Dwyer, 1983, Kessler & Greenberg, 1981, or Williams & Podsakoff, 1989). A reanalysis of Frese’s (1985) data is presented below because it also looks at the issue of spuriousness.

The Treatment of Third Variables in Longitudinal Studies

The treatment of third variables in longitudinal designs is not easy. Whereas in cross-sectional studies researchers are well aware of this problem, little attention is paid to ruling out third variable explanations in longitudinal investigations (Link & Shroot, 1992). Effects in two-variable panel designs can always be explained by models that assume no causal relationship between these variables. That is, without additional assumptions, it is difficult to demonstrate unambiguously that causal relationships exist in two-wave panel studies (Dwyer, 1983; Link & Shroot, 1992).
A typical path model representing causal effects in a two-wave design was shown in Figure 3. An alternative model representing spuriousness in such a design is presented in Figure 4.

On the basis of statistics alone, one cannot reject the spuriousness model of Figure 4 (Dwyer, 1983; Kenny, 1975, 1979; Link & Shout, 1992) because seven parameters (a–g in Figure 4) are involved in the associations of the observed variables x₁, x₂, y₁, and y₂. However, a cross-legged panel design provides only six observed correlations. Consequently, the model in Figure 4 cannot be estimated unambiguously from the empirical data. Such a model is unidentified. As a result, more than one causal model is consistent with a set of observed relationships. Therefore, any assessment of causality of a panel design rests on the logic of the inquiry and the persuasiveness of tests proposed to rule out alternative hypotheses (Link & Shout, 1992).

In LISREL analyses several theoretical models can be built and tested against each other (Bollen, 1989; Jöreskog & Sörbom, 1989). We have reanalyzed Frese’s (1985) data using the LISREL approach and tested a series of models. The results are presented in Table 2. Model A, the null model, encompasses all variables and provides a correlation at Time 1, but no lagged or synchronous causal paths between stressors and complaints. In Models B through E, there is one additional causal path in each case: a lagged causal effect of stressors on complaints in Model B, the reverse lagged causal effect of complaints on stressors in Model C, a synchronous causal effect of stressors on complaints in Model D, and a synchronous reverse effect in Model E. The comparison of the overall fit indexes shows that Model D with a synchronous causal effect of psychological stressors on psychosomatic complaints has an excellent model fit with the goodness of fit and adjusted goodness of fit equal to 1. A comparison of the four causal Models B through E shows that the models with an effect of stressors on complaints generally fit better than the models with an effect of complaints on stressors. It should be noted that the result of the CLPC analysis in the original article led Frese to conclude that there is a lagged effect of stressors on complaints. The LISREL analysis clearly demonstrates that there is a synchronous effect. We believe that this is a typical finding: CLPC results are usually interpreted to indicate lagged effects if the lagged correlations are significantly different. LISREL analysis, however, is able to reveal synchronous effects.

In addition, Table 2 comprises five third variable models that is only a selection of many more possible models. The first four models assume that stressors and complaints are indicators of one latent construct (cf. z in Figure 4). In addition, there are the following restrictions: The variances of the latent variables are constrained to 1 (i.e., the variance does not change over time); in addition, there are equal autoregressions of stressors and complaints (f = g in Figure 4). The four models differ with respect to stationarity and the stability of the common factor (i.e., the latent third variable). Model 1 comprises proportional stationarity and free stability (i.e., the ratio of the factor loadings of stressors and complaints are equal; \( ab = cd \) in Figure 4), and the stability of the common factor is estimated (e in Figure 4). Model 2 differs from Model 1 in that there is perfect stationarity (i.e., the factor loadings of stressors and complaints on the common factor are equal at Time 1 and Time 2; \( a = c \) and \( b = d \) in Figure 4). Models 3 and 4 differ from Model 1 in that there is one common factor that does not change over time (\( e = 1 \) in Figure 4).

In terms of content, the models described correspond, for example, to the critique of observed overlap by Kasl (1978): Stressors and complaints are indicators of a time-variant construct (without chain of mediation). A critique of W. M. J. (1988), and M. M. T. (1988) is that personality traits are a good predictor for stressors-stressors and complaints-stressors of the Frese (1985) models show. Model D with complaints. N indicates that the models in Table 2 are not only defined as the original lagged effect complaints. If the 

![Figure 4. Spuriousness model.](image-url)

Table 2

<table>
<thead>
<tr>
<th>Model</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Null model</td>
</tr>
<tr>
<td>B</td>
<td>Stressor 1 — Complaints</td>
</tr>
<tr>
<td>C</td>
<td>Complaints — Stressor 2</td>
</tr>
<tr>
<td>D</td>
<td>Stressor 1 — Complaints — Stressor 2</td>
</tr>
<tr>
<td>E</td>
<td>Complaints — Stressor 2</td>
</tr>
</tbody>
</table>

Synchronous constraints:
- Proportionality:
  - Perfect: 2
- Autoregression:
  - Equal: 3

Note: GFI = Goodness of Fit Index

* See Link and Shout (1992)

* Indicates statistically significant effects

** See Dwyer (1983)

### A Review of Frese’s Study
In the beginning, Frese (1985) assumed the need for longi...
Table 2
Reanalysis of the Frese (1985) Data Using Structural Equations

<table>
<thead>
<tr>
<th>Model and type</th>
<th>Overall fit index scales</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\chi^2$</td>
</tr>
<tr>
<td>Causal models</td>
<td></td>
</tr>
<tr>
<td>A. Null model</td>
<td>9.75</td>
</tr>
<tr>
<td>B. Stressor $\rightarrow$ Complaints 2</td>
<td>4.12*</td>
</tr>
<tr>
<td>C. Complaints 1 $\rightarrow$ Stressor 2</td>
<td>9.49</td>
</tr>
<tr>
<td>D. Stressor 2 $\rightarrow$ Complaints 2</td>
<td>0.27**</td>
</tr>
<tr>
<td>E. Complaints 2 $\rightarrow$ Stressor 2</td>
<td>8.65</td>
</tr>
<tr>
<td>Third variable models</td>
<td></td>
</tr>
<tr>
<td>Synchronous common factor models*</td>
<td>4.44</td>
</tr>
<tr>
<td>1. Proportional; free\b</td>
<td>5.30</td>
</tr>
<tr>
<td>2. Perfect; free\b</td>
<td>7.31</td>
</tr>
<tr>
<td>3. Proportional; 1\b</td>
<td>7.70</td>
</tr>
<tr>
<td>4. Perfect; 1\b</td>
<td>5.71</td>
</tr>
</tbody>
</table>

Note. GFI = goodness of fit; AGFI = adjusted goodness of fit; sRMR = standardized root mean squares residuals; AIC = Akaike information criterion.
* See Link and Strout (1992).
\b Indicates stationarity and stability of common factor.
\c See Dwyer (1983).

* $p < .05$. **$p < .01$. (Significantly better than the Null model.)

indicators of a latent construct that either changes over time or does not change. The latter model (without change), for example, corresponds to the critique of Watson and Clark (1984), Brief et al. (1988), and M. J. Burke et al. (1993) that a (stable) personality trait "negative affectivity" is responsible for stressor-strain correlations. In the particular case of the Frese (1985) data, none of the third variable models show comparable fit indexes as the causal Model D with its synchronous effect of stressors on complaints. Note, however, that the third variable models are more parsimonious. Most of their fit indexes are, for example, better than the parameters of the originally suggested causal Model B with a lagged effect of psychological stressors on complaints. If the correlational pattern is less clear than in the present case, third variable models usually produce similarly good fit indexes and clear decisions between the models are impossible. Finally, note that the occasion factor model (Model 5) is clearly inferior compared with the causal model of stressors on complaints (Model D). Thus, the hypothesis of mood or other occasion factors being responsible for the correlational pattern has to be rejected.

A Review of Longitudinal Studies

In the beginning of this article we mentioned the need for longitudinal studies in the area of stress at work. We wanted to know whether research had actually advanced and produced more longitudinal studies lately. With the help of database queries, review articles, and incidental knowledge of articles, we collected longitudinal studies using the following criteria: (a) they should be quantitative; (b) they should include more than one measurement point; (c) they should measure job-related variables such as work stressors, social stressors at work, job content variables, or work-related social support; (d) they should use variables of mental health as dependent variables. In the following discussion we concentrate on studies with a passive longitudinal design, that is, quasi-experimental designs are not discussed. We also excluded articles based on students' samples.

Furthermore, we excluded longitudinal studies on unemployment and health from our review. Space reasons and the fact that unemployment studies usually do not include more specific organizational stressors may justify why these articles were not considered. Moreover, it is our impression that the necessity for doing longitudinal research has been best understood in this field, and most of the reviews in the field of unemployment concentrate on the discussion of longitudinal research (cf. Feather, 1989; O'Brien, 1986; Warr, 1987). Studies that are based on repeated measurements but were not really interested to draw causal inferences were also excluded (e.g.,
Theorell, Orth-Gomér, & Eneroth, 1990). One publication was not considered because we were not able to reconstruct what the authors had really done. Finally, one article was excluded because the results of the longitudinal study were already published in another article by the same author. This literature search led to the consideration of 43 studies published in 45 articles (see Appendix A). An increase in the publication of longitudinal studies can be observed from 10 studies published until 1985, 16 studies published between 1986 and 1990, and 19 studies published from 1991 onward. However, given that the need for longitudinal designs has been emphasized repeatedly during the last two decades, and given the many advantages of using longitudinal research for the analysis of causal stressor-strain effects and the hundreds of publications in organizational stress in general, the number of longitudinal publications is less than one would expect.

Some characteristics of longitudinal studies are presented in Table 3. They are applied to the longitudinal stress studies in Appendix B. In the following, we summarize the stress studies from a methodological standpoint.

**Time Lag**

We identified time lags of 1 month (Daniels & Guppy, 1994; Theorell et al., 1994), 3 months (6 studies), 6 months (11 studies), and 1 year (13 studies). A time lag of about 18 months occurred five times and a time lag of 2 years occurred four times. There were only two studies with a lag of 5 years, three studies with a lag of 6 years, one study with a lag of 8 years, and 2 studies with a lag of 10 years (one study included more than 10 years). Thus, most studies analyzed time lags up to 1 year. We got the impression from the reports that organizational reasons were much more important for choosing a particular time lag than theoretical considerations. There were only a few publications that discussed the time lag problem in detail.

**Designs**

Most studies (25) used a full two-wave panel design (i.e., with both independent and dependent variables measured at Time 1 and Time 2). However, there were also seven investigations that measured either the independent or the dependent variable only at one time point. Five studies used a three-wave design. Finally, there were four prospective studies that collected data at Time 1 and predicted a certain event, for example, the occurrence of coronary heart disease several years later. Most of the research with a long time lag fell into this category. Within this design, two groups (event yes vs. event no) were usually distinguished and compared with the help of logistic regression or hazard analysis methods.

**Statistical Analysis**

Six studies regressed strain Time 2 on stressors Time 1 (most of them also controlled for background, nonconstant variables, or prior strain). In 17 cases hierarchical multiple regression of $y_2$ on $y_1$ and $x_1$ or some similar method was used. One investigation also considered $x_2$ in the regression of $y_2$ on $y_1$ and $x_1$ (Article 17 in Appendix B). The CLPC approach was used seven times and either Kenny’s (1975) or Pelz and Andrew’s (1964) recommendations were followed (Articles 7, 9, 14, 30, 31, 44, and 45 in Appendix B). Finally, 10 studies used structural equations (Articles 3, 11, 12, 15, 26, 27, 28, 29, 32, and 40; Articles 28, 29, and 32 omitted measurement models), mostly applying the computer program LISREL (Jöreskog & Sörbom, 1989, 1993).

**Test of Reverse Causation**

As explained above, there are basically two reasons why researchers should use longitudinal designs: First, to make a decision between hypotheses of opposite causation, and, second, to improve the possibilities to reject third variable explanations. We discuss these points in the following sections.

We were surprised to see that in many cases, the problem of reversed causal relationships between stressors and strains was not discussed. This confirms James and James (1989) who argued that reciprocal causation has not been a popular hypothesis in organizational research. It is obvious that the six studies of the type "regression of $y_2$ on $x_1$" can tell researchers little about causal effects of strains on stressors. But even studies that used a hierarchical multiple regression approach did not usually test for reverse causation hypothesis (exceptions are articles 2 and 13 in Appendix B). Even some LISREL analyses did not take reverse causation into account (e.g., Dignam & West, 1988), although they tested a series of models. The reverse effect was, of course, tested in the 7 publications using the CLPC approach,
Table 3
Causal Effects in Longitudinal Studies With Reference to Various Statistical Techniques

<table>
<thead>
<tr>
<th>Technique</th>
<th>Causal effects</th>
<th>Reversed causation</th>
<th>Reciprocal causation</th>
<th>Background variables</th>
<th>Nonconstant variables</th>
<th>Occasion factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression*</td>
<td>Calculate hierarchical regression, control for prior symptoms, and examine regression coefficients</td>
<td>Calculate reversed regressions and examine regression coefficients</td>
<td>Remains unclear</td>
<td>Measure and introduce into equation before stressors are entered</td>
<td>Measure and introduce into equation before stressors are entered</td>
<td>Partial out dependent variable at preceding time points</td>
</tr>
<tr>
<td>CLPC</td>
<td>Compare the cross-lagged correlation coefficients statistically; calculate both CLPC with and without stabilities partialled out</td>
<td>Remains unclear</td>
<td>Remain problematic if their influences change over time</td>
<td>If (proportional) stationarity is violated, influence cannot be reliably assessed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LISREL</td>
<td>Calculate models with synchronous and lagged effects and examine overall fit indexes and structural coefficients</td>
<td>Calculate reversed LISREL model with synchronous and lagged effects and examine overall fit indexes and structural coefficients</td>
<td>Calculate any combination of synchronous and lagged reciprocal models and examine overall fit indexes and structural coefficients</td>
<td>Measure and introduce into the model</td>
<td>Measure and introduce into the model; unmeasured nonconstant variables can theoretically be modelled when number of waves exceeds two*</td>
<td>Allow within-time correlation of structural disturbances</td>
</tr>
</tbody>
</table>

Note. CLPC = cross-lagged panel correlations.
* If predictor variables are measured with error, the regression approach may lead to heavily distorted coefficients and is therefore questionable when causal inferences are drawn from such results (e.g., Dwyer, 1983; Kenny, 1979).
* There seems to be no practical experience with this approach; see Dwyer (1983) for more details.
The Treatment of Third Variables

As argued before, the third variable problem is not automatically solved by doing longitudinal research. We later discuss hierarchical regression, CLPC, and structural equations approaches to this problem.

In the traditional hierarchical regression approach, third variables have to be included to control for their effects; this is similar to cross-sectional research. This was done in less than half of all cases (the studies that controlled for third variables were Articles 1*, 2*, 4*, 5*, 6*, 8*, 17*, 18*, 19*, 20*, 33**, 36**, 38**, and 40** of Appendix A; studies marked with a plus sign controlled for third variables measured only once; studies marked with an asterisk controlled for third variables measured at different time points).

In CLPC designs, the effects of variables that can reasonably be considered to be stable (age, sex, and marital or social status for shorter time lags), and that affect stressors and strain equally at Time 1 and Time 2, can be ruled out by the CLPC technique (rejection of the synchronous common factor model, cf. Kenny, 1975). However, third variables that can change over time, such as social desirability, remain a problem. As already stated, it is also not possible to rule out models that assume stressors and strains to be measures of the same underlying construct, for example of negative affectivity (at least not as long as Time 1 health measures are not partialed out). Therefore, the application of CLPC remains limited in this respect. In comparison, structural equations make it possible to test a variety of third variable models and compare them with various causal models.

In most prospective research, it was not necessary to test the reverse causation hypothesis because, for example, an acute heart disease or mortality cannot itself affect the working conditions. However, third variable explanations are as plausible as in other designs. An example: One’s reduced physical efficiency is related both to bad working conditions and to the occurrence of a heart disease. Thus, third variables have to be explicitly considered, and all prospective studies did so.

There were some reports (Dignam & West, 1988; Dornmann, Zapf, & Speier, 1995; Marcelissen et al., 1988) that modeled the occasion factors described above. Dornmann et al. (1995) added occasion factors that led to a more significant causal effect of social stressors on depression compared with a model without occasion factors. Although occasion factors are often an alternative explanation for true effects in cross-sectional research, it should be noted that the introduction of such factors may introduce new problems that are not present in the original models. Also, the use of occasion factors can lead to overfitting of the model.

Summary of Methodology

To make the results of this study more convincing, we recommend that future studies be conducted using a variety of methods. Their results can be more easily interpreted by considering the potential confounding factors. The authors who designed the study need to provide more information about the study design and the potential confounding factors. This will help future researchers to better understand the results of the study.

Leiner (1995) suggests that authors who use longitudinal data should consider the possibility of methodological and statistical confounding factors. He recommends that researchers should consider the potential confounding factors in their study design and analyze the data using appropriate statistical methods. This will help future researchers to better understand the results of the study.
problem is not
nal research.
CLPC, and
problem.
approach,
trol for their
research. This
is a wave of
tudes that can
age, sex, and
gs), and that
1 and Time
(c, ibby, Kenny,
change over
problem. As
t rule out
rens to be
struct, for
as long as
talled out).
ains limited
ations and
ariable ca
ecessary be
arity cannot
er, third
as in other
ysical effi-
itions and
r, third
red, and all

West, 1988;
lelsen et al.,
's described
factors of
ocial h
model
factors
ue effects in
ed that the
introduction of these occasion factors in longitudinal
studies may, in some cases, reveal causal relationships
that otherwise would remain undetected (cf. also Marcellissen et al., 1988, who obtained structural
coefficients higher than the original correlation when
occasion factors were introduced).

Summary of Four Model Studies

Methodological arguments are often quite abstract.
To make them more concrete, we conclude this article
by summarizing four reports that we find particularly
convincing from a methodological point of view.
Their results are of particular importance because one
can be more confident of their conclusions than of
results of studies with a lower degree of methodologi-
cal rigor.

Leitner (1993) and Frese (1985) were the only
authors who reported using observational interviews
to measure the stressors. Because much has been said
about the study of Frese (1985) already, we do not
consider it here again. Through the use of observa-
tional measures, Leitner (1993) avoided common
method variance between stressors and self-reported
health. This makes it less important to consider most
third variables. The stressor—additional effort—was
analyzed with respect to nine health indicators using
CLPC. Significant effects were found for psychoso-
matic complaints, irritation, depression, eye prob-
lems, allergy, and life satisfaction. Thus, six of nine
analyses revealed effects in the hypothesized direction.
No reversed causation was found: Controlling
for age, job tenure, school education, and professional
education did not change these findings but led to an
insignificant maximum reduction of the lagged
 correlations of .04. LISREL analyses, although only
mentioned but not reported in the Leitner article,
came to the same results.

Kohn and Schooler's (1982) LISREL study is
well-known and, therefore, can be treated very
briefly. They analyzed a full set of third variables and
calculated various models of reciprocal causation.
Because they only had data from two waves, it was
not possible to analyze all relevant models. However,
the authors skillfully used structural equations to get
the most information out of a two-wave longitudinal
design. In their final analysis, a combined distress-
well-being indicator (including trustfulness, self-
deprecation, idea-conformity, self-confidence, and
anxiety) was related to 14 different job conditions.
Six of these job conditions could be assessed as a
cause of well-being (closeness of supervision, position
in hierarchy, dirtiness, hours of work, job
protections, and job income); there were three
reversed effects (time pressure, heaviness, and "held
responsible").

Marcellissen et al. (1988) used a three-wave design
to analyze the reciprocal effects between several
strain variables and coworker or supervisor support.
Their LISREL analyses did not include measurement
models, and only synchronous but no lagged effects
were analyzed. Third variables were also not
included. However, occasion factors were considered
and separate analyses for lower and higher occupa-
tional levels were conducted. Three out of 16 causal
effects from supervisor support on strains were found:
regular health complaints (lower occupational level),
worries concerning one's job (lower and higher level),
but no reversed effects. For coworker support
the results were quite the opposite. There were no effects
from coworker support on any strain variable, but 5 of
16 reversed causal coefficients reached significance:
stressful strain (low and high level), worry concern-
ing one's job (low), diastolic blood pressure (low),
and regular health complaints (high occupational
level).

Schonfeld (1992) was the only one who controlled
for preemployment symptoms. His LISREL analyses
tested reciprocal models with different time lags but
did not incorporate an occasion factor or other third
variables. School environment stressors were synchro-
nously related to depressive symptoms but there was
no significant reversed effect.

In summary, 16 of the 50 tested causal hypotheses
and 8 of 50 tested reversed causation models of these
four studies were supported. On average, the causal
effect was .12. While this coefficient is not high, it is
in line with our arguments presented in the beginning
of this article.

However, one can point even in this set of excellent
articles to methodological considerations that are
necessary but missing: testing lagged and synchro-
nous effects, controlling for third variables, and
testing reversed and reciprocal causation (the only
exception was Kohn & Schooler, 1982, who did all of
this).

Conclusions

In summary, the lack of common standard proce-
dures to analyze longitudinal data is reflected in the
empirical articles reviewed for this article. The
concerns of Williams and Podsakoff (1989) that many
researchers unjustifiably feel that applying a longitu-
dinal design automatically solves many problems
inherent in cross-sectional approaches can be sup-
ported by our review.
The following recommendations can be made with regard to methodological issues of longitudinal stress research:

1. All variables should be measured at all time points using the same measurement method for the respective variables. If this is not done, one cannot examine reverse or reciprocal causation hypotheses that consider strain Time 2 as the dependent variable. Additionally, certain third variable hypotheses such as occasion factor hypotheses require all variables measured at each time.

2. As in cross-sectional studies, researchers should carefully consider third variables as potential confounders of the stressor-strain relationship and include them in their design. As mentioned above, structural equation models possess the capability to take unmeasured third variables into account. However, such models usually require that certain constraints be put on the models (e.g., stationarity restrictions). Because there is still little practical experience with complex common factor models and because estimation problems occur even in simpler analyses, it is better to explicitly measure third variables whenever possible.

3. The time lag should be thoroughly planned (Kessler, 1987). Simulation studies show that a time lag that is too long is less of a problem than one that is too short (cf. Dwyer, 1983). Time lags that are too short may lead to the conclusion that no causal effects exist, whereas a time lag that is too long solely leads to an underestimation of the true causal impact. It would be best to conduct a multivariate design with equal time lags. If short-term effects are adequate then it should be possible to replicate them as synchronous paths within the same model. If there are long-term effects, the appropriate time interval can be assessed by comparisons of different time lag models (e.g., Dornmann et al., 1995).

4. Assumptions about the time course (Frese & Zapf, 1988) of the variables under study should be made. For example, if strong adaptation processes are expected, then it should be wise to study participants beginning their jobs, that is, before they have the opportunity to adapt to the working conditions (cf. Schonfeld, 1992).

5. A linear structural equations approach is recommended to analyze the data (cf. Dwyer, 1983; James & James, 1989; Link & Shrout, 1992; Williams & Podsakoff, 1989).

6. Measurement models should be included in the models. Errors in measurement that may attenuate relationships across variables can be accounted for by the introduction of measurement models.

7. Multiple competing models should be tested (James, Mulaik, & Brett, 1982). The rationale for this recommendation is that models that survived competing tests with a series of alternative models can be trusted more. The following recommendations are useful: (a) To reduce the complexity, one should test measurement models separately before testing the structural model (two-step approach, cf. Anderson & Gerbing, 1988). This is especially true for longitudinal models because several measurement models that express different "behaviors" of construct-indicator relationships over time can be tested. This task becomes too complicated if done together with procedures to test causality assumptions. Although this approach has been criticized (e.g., Fornell & Yi, 1992) because measurement models can change with variations in the structural part of the equation system, our practical experience suggests that in most cases these changes are minor. (b) Causal effects including reversed effects should be systematically introduced into the models. (c) Effects of third variables should be systematically tested. (d) Occasion factors should be tested. Occasion factors are easy to model, they are always identified, and they can be combined with several kinds of causal models. This allows to test whether or not a certain causal model holds in spite of this special kind of third variables.

When we started reviewing the literature on longitudinal studies it was our hope to find a clear trend to methodologically sounder studies in recent years. This trend is not as clear as we expected. We hope that this article contributes to more systematic longitudinal designs of organizational stress studies that make it possible to test reverse causal hypotheses and a series of third variable explanations. If we have then a better understanding of causal stressor-strain relationships, then time and effort to write this article was well invested.

References


beitswissenschaft, 47, 98–107.


(Appendices follow on next page)