The untapped potential of causal inference in cross-modal research

Pan, Jian; Mahdavi, Ardeshir; Mino-Rodriguez, Isabel; Martínez-Muñoz, Irene; Berger, Christiane; Schweiker, Marcel

Published in:
Building and Environment

DOI (link to publication from Publisher):
10.1016/j.buildenv.2023.111074

Publication date:
2024

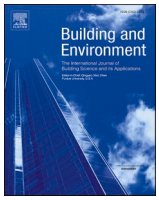
Document Version
Publisher's PDF, also known as Version of record

Building and Environment 248 (2024) 111074

Contents lists available at ScienceDirect

Building and Environment

journal homepage: www.elsevier.com/locate/buildenv



The untapped potential of causal inference in cross-modal research

Jian Pan [a,*], Ardeshir Mahdavi [b], Isabel Mino-Rodriguez [c], Irene Martínez-Muñoz [b], Christiane Berger [d], Marcel Schweiker [a]

[a] Healthy Living Spaces Lab, Institute for Occupational, Social and Environmental Medicine, Medical Faculty, RWTH Aachen University, Pauwelsstr. 30, 52074, Aachen, Germany
[b] Institute of Building Physics, Services, and Construction, Faculty of Civil Engineering Sciences, TU Graz, Lessingstraße 25/III, 8010, Graz, Austria
[c] Building Science Group, Faculty of Architecture, Karlsruhe Institute of Technology, Englerstr. 7, 76131, Karlsruhe, Germany
[d] Department of Architecture, Design and Media Technology, Human Building Interaction, Aalborg University, Rendsburggade 14, 9000, Aalborg, Denmark

ARTICLE INFO

Keywords:
Causal inference
Cross-modal research
Multi-domain research
Methodology
Estimand

ABSTRACT

Cross-modal effects have recently become a popular topic in building science. However, studies in this area frequently neglect causal inference, leading to a lack of valid causal results. To address this problem, we specifically highlight causality and its importance to cross-modal research. We present three general guidelines, and describe them using toy examples, for appropriately conducting causal cross-modal research. The guidelines originate from the methodological framework for quantitative social science by Lundberg et al. (2021). They are as follows: i) specify the theoretical estimand as the target of causal inference; ii) specify the empirical estimand that is informative for the theoretical estimand based on causal assumptions; iii) select the estimation strategy empirically to estimate the empirical estimand. In light of these guidelines, we discuss some common methodological pitfalls in current research practices that can jeopardize causal inference. Moreover, we offer certain recommendations to avoid such pitfalls. The general objective of this paper is to promote transparent causal cross-modal research by raising the awareness of causal inference in view of appropriate causality-related methodological choices.

1. Background

Human perception of the indoor environment is simultaneously influenced by multiple sensory modalities. In the past, numerous multi-domain studies have focused on thermal, visual, acoustic, and air quality domains and investigated cross-modal effects. These effects pertain to circumstances where a stimulus from one domain influences a response from another domain [1].

Multi-domain studies of people's evaluation of indoor-environmental exposure situations have been reviewed in the past. For instance, Schweiker et al. [2] reviewed 219 papers in detail. A key conclusion of their extensive review pointed to a certain level of inconclusiveness of the findings: In many instances, participants' responses could not be suggested to display signals clearly above the noise level associated with the experimental uncertainties. This inconclusiveness was furthermore underlined by the circumstance that studies of similar combination of exposure elements sometimes appeared to provide conflicting results. Another fairly comprehensive review [1] provided further indications of inconsistency in methodological approaches and documentation of multi-domain studies. Specifically, this review identified the lack of consistency in research design, study set-up, data collection, statistical analysis, and results reporting as responsible for the fact that most studies do not facilitate the generation of cumulative knowledge. Furthermore, a recent high-level analysis of past studies [3] identified multiple factors that have limited the reliability of past multi-domain studies on indoor-environmental quality. These factors included certain inherent limitations of short-term controlled laboratory studies, the insufficiently established utility of the studies for practical inquiries, the inconsistency in the use (and the absence of validation) of deployed constructs and scales, the absence of foundational theories, and the lack of consideration for the informational component of exposure situations.

In this context, the present paper suggests that many shortcomings identified by the past reviews may stem from the absence of a general tightly structured procedural approach to the specification of the concrete research targets, to the explicit methodological step toward the

* Corresponding author.
E-mail address: jpan@ukaachen.de (J. Pan).

https://doi.org/10.1016/j.buildenv.2023.111074
Received 10 July 2023; Received in revised form 22 November 2023; Accepted 25 November 2023
Available online 29 November 2023

design of empirical quantities that appropriately approximate the research targets, and to the selection and execution of suitable estimation strategies. The authors further suggest that the pursuit of causal inference methods could have provided remedies for some of the primary shortcomings identified in the aforementioned reviews. This observation, and the general absence of any specific reference to causal inference in studies reviewed by Refs. [1,2], provides the primary motivation for the present study, which aims to bring the attention to the untapped potential of causal inference in the cross-modal studies.

To start with, we posit that one fundamental problem may be underlying the inconclusiveness of cross-modal research, namely a lack of valid causal results due to the widespread neglect of causal inference. The contention is that appropriate causal inference is routinely missing in cross-modal research regarding indoor environment, resulting in studies that are difficult to interpret, offering instead a mix of spurious and causal associations.

The above-mentioned unsatisfactory status of contemporary research could stem from multiple circumstances. First, researchers might be unaware of causality and the difference between causal and predictive research. Second, researchers aware of causality might undervalue its importance for cross-modal research. Third, researchers might appreciate the importance of causality, but fail to appropriately conduct causal inference.

To address these issues, we first briefly introduce causality and differentiate causal research from predictive research. Next, we explain why causality is important for cross-modal research. Subsequently, we provide recommendations for appropriately conducting causal cross-modal research. These recommendations are informed by three guidelines that originate from recent advances in causal inference for social science [4]. The guidelines toward conducting sound causal cross-modal research are explained using toy examples from cross-modal research questions. Moreover, some common methodological pitfalls that negatively affect causal inference are discussed and related recommendations are offered.

The main objective of this contribution is to promote causal cross-modal research by drawing attention to causal inference and to inform causality-related methodological choices. The presented guidelines are meant not only to aid planning, implementing, and justifying causal research, but also to serve as a structured basis for systematic evaluations of existing and future studies in view of their potential to yield valid causal results. Using these guidelines, researchers can transparently and effectively discuss contentious research issues and engage in productive distributed collaboration among each other [5]. We believe that such collaborative evaluations and discussions can improve the consistency and cumulative depth of cross-modal research and more precisely identify research gaps, which cannot be bridged via isolated studies.

2. Introduction to causality

This section first provides a pragmatic introduction to causality and the difference between causal and predictive research. Next, we explain why causality is important for cross-modal research. Finally, we discuss the undesirable consequences of neglecting causal inference.

2.1. Causality

Over decades, causality has been under active discussion in various fields, such as philosophy, statistics, and informatics. Given the limited scope of this article, we refrain from philosophically discussing (e.g., Refs. [6,7]) or mathematically defining (e.g., Ref. [8]) the epistemologically broad concept of causality. Rather, we approach it in a simple and pragmatic way: Causality (synonym "causation") tells us about the consequences of an intervention [9–11]. An intervention actively alters the value of a variable. Given causality, a (hypothetical) intervention on a variable (i.e., the cause) will lead to a change in another variable (i.e., the consequence).[1]

As researchers in building science, we are often interested in causal questions regarding the consequences of certain environmental interventions,[2] such as the impact of increased indoor temperatures conditions on occupants' perception of thermal comfort (e.g., Ref. [12]), or the impact of traffic noise on occupants' task performance (e.g., Ref. [13]). Answers to these kinds of questions can inform design choices and technical standards that require environmental interventions (e.g., installation of shading devices to prevent overheating, or installation of acoustically performant windows to reduce traffic noise transmission). Such questions all involve causality because they target the consequences of interventions.

2.2. Directed acyclic graph and fundamental causal structures

Causal inference methods help us to appropriately investigate causal effects based on empirical data [8,9]. One popular tool that aids causal inference is the directed acyclic graph (DAG; [14]). Given the limited scope of this paper, we restrict our introduction to DAG to only the essentials that will be relevant for later sections.

DAGs visualize causal relationships with nodes and arrows (see Fig. 1 for a hypothetical example[3]). Nodes represent variables (e.g., temperature and weather). Directed arrows connect nodes from the cause to the consequence (e.g., weather causally influences indoor temperatures in naturally ventilated buildings). DAGs are non-parametric, that is, there is no assumption regarding the functional form of causal relationships (e.g., linear, polynomial, or exponential). Interactions are not explicitly depicted in DAGs, but variables that jointly influence another variable may have any form of interaction (e.g., weather and temperature may interact regarding their effects on visual comfort[4]).

Causal inference literature differentiates three fundamental causal structures: confounder, collider, and mediator (see Ref. [10] for a more detailed introduction). A confounder is a common cause for two other variables and induces a spurious (i.e., non-causal) association between these two variables. In the presence of a spurious association, the confounded effect estimates will deviate from the true causal effects. Controlling for (e.g., through statistical control as covariate, through

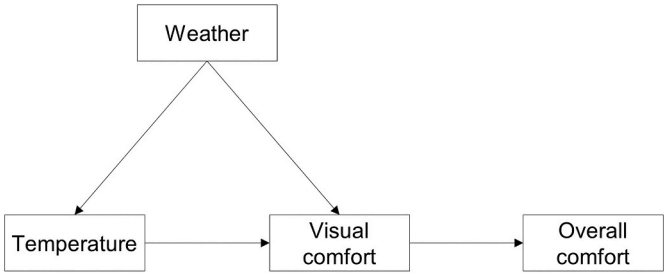


Fig. 1. Example causal assumptions regarding the relationships among indoor temperature, weather, visual comfort, and overall comfort.

[1] Following this definition, when we say variable x (e.g., temperature) causes outcome y (e.g., visual comfort), we mean that a (hypothetical) intervention on x will lead to a change in y, without implying any binary meaning (e.g., comfort or no comfort).

[2] With environmental interventions, we mean interventions on the (indoor) environment, for example opening the window or turning off the light.

[3] All examples in this paper are for demonstration purposes only and may not reflect reality.

[4] For our purposes, we understand the psychological construct "visual comfort" as an umbrella term. As a construct, its levels can range between extreme discomfort and extreme comfort. The specific visual comfort level may be caused by glare, color perception or further factors like illuminance levels.

stratification of the sample, or through selection process during data collection) the confounder will remove the spurious association. In our example, weather is a confounder because it causes both indoor temperature and visual comfort. Thus, weather induces a spurious association between temperature and visual comfort (in addition to their causal association).[5] This non-causal association can be removed by controlling for weather.

A collider is a common consequence of two causes. Contrary to the confounder, the collider only induces a spurious association between its causes when the collider is controlled for. For example, visual comfort is a collider for weather and temperature because it is caused by these two variables. By default, thermal comfort does not induce a spurious association between weather and temperature. However, when thermal comfort is controlled for, a spurious association will emerge between weather and temperature.

A mediator is the consequence of a cause and at the same time the cause for another consequence. Controlling for the mediator will remove the causal association that is transmitted (i.e., mediated) by the mediator between its cause and its consequence. For example, visual comfort is a mediator because it is the consequence of temperature and the cause of overall comfort. The causal association between temperature and overall comfort that goes through visual comfort will be removed, once visual comfort is controlled for.

2.3. Difference between causal and predictive research

In contrast to causal research that targets causality, predictive research aims to forecast outcomes as accurately as possible. For this purpose, any statistical associations, whether causal or not, are exploited [15,16]. Common threats to causal inference such as confounders and colliders are useful in predictive models because their non-causal relationships with the outcome improve predictions [10]. However, predictive research assumes, often implicitly, that predictions will be made under stable settings where there are no interventions or changes [15]. If interventions not available in the data used to develop the model are introduced into the settings, predictive models will likely become biased and misleading. In that case, predictions may have a poor performance, because predictors in predictive models do not necessarily have a causal relationship with the outcome and may not be useful for predictions under interventions [15].

Suppose we aim to predict overall comfort in an open-plan office with stable environmental settings. We could use observational data from a representative sample and develop an accurate predictive model. The model may take all available predictors, such as temperature, workers' salaries, and even their shoe sizes, into consideration and predict the workers' comfort in this office accurately. However, if we actively manipulate the environmental settings (e.g., due to installation of a new building automation system) and try to predict the resultant comfort levels, the predictive model is likely to fail, because it utilizes non-causal associations that will not hold under interventions, i.e., under the altered conditions.

2.4. Potential of causal inference in cross-modal research

Causality is important for cross-modal research due to three main reasons. First, cross-modal research often aims at informing intervention-related decisions, for example regarding how to design indoor environments. As explained above, causal results are needed when we want to predict what happens under interventions. Therefore, cross-modal research needs causal inference to achieve valid causal results.

Second, causality promotes theory development. Researchers have highlighted a lack of theories for cross-modal research (e.g., Ref. [2]). Causality helps us identify the underlying causes and understand the mechanisms of cross-modal effects. Such causal knowledge lays the foundation for theory development and guides future research. On the other hand, causality is also needed to appropriately examine proposed theories, because non-causal results might be variously biased and thus incorrectly confirm or falsify theories.

Third, causality is necessary for generalizing results. Very often, the sample in cross-modal research is not representative of the target population. This threatens the generalizability of the results from a sample. Advances in causal inference have shown that generalization depends on the causal relationships among variables and on the mechanisms underlying the difference between the study population and the target population a result is planned to be generalized for [17,18]. To get generalizable results for the target population, we need causal inference to appropriately model the sampling process and accordingly adjust the estimates (see later sections on theoretical estimand and generalizability). Thus, regardless of whether our research pursues causal effects or just descriptive differences, causal inference methods are relevant as long as the results require generalization, for example from a sample to the target population or even across different populations and settings (e.g., in cross-cultural research).

Despite the importance of causality, appropriate causal inference has been generally absent in existing cross-modal research. Consequently, existing literature demonstrates various methodological issues from the causality perspective, including confusing causal inference with predictive inference. Such confusions have led to conflicts among research aims, analysis methods, results, and interpretations. For example, a study might aim at understanding the causal effect of a thermal intervention on visual perception. However, it analyzes the data in a predictive way, for example by misusing certain machine learning techniques that are actually meant for non-causal predictions, or by adding all available variables to a regression under the incorrect assumption that the more covariates, the better the causal results.[6] The study then reports the predictive results but misinterprets them as causal.

Another common methodological issue is the application of *p*-values[7] and numerous predictive model selection criteria[8] (e.g., Bayesian information criterion [19]) for selecting the "correct" causal relationship. This practice has been repeatedly criticized by methodological literature in other fields (e.g., Refs. [10,20,21]), because these statistical tools alone do not suffice for inferring causal relationships.

These examples above also imply that randomized experiments alone, the generally known "gold standard" for causal effects, do not guarantee valid causal results because the above-mentioned issues cannot be dealt with solely by experimentation and randomization. Great efforts are still needed to appropriately handle diverse aspects related to causal inference, such as measurement errors, estimation strategies, imperfect randomization, and generalizability.

For cross-modal research, neglecting causal inference has multifold consequences. On the academic side, inappropriate causal inference can systematically bias both the direction and the size of effect estimates (see

[5] Suppose the causal effect of temperature on visual comfort is 1 unit. Given the DAG with confounder weather, if we control for weather and then estimate the effect of temperature on visual comfort, we will get roughly 1 (i.e., the true causal effect) back. However, if we don't control for weather, it will confound the effect estimate, so that we will get, say 3, back. This deviation from 3 (confounded effect estimate) to 1 (the true causal effect) is the spurious association.

[6] This typical non-causal approach is known as "garbage-can regressions" (e.g., Ref. [69]). See Ref. [22] for examples of bad control variables that bias causal inference.

[7] *p*-values are designed to control false positive rates and do not imply causation [10].

[8] Predictive model selection criteria are predictive and often choose models with spurious relationships over causal models [10].

Ref. [22] for various biases). The biases can lead to misidentified influencing factors (i.e., false positive errors) and neglected important aspects[9] (i.e., false negative errors). Biases can also lead to inconsistent findings as highlighted by Ref. [2]. Furthermore, cumulative research is threatened because it becomes unclear whether the inconsistency reflects genuine differences in causal effects, methodological problems, or a mix of both. On the practical side, resources are wasted when follow-up research is based on invalid causal results. Furthermore, invalid results will mislead intervention-related decisions, resulting in ineffective or even counterproductive measures.

3. Three guidelines for causal cross-modal research

Below, we introduce three guidelines regarding how to appropriately conduct causal inference in cross-modal research. The guidelines are largely based on the methodological framework for quantitative social science by Ref. [4]. These guidelines are presented here in a rather general manner in order to maintain their general applicability.

3.1. Theoretical estimand

The first guideline is to precisely specify the theoretical estimand[10] at the research start. A theoretical estimand, following [4], defines the target of causal inference. It states what researchers want to know. A precisely defined theoretical estimand guides the methodological choices in later steps of the research.

A theoretical estimand consists of a unit-specific quantity and a target population. The unit-specific quantity is the difference in the interested outcome under different intervention conditions for a single unit (e.g., an individual) from the target population. It may involve unobservable quantities such as latent constructs (constructs that cannot be observed or measured directly) or counterfactuals (potential outcomes under a hypothetical intervention condition). It should be independent of any statistical model (e.g., not defined as a coefficient in a regression model), as the selection of statistical models belongs to estimation strategies addressed by the third guideline. The target population is the group of units that the unit-specific quantity is aggregated over. It specifies to whom we want to apply the results from the study.

For example, we might be interested in the causal effect of temperature on visual comfort, as postulated by the hue-heat-hypothesis (e.g., Refs. [23,24]). The theoretical estimand may be the difference between an individual's latent visual comfort under 20 °C versus under 15 °C (i.e., unit-specific quantity), averaged over all German citizens (i.e., target population).

Researchers should argue why a theoretical estimand is of interest and worth investigating by linking it to research goals, previous findings, related theories, and practical implications. When specifying a theoretical estimand, researchers need to consider both the theoretical implications and practical constraints such as operationalization, sampling, and confounding. A theoretical estimand that can be straightforwardly operationalized might lack theoretical relevance, while a theoretical estimand that is theoretically important might suffer from restrictions on feasibility. Researchers need to make a balanced choice between these two aspects.

3.2. Empirical estimand

The second guideline is to specify the empirical estimand that is informative for the theoretical estimand based on causal assumptions. Following [4], an empirical estimand defines the target of the statistical analysis. In contrast to a theoretical estimand that may involve unobservable quantities, an empirical estimand involves only observable quantities. For instance, we cannot directly observe the individual-level change of latent visual comfort nor the effect for the whole German population, but we could estimate them using observable comfort ratings under several manipulated thermal conditions from a selected sample.

Because of practical constraints (e.g., limited samples) and complexities (e.g., confounders) in the real world, the data we collect generally cannot be one-to-one mapped to the theoretical estimand. Thus, researchers need to transparently specify the causal assumptions about how the empirical estimand approximates the theoretical estimand. Causal assumptions may be embodied in a generative model regarding the causal processes that generate the observable data [25]. DAGs provide an intuitive way to graphically represent generative models [10].

When making causal assumptions, researchers should strive, to the extent possible, for a causal DAG that includes all common causes (i.e., confounders), whether observed or unobserved, between any pair of included variables [26]. Researchers may follow the recommendations by Ref. [10] to derive a causal DAG: First, draw the path representing the target causal effect. Next, draw the paths for competing causes (i.e., other variables that influence the outcome). Subsequently, draw the paths representing the relationships among the causes. Finally, draw confounders, whether measured or unmeasured, among the variables.

Given a causal DAG, researchers may use software (e.g., Ref. [27]) to algorithmically[11] determine whether the target causal effect can be estimated given the causal assumptions (e.g., a causal effect cannot be estimated if it is assumed to be confounded, but we cannot adjust for the confounders because they are unmeasured). They must also determine which variables must be controlled for and which ones can or must be ignored in order to derive a valid causal conclusion (cf. confounders and colliders). Overall, causal assumptions and causal inference tools provide a principled way to design, improve, and justify an empirical estimand for the theoretical estimand.

We will illustrate empirical estimand and causal assumptions with the example regarding the effect of temperature on visual comfort. Our empirical estimand may be the difference between an individual's visual comfort rating under 20 °C versus under 15 °C, averaged over a representative sample. One key assumption for this empirical estimand to be informative for the theoretical estimand is that there are no variables that confound the relationship between temperature and visual comfort. We may represent this causal assumption in a generative model as Fig. 2, where visual comfort is caused by temperature in absence of any confounders. If data came from a perfect experiment where the temperature was precisely manipulated and everything else, such as weather and time at circadian rhythm, was kept constant or successfully randomized, this causal assumption would be plausible and our empirical estimand would be justified. However, if data came from an observational study or an experiment that did not successfully control or randomize relevant aspects, this assumption would be implausible because there are likely confounders such as weather and daytime.

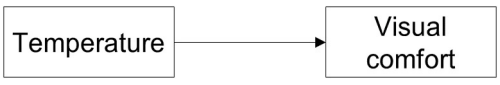


Fig. 2. Example causal assumption that temperature causes visual comfort without any confounder.

[9] Under specific conditions, a causal effect might be biased towards zero. Thus, in complement to the well-known wisdom that association does not necessarily imply causality, causality also does not always imply association.

[10] A study may have multiple theoretical estimands depending on the research questions. For better readability, this article generally uses the singular form.

[11] We refer interested readers to the d-separation [9] and the do-calculus [8] for the underlying algorithms.

To take weather into consideration, we may improve our empirical estimand to be the difference between an individual's visual comfort rating under 20 °C versus under 15 °C, controlling for weather, averaged over the sample. We may propose the new causal assumptions as Fig. 3. To justify this empirical estimand, we need to defend our causal assumptions and argue why we assume that weather confounds the relationship and why there are no further confounders.

Every empirical research that aims at causality needs causal assumptions about how observable data provide information about the target of causal inference because data only contain statistical associations and cannot reveal causal relationships without causal assumptions [8,10]. Researchers need to clearly specify their causal assumptions and plausibly defend them rather than keeping them hidden or intransparent. The assumptions may be defended with related theories and subject matter knowledge such as the temporal ordering of the variables (i.e., a cause precedes its consequence) and design characteristics of the data collection process (e.g., randomization and double blinding).

Often, once the theoretical and empirical estimand are specified, the causal assumptions will readily show that the observable evidence is not suitable for supporting the causal interpretations that we are interested in. Clarity and transparency regarding the estimands allow researchers to realize misalignments between their research goal and empirical evidence and to accordingly make improvements. Moreover, other researchers will be able to assess how plausible the causal assumptions are and whether the empirical evidence validly provides information regarding the research goal.

3.3. Estimation strategy

The third guideline is to empirically choose the strategy to estimate the empirical estimand from data. Following [4], the same empirical estimand can be estimated by diverse estimation strategies, such as parametric models, semi-parametric models, non-parametric models, and machine learning models. Researchers should design candidate estimation strategies that recover the empirical estimand from available data. Researchers need to consider how well the statistical assumptions of the respective estimation strategy hold for data (e.g., homoscedasticity and independence of observations). Admittedly, it is often difficult, if not impossible, to justify all aspects of the estimation strategy a priori. For example, the functional form (e.g., linear, quadratic, or stratified non-parametric) of the estimation strategy can be hard to defend based on theories alone, because theories rarely involve such information.

Instead of conceptually arguing among different estimation strategies that all serve as estimators for the same empirical estimand, they may be selected in a largely data-driven way. For this purpose, researchers should develop appropriate performance metrics for assessing the candidate estimation strategies. Per suggestions by Refs. [4,10], a useful metric may be the out-of-sample predictive performance tailored for the theoretical estimand and the empirical estimand. Again, clearly specified estimands will guide the research because they tell us what kind of out-of-sample predictions we aim to make. This is especially relevant when the sample is not representative, so that reweighting is needed when making out-of-sample predictions for the target population.

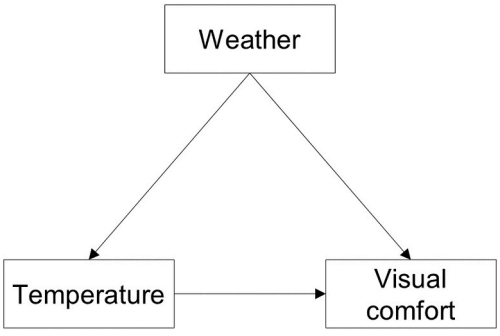


Fig. 3. Example causal assumption that (only) weather confounds the effect of temperature on visual comfort.

In our example, to estimate the effect of temperature on visual comfort while controlling for weather, we could consider candidate estimation strategies among machine learning algorithms and regression models that may differ by higher order terms and interaction terms. These candidate estimation strategies need to be designed under consideration of the theoretical estimand and the empirical estimand. Besides, their statistical assumptions should hold for the available data. To select the best estimation strategy, we may choose the one that minimizes expected squared errors in out-of-sample predictions for a representative sample.

4. Common pitfalls and recommendations

Below, we address a selection of common methodological pitfalls in cross-modal research that threaten causal inference. Based on the above-presented guidelines, we grouped these pitfalls into separate areas of concern. For each group, we discuss major problems and provide corresponding recommendations.

4.1. Theoretical estimand

The first group of pitfalls concerns the theoretical estimand. We will discuss underspecified theoretical estimands and questionable generalizability.

4.1.1. Underspecified theoretical estimands

In current cross-modal research, one major problem could be the insufficient specification of theoretical estimands, especially at the beginning of the research. This is the case when researchers first spent considerable resources collecting potentially messy data and only afterwards consider the target of causal inference and analysis methods. As Sir Fisher [28] said:

> "To consult the statistician after an experiment is finished is often merely to ask him to conduct a post mortem examination. He can perhaps say what the experiment died of."

Conducting experiments without thorough consideration of theoretical estimands (and the following empirical estimands and causal assumptions) may lead to serious causal inference problems that are unrecoverable by statistics [10,29]. For example, if causal assumptions imply important confounders, but these confounders are not registered in data, we may be unable to identify the target causal effect regardless of the estimation strategy. In such cases, the empirical evidence cannot be plausibly mapped to the theoretical estimand. Furthermore, data might be collected in a way that does not match the causal assumptions that researchers impose on the study. For instance, an unblinded experiment may cause participants' reactivity to confound the effect of interest as suggested by the Hawthorne effect [30]. Researchers might neglect the reactivity and incorrectly assume no confounders. In such situations, any subsequent estimation will be misleading as the empirical estimand cannot appropriately approximate the theoretical estimand because of flawed causal assumptions.

In our experience, the research targets of existing cross-modal studies are often obscure and cannot be translated into unambiguous unit-specific quantities and target populations as required by the above-presented first guideline. Insufficient specification of theoretical estimands may have contributed to the prevalent confusion between causal and predictive research highlighted above. Without a clear theoretical estimand, the research goal is obscure. It becomes difficult to determine whether the target of the inference are causal effects or merely predictions under stable settings.

Moreover, underspecified theoretical estimands make research intransparent and hard to interpret. Often, we could reconstruct the empirical estimand based on the analysis conducted. However, the same empirical estimand may be used to approximate different theoretical estimands, usually with different degrees of plausibility. For example, when there are interactions among covariates and independent variables, the effects of the independent variables will depend on the specific levels of the covariates. Thus, the empirical effects conditional on certain levels of the covariates may not plausibly represent the effects of interest given very different levels of covariates. However, we can only assess such discrepancies when the target levels of covariates are clearly specified. More generally, if the theoretical estimand is underspecified, it often becomes unclear whether there is a gap between the empirical evidence and the target of causal inference and how valid the conclusions are. In such cases, the results will be hard, if not impossible, to interpret and evaluate. It is thus important to recall the first guideline, suggesting that researchers should clearly specify the theoretical estimand, particularly at the start of their study.

4.1.2. *Questionable generalizability*

Given the well-recognized importance of generalizability (e.g., Refs. [31–33]), it deserves increased attention in cross-modal research. Thereby, three major problems need to be addressed. First, the target population is often not specified. Second, generalizability is often incorrectly assumed as given. Third, advanced methods for transporting results across populations have not been generally considered.

We first address the underspecification of the target population. Currently, research publications generally describe the sample characteristics. However, a clear statement of the target population is often missing. This is problematic because describing the sample cannot replace a specification of the target population. If unspecified, the target population may be the sample, the population where the sample comes from, or some other population [18]. Such lack of clarity can render research intransparent, lead to misapplication of results to unsuitable populations, and hinder the evaluation of the results' generalizability. As required by the first guideline on theoretical estimand, researchers should always specify their target population and explain why this population is of interest. Transparency regarding the target population facilitates the consideration of generalizability and sheds light on the underlying inferential problems (see below).

The second problem is related to the possibly incorrect generalizability assumption. Besides specifying the target population, researchers need to convincingly argue for the generalizability of the results from the available sample to the target population. At times, researchers may implicitly assume generalizability of their own or other authors' results as they interpret the results as applicable for the target population without providing arguments. Such interpretations can be implausible for many reasons. For example, many cross-modal studies have samples of a very limited size.[12] Also, they are often recruited from a special subpopulation (e.g., opportunity sample). Such samples are unlikely to be representative of a broader target population and do not justify direct generalization [34].

Researchers may incorrectly appeal to the diversity of the sample and argue for the applicability of the results for a broader population. However, diversity does not imply representativeness. As methodological literature (e.g., Refs. [17,18]) has pointed out, if the diversity in the sample does not match the diversity in the target population, the results from the sample may still be invalid for the target population. Similarly, generalizability cannot be simply assumed for a large sample because a large sample size does not guarantee representativeness [34].

Finally, we should mention that advanced methods for generalization are often not taken into account. Current research practices sometimes include disclaimers about how the sample may not be representative of the target population and how the results may not generalize. However, the focus on threats to generalizability only mentions problems instead of solving them [18]. Advances in causal inference have brought us the key insight that valid estimates for the target population may still be possible even if the sample is not representative. From the causal inference perspective, generalizability depends on the mechanisms by which the populations differ and the causal relationships among the variables [17,18]. Various methods have been developed for transporting results across populations, but they are often neglected in cross-modal research.

In this context, it would be useful to briefly introduce selection diagrams [17] and post-stratification [35,36]. Selection diagrams extend DAGs and represent the mechanisms underlying differences in populations [17]. We can use selection nodes to indicate variables that are assumed to distribute differently across populations. These additional causal assumptions allow us to determine whether generalization is possible and derive better empirical estimands for the target population.

Once the empirical estimand is specified based on selection diagrams, we may apply post-stratification in our estimation strategy to adjust for the differences between the populations [10,35,36]. Specifically, estimates from the subpopulations in the sample are reweighted based on their relative frequencies in the target population regarding relevant variables specified by the selection diagrams. Such procedures are expected to result in better estimates.

For example, we may be interested in the effect of lighting on thermal comfort and assume that only sex modifies this effect (Fig. 4). Our sample may mainly be female, but our target population are all German citizens which has a more balanced sex distribution. Apparently, the sample differs from the target population regarding the sex distribution. We thus add a selection node to the node sex. Based on the selection diagram, we derive an empirical estimate that includes sex. To get a valid estimate for the target population, we may apply post-stratification to adjust the results by reweighting the sex to its distribution in the general German population.[13]

Advanced generalization methods open more possibilities to cross-modal research. For example, studies are generally more interested in effects in the real world, but out of practical constraints, the data often come from the laboratories. Transportability, a general framework for generalizability [17], offers us a systematic way to deal with whether and how causal effects may be transported from experimental settings to the real world.

Cross-cultural differences and generalizability are popular topics in building science (e.g., Refs. [37,38]) and can promisingly extend

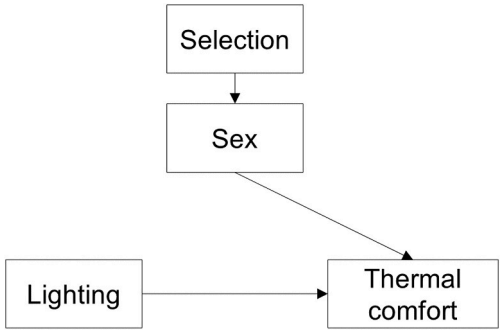


Fig. 4. Example selection diagram assuming that only sex modifies the effect of lighting on thermal comfort.

[12] For example, previous reviews (e.g., Ref. [2]) have criticized that many cross-modal studies had less than ten participants.

[13] Such statistics may be available from other sources like the German Federal Statistical Office or related studies.

cross-modal research. In cross-cultural studies, the differences or similarities between different populations are easily misinterpreted because of the diverse contexts involved and additional threats such as measurement inequivalence[14] and demographic differences. By adopting methods for cross-cultural generalizability (e.g., Ref. [18]), we can address these problems in a principled way and appropriately conduct cross-cultural cross-modal research.

In summary, the three recommendations offered can mitigate some of the current problems regarding generalization in cross-modal research. First, clearly specify the target population as required by the first guideline. Second, model the causal assumptions regarding how populations differ with tools such as selection diagrams. Third, use methods such as post-stratification to adjust sample estimates for the target population.

4.2. Empirical estimand

The next group of pitfalls concerns the empirical estimand and causal assumptions. We will discuss the mismatch between theoretical and empirical estimand and the neglect of measurement-related inferential problems.

4.2.1. Mismatch between theoretical and empirical estimand

Another pitfall that threatens cross-modal research are mismatches between the theoretical estimand and the empirical estimand, which are often implicitly implied by the research goal and the conducted analysis. Diverse problems may underlie such mismatches. Here, we focus on unspecified causal assumptions, inappropriate control variables, and a lack of theory for deriving causal assumptions.

Currently, hardly any cross-modal study explicitly specifies causal assumptions. However, as mentioned above, any causal results analyzed from empirical data always involve causal assumptions. In many studies, the causal assumptions implied by their analyses are implausible. For example, many studies apply procedures which do not control for any confounder (e.g., ANOVA, *t*-test, or Friedman test). Such analyses implicitly assume no confounders between the exposure and the outcome. For non-experimental studies, this assumption is generally unjustified because of a lack of experimental manipulation and randomization. However, even experiments in cross-modal research can suffer from imperfect manipulation and randomization, because experimental manipulations often have multiple environmental side effects. For instance (Fig. 5), changing ventilation rates to manipulate temperature may change CO_2, water vapor concentration, as well as further confounders. For demonstration purposes, we assume that these environmental side effects causally influence visual comfort. If not controlled for, they will confound and systematically bias the target causal effect. In such cases, the underlying causal assumption of no confounders is implausible and any empirical estimand based on this assumption would mismatch the theoretical estimand.

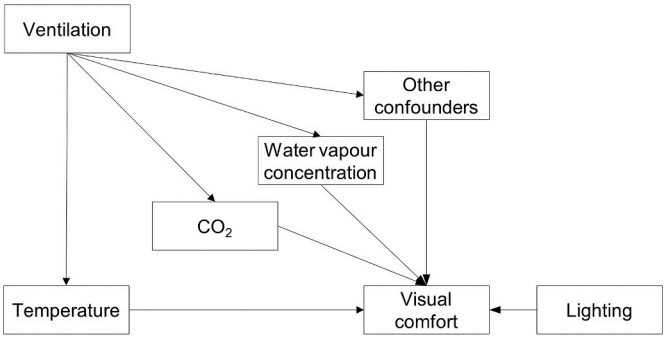


Fig. 5. Example causal assumptions on the effects of temperature and lighting on visual comfort with confounders[15] introduced by ventilation such as CO_2 and water vapor concentration.

Besides omitting important control variables, inappropriate inclusion of control variables may also cause a mismatching empirical estimand and lead to systematically biased results [29]. As introduced above, including a collider will lead to spurious relationships. Suppose temperature influences thermal comfort, and overall comfort is both influenced by temperature and thermal comfort (Fig. 6). If an analysis regarding the effect of temperature on thermal comfort includes overall comfort as a covariate, it introduces overall comfort as a collider and biases the estimation with unpredictable magnitude and direction.

In this example, researchers may also induce the overcontrol bias [22] by controlling for thermal comfort while investigating the effect of temperature on overall comfort. Here, thermal comfort is a mediator. Controlling for it will block the causal effect from temperature to overall comfort through thermal comfort, thus biasing the results. Beyond these examples, inappropriate control variables can also induce numerous other types of biases (see Ref. [22] for further examples).

It is at times assumed that a large sample could resolve various biases. For example, larger samples might be mentioned as a way to check the reliability and validity of the findings for future research. One might also incorrectly assume that if an effect is consistently found within large samples, the results can be considered as robust and trustworthy. However, sheer sample size does not address the problems in causal assumptions. Bad control variables systematically introduce biases to the results regardless of the sample size because the empirical estimand mismatch the theoretical estimand. Under certain conditions, a large sample can even amplify biases [22,34]. The consensus in causal inference is that wrong causal assumptions can be fixed neither by data nor by estimation strategy [10,25]. Biases need to be assessed with causal assumptions and corrected by improving empirical estimands (e.g., with an adjusted set of control variables).

To counteract these problems, one should explicitly state and defend the causal assumptions underlying the empirical estimand as per the second aforementioned guideline. Since defending causal assumptions requires extensive related theories and subject matter knowledge, it is admittedly very difficult to plausibly argue for causal assumptions, particularly given the current state of research where theories for deriving causal relationships and valid causal results are generally lacking. Thus, a pressing task for the cross-modal research community is the development of formal theories to accompany and guide our research. A theoretical framework for cross-modal research is currently being developed by the authors. Note that researchers do not need to, in fact generally cannot, propose perfect causal assumptions, since no research can take every influencing aspect into consideration. However, researchers should always keep their assumptions transparent to

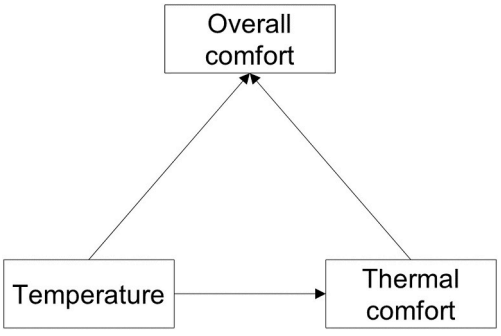


Fig. 6. Example causal assumptions regarding the relationships among temperature, thermal comfort, and overall comfort.

[14] I.e., measurement instruments may demonstrate different properties (e.g., reliability and validity) across populations (e.g., Refs. [37,70]).

facilitate independent evaluation and interpretation by other researchers.

4.2.2. Measurement-related inferential problems

Measurement-related inferential problems like measurement errors and validity are further pitfalls in cross-modal research and the focus of this section. It is widely recognized that all measurements involve errors. The errors may stem from, for example, instrument inaccuracy and data entry mistakes. But when researchers, for instance, aim at the effects between latent constructs (e.g., visual and thermal comfort) but directly conduct regressions on observed values (e.g., subjective ratings) without adjusting for reliability, their analysis is conducted as if there were no measurement errors.

However, measurement errors are not a trivial matter for causal inference. There is a vast literature showing how neglecting measurement errors can attenuate or exaggerate results (e.g., Refs. [39–41]). The strength and direction of the bias induced by measurement errors depend on the specific constellation [42,43]. For example, if we directly use observed values to adjust for confounding effects, we will not completely remove all confounding, as these values are an error-loaded proxy for the latent construct, and hence residual spurious relationship will remain in results. The spurious relationship might then bias the results upwards or downwards with unpredictable magnitude.

Following [42], we recommend representing measurement errors in DAGs while specifying causal assumptions. Researchers may then make assumptions regarding the structure and magnitude of measurement errors and conduct statistical error corrections (for introductions to assumptions and error models, see Refs. [42–44]). For instance, independent measurement errors assume that the errors for different constructs do not influence each other. Errors may also be assumed as non-differential when the latent value of the construct does not influence its errors.

In Fig. 7, we illustrate an example DAG where the observable comfort ratings are both determined by the latent comfort and measurement errors. We assume the errors for thermal comfort and visual comfort to be dependent as personality affects both ratings (e.g., through individual response tendencies). The errors are assumed as non-differential, because they are both independent of the latent comfort. Given these causal assumptions, we may empirically estimate or directly assume the measurement reliability and use that value to adjust the observable ratings for latent comfort. To address the dependence of errors, we may use repeated measures and multilevel modeling to control for

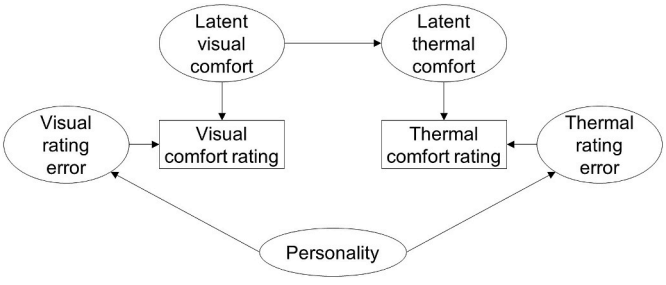


Fig. 7. Example causal assumptions for the effect of visual comfort on thermal comfort. The latent comfort and errors determine the observable comfort ratings. The errors are dependent because of the confounder personality.

[15] Although we emphasize the importance of being as precise as possible when making causal assumptions, we acknowledge that researchers may often encounter cases where not all confounders can be or have been observed. In causal inference, it is a common practice to indicate the presence of further unobserved confounders in DAGs (e.g., Ref. [10]).

personality.

Another major problem related to measurement is the questionable validity. When the scales used in cross-modal research (regarding, for example, comfort or satisfaction) are not validated, it is unclear what latent constructs are being measured and how successfully the operationalizations reflect the target constructs (i.e., construct validity [45]). This can lead to gaps between the empirical estimand and the theoretical estimand, for instance, if the target construct is comfort but the scale mainly measures social desirability (i.e., the tendency to response in a socially favorable way [46,47]).

Researchers from other fields have been intensively discussing problems related to validity. For example, Yarkoni [31] highlighted poor alignments between hypotheses and quantitative inference in psychology and argued that results and generalizations are invalidated by seemingly arbitrary operationalizations of broad constructs. Similarly, Eronen and Bringmann [48] also deem questionable construct validity as one of the fundamental difficulties in psychological research. For cross-modal research, we recommend the community consider and discuss the validity of the scales they apply. Because of a limited scope, we refrain from further discussions on this topic and refer interested readers to our upcoming review on the constructs and scales used in cross-modal research.

4.3. Estimation strategy

The final group of pitfalls concerns the estimation strategy. We will discuss violations of statistical assumptions, inadequate analysis of rating scale data and interaction-related inferential problems.

4.3.1. Violations of statistical assumptions

Inappropriate statistical practices, such as violations of statistical assumptions, have been repeatedly highlighted in many fields over decades (e.g., Refs. [49,50]). Statistical procedures are generally designed based on statistical assumptions. If the assumptions underlying the procedure are seriously violated, the results may be completely invalidated.

The risk of getting invalidated results is increased if researchers ignore statistical assumptions, incorrectly examine statistical assumptions, or make implausible statistical assumptions. For example, one assumption for the linear regression model is the normality of residuals. Neither outcomes nor predictors are required to distribute normally [49,51]. However, researchers may incorrectly test for the normality of variables without checking residuals. Even when variables deviate from normality, such tests do not tell us whether the normal residuals assumption is violated, and thus do not justify any consequent change to other procedures such as non-parametric tests.

Another example is the application of procedures that assume independent observations (e.g., Pearson's correlation, *t*-test, and multiple regression) on data from repeated measures (such data are termed pseudoreplications [52,53]). If within-subjects designs are applied, as in many cross-modal studies, but the analyses neglect the non-independence of data, the results can be invalidated because each repetition would be mistaken as an independent sample. For a data set with 50 participants, each with 10 measurement points, procedures that assume independent observations would mistreat the data as if there were 500 (=50 × 10) participants. The degree of freedom and the variance decomposition within and between subjects will be incorrect, resulting in misleading standard errors and confidence intervals. The significance level will also be biased. Importantly, the results become uninterpretable as the bias can differ in magnitude and direction depending on the specific constellation (see Ref. [53] for simulated examples).

More generally, the assumption of independence is violated when residuals are correlated [52,53]. This can happen both at the individual level (e.g., from repeated measures or response tendencies) and at the group level (e.g., from participants tested in groups). In cross-modal

research, participants are often grouped during experiments (e.g., being in the same room at the same time). The spatial and temporal dependence induced by groupings will cause individuals to be more similar to each other, so that their residuals become correlated.

More advanced statistical procedures such as multilevel analysis [54,55] may be applied to model non-independent data. Some researchers may argue that non-independence, along with many other methodological problems, may not bias the results much and can be ignored. However, researchers should use the best methods available. Ignoring well-developed solutions to widely acknowledged methodological problems is not a hallmark of rigorous scientific practices. Furthermore, it is necessary to empirically demonstrate that methodological problems such as correlated residuals do not matter for the specific case. For this purpose, we may employ different methods and compare their results. However, simply assuming no difference between the methods does not suffice.

Because of a limited scope, we refrain from discussing further assumption violations. Generally speaking, researchers should thoroughly consider all statistical assumptions underlying their analyses and explicitly state whether their assumptions are satisfied.[16] Such transparency brings about awareness with regard to assumption violations and allows others to assess how appropriate the statistical analyses are and whether the results are valid.

4.3.2. Inadequate analysis of rating scale data

Rating scales are ubiquitous in cross-modal research. However, they can often be improperly analyzed and interpreted as metric and unbounded. Such analyses can systematically lead to diverse biases that differ in magnitude and direction depending on the constellation (see Ref. [56] for examples). Two important properties of rating scale data underlie these biases, namely, the ordinal measurement scale and boundedness. Rating scales are ordinal because the response levels have a natural order, but the differences between the levels are not necessarily equal [56,57]. Thermal comfort researchers, for instance, have questioned the equidistant assumption of common thermal sensation scales and empirically demonstrated that most people do not perceive the distances between scale categories as equal [37,58,59]. More generally, assuming a scale with 2 = uncomfortable, 3 = comfortable, and 4 = very comfortable, the change from 2 to 3 likely does not equal to the change from 3 to 4. Although the response levels are assigned with consecutive integers, these numbers only indicate order. The numerical assignment may be arbitrarily changed as long as the order is kept.[17]

In contrast, metric data have constant distances between adjacent values [56]. Common metric statistics such as means, standard deviations and Pearson's correlation require equidistant data. Applying metric statistics to rating scales requires the implicit assumption of equidistance between all adjacent response levels. If this assumption is violated, the results will be misleading and uninterpretable. For the previous example, a mean value of 2.3 (between uncomfortable and comfortable) does not have a clear interpretation if we cannot assume equidistant levels.

Rating scales are also inherently bounded, namely, there are limits on both ends. The limited range will squeeze a wide range of extreme values on a hypothetical latent scale (e.g., latent comfort) into the few response levels on the ends of a rating scale [10,60]. Thus, ratings are often denser on the ends and demonstrate floor and ceiling effects. By contrast, common metric methods assume data distribution over an unlimited range. These methods inadequately analyze rating scale data because they ignore the boundedness and floor and ceiling effects, thereby biasing the results.

For example, a room with an ambient temperature of 30 °C may generally be rated as thermally very uncomfortable, the lowest level of a scale. However, a room in an even higher temperature (i.e., worse regarding latent comfort) will still be generally rated as very uncomfortable because of the boundedness of the scale. This floor effect will suppress the effect of temperature, especially when the analysis is conducted with a linear model on the metric scale.

Some researchers are aware of the ordinal scale and boundedness but still want to apply metric methods on rating scale data. They may replace rating scales with visual analogue scales or percentages. However, this presumed "workaround" does not solve the underlying problems because the data remain bounded and possibly ordinal.[18] Other researchers may resort to common non-parametric procedures which only require ordinal data, such as Wilcoxon test and Friedman test. However, non-parametric procedures have several general limitations. For example, it is often hard or impossible to extend these procedures with necessary control variables derived by the causal assumptions and with multilevel structures to address correlated residuals. In addition, common non-parametric procedures generally cannot analyze cross-modal interactions that are often of major research interest. Furthermore, non-parametric procedures have less power than parametric ones [61]. Moreover, such non-parametric results need to be interpreted on the ordinal scale and are thus less informative for intervention-related decisions than the metric results from parametric procedures.

For analyzing outcome variables from rating scales, we recommend ordinal regression (see Ref. [62] for introductions). Loosely speaking, ordinal models transform ordinal and bounded data into a latent metric unbounded scale and then conduct regression on that scale. Explanatory variables in ordinal regressions can be interpreted similarly as in linear regression. Besides, ordinal models are parametric and allow studying interactions. They can also incorporate control variables, multilevel structures, and further statistical extensions.

In addition to outcome variables, cross-modal research often involves ordinal and bounded explanatory variables. For example, ratings may be used as explanatory variables when studying the effect of thermal comfort on visual comfort. Furthermore, ordinal variables such as education level and socio-economic status may be included as control variables. As with ordinal and bounded outcome variables, mistreating these explanatory variables as metric and unbounded can also bias the results in an unpredictable way [63]. For appropriate modeling, we recommend monotonic regression (also known as isotonic regression [64]). For example, researchers may apply Bayesian models using Dirichlet distributions as priors for the ordinal explanatory variables (for further information, see Refs. [10,63]).

4.3.3. Interaction-related inferential problems

A central target of cross-modal research is cross-modal interaction, that is whether a stimulus from one domain influences the effect direction and magnitude of a stimulus from another domain [1]. In causal inference, a causal interaction means that an intervention on a variable would influence the effect of another variable on an outcome [65,66]. In this section, we focus on three inferential problems related to causal interactions, namely, confounded interactions, scale dependence, and interactions induced by floor and ceiling effects.

One major inferential problem is that confounders, which induce non-causal relationships between variables, may also induce spurious interactions [65,66]. In a spurious interaction, the effect of variable x varies with variable y, but intervention on y does not change the effect of x, because y is confounded with a third variable z (i.e., confounder) that causally interacts with x. In Fig. 5, suppose lighting (x) only causally

[16] We acknowledge the word limit in journals and recommend making such statements in footnotes or supplementary materials.

[17] For example, the example scale levels may be remapped to 1, 9.94 and 10.16.

[18] Visual analogue scales and percentages often need to be treated as ordinal rather than metric because in practice, a difference near the middle of the scale (e.g., between 49 % and 50 %) may qualitatively differ from the same numerical difference at the end [71] (e.g., between 98 % and 99 %).

interacts with humidity (z) regarding visual comfort. Because the manipulation changes humidity along with temperature (y), humidity confounds the interaction between temperature and lighting. Thus, even if temperature does not causally interact with lighting, we can still find a significant spurious interaction between them because of the confounding between temperature and humidity.

In cross-modal research, if studies do not consider the possibility of spurious interactions or inappropriately control for confounding interactions, such neglect may result in interaction estimates that are biased with unpredictable magnitude and direction [65,66]. To properly remove spurious interactions for causal inference, researchers need to control for the interaction between any relevant confounder and the independent variable of interest [66]. For the previous example, we may include the interaction term between humidity and lighting as a covariate. Importantly, simply controlling for the confounder, for example by including standalone humidity term instead of its interaction term as a covariate, does not remove spurious interactions [66].

Another generally overlooked problem inherent to interactions is the scale dependence [4,65,67], namely, the direction and strength of an interaction depend on the analysis scale (e.g., additive or multiplicative). For example (Fig. 8 left), if we analyze the effect of variable x (e.g., illuminance) on outcome y (e.g., visual comfort) over variable z (e.g., temperature) on an additive scale (e.g., using the absolute values), both effects of x appear parallel, indicating no interaction between x and z. However, if we analyze the interaction on a multiplicative scale, for example by log-transforming y to the rates of change, the effects of x over z become non-parallel (see right). This means that an interaction between x and z emerges purely through the scale transformation. Thus, a null result regarding an interaction using an estimation strategy on an additive scale may become significant on a multiplicative scale. Similarly, changing the analysis scale may also cause a significant interaction to differ in size and direction or even disappear.

The scale dependence of interaction is particularly problematic when the observed measurement and the underlying construct do not have a natural single mapping [65,68]. This can particularly be the case for rating scales ubiquitous in cross-modal research. As discussed previously, rating scale data should not be analyzed and interpreted as metric on the absolute scale because they are ordinal and bounded. Appropriately analyzing them for interactions generally requires transformation to a latent scale. Because ratings are often more densely distributed on both ends rather than evenly distributed over the whole range, a linear mapping of the ratings for the latent scale is not plausible and non-linear mappings are required. However, as rating scales differ greatly in their properties, there is no single natural mapping for them, so that assumptions for such non-linear mappings are necessary [67]. Because of the scale dependence, by assuming different non-linear mappings, the direction and magnitude of the interaction will change accordingly. Overall, we recommend researchers consider the analysis scale for the interaction in the estimation strategy, specify how and defend why they choose the scale of interest and interpret the meaning of the interaction in alignment with the chosen scale (e.g., avoid misinterpreting multiplicative results as additive).

When studying cross-modal interactions with rating scales, researchers also need to deal with above-mentioned floor and ceiling effects that are inherent to ratings because of the boundedness, as these effects can also induce spurious interactions [10,65]. In Fig. 9 left, there is no interaction as the two lines are parallel. But when there are floor and ceiling effects, there is less room for change near the ends and all data over the limit are assigned with the highest or lowest rating. In a linear model assuming an unlimited range (e.g., ANOVA or multiple linear regression), this boundedness induces methodological artifacts and results in a spurious interaction, shown by the lines on the right which are no longer parallel. This highlights again the fact that a significant interaction is not necessarily a causal interaction. When investigating causal interactions, the boundness requires a more appropriate statistical model than methods that assume unlimited distribution. As mentioned previously, for bounded data, ordinal regression would be preferable [63].

5. Conclusion

This paper addressed a fundamental problem underlying cross-modal research in building science, namely the widespread neglect of causal inference. To this end, we first discussed causality and differentiated causal research from predictive research. We then presented three guidelines that originated from Ref. [4] for appropriately conducting causal cross-modal research: first, specify the theoretical estimand as the target of causal inference; second, specify the empirical estimand that is informative for the theoretical estimand based on causal assumptions; third, select the estimation strategy empirically to estimate the empirical estimand. Finally, we discussed common methodological pitfalls and offered corresponding recommendations.

The presented guidelines and recommendations were meant to raise the awareness of causality in the relevant research community and to encourage reflections on our research practices. Moreover, they can assist researchers to plan, implement, justify, and evaluate causal studies in a principled way. We acknowledge that conducting causal inference appropriately is extremely difficult and there is no universally applicable standard procedure that guarantees valid causal inference. For example, empirical studies cannot take every conceivable aspect into consideration and will inevitably face alternative explanations that cannot be ruled out and assumptions that are unsubstantiated but necessary. Thus, we concede that the presented guidelines and recommendations will not solve all problems. Rather, the assumption is that adopting these guidelines will be an initial yet important step toward transparent causal cross-modal research. Transparency is arguably a

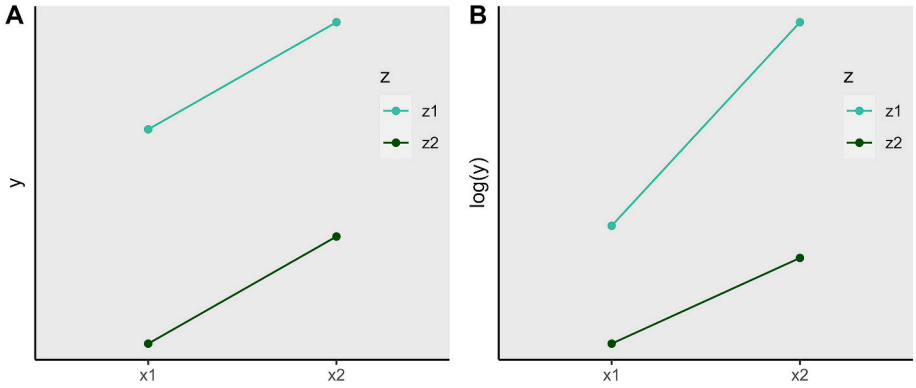


Fig. 8. Left: Two parallel lines indicating no interaction between variable x and z on an additive scale of outcome y. Right: Two non-parallel lines indicating an interaction between variable x and z on a multiplicative scale of outcome y. Figure inspired by Ref. [67].

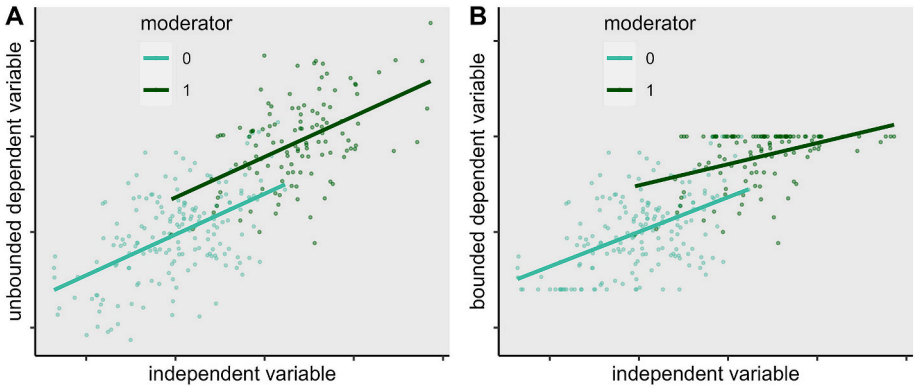


Fig. 9. Left: Two parallel lines indicating no interaction between the independent variable and the moderator when the dependent variable is not bounded. Right: Two non-parallel lines indicating an interaction between the independent variable and the moderator when the dependent variable is bounded. Figure inspired by Ref. [65].

necessary condition for research to generate independently verifiable and criticizable results and promote cumulative cross-modal research, which is expected to gradually resolve the inconsistency issues in the current literature.

Note that highlighting suboptimal research practices in the field can also reveal remarkable opportunities for future research. For example, we urgently need formal theories to guide our research so that plausible causal assumptions can be specified in cross-modal research. Needless to say, great efforts are also needed from the entire research community involved to conduct rigorous causal inference and establish valid causal results with regard to every aspect of human exposure to indoor-environmental conditions.

Although we mainly addressed cross-modal research, the presented guidelines and recommendations apply more broadly to other research fields. Specifically, research with regard to occupancy, occupant behavior and single-domain perception in the built environment can benefit from these recommendations, given the prevalence of challenges similar to those in cross-modal research, such as inattention to causality and deficient causal inference practices. As such, improved causal inference is likely to enhance not only cross-modal research, but also general research efforts in occupant-centric building design and operation.

CRediT authorship contribution statement

Jian Pan: Writing – review & editing, Writing – original draft, Visualization, Project administration, Methodology, Conceptualization. **Ardeshir Mahdavi:** Writing – review & editing, Funding acquisition, Conceptualization. **Isabel Mino-Rodriguez:** Writing – review & editing, Visualization, Conceptualization. **Irene Martínez-Muñoz:** Writing – review & editing, Conceptualization. **Christiane Berger:** Writing – review & editing, Conceptualization. **Marcel Schweiker:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgement

Jian Pan, Marcel Schweiker, and Isabel Mino-Rodriguez are funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under the project number 498385769.

Marcel Schweiker is supported by a research grant (21055) by VILLUM FONDEN.

Ardeshir Mahdavi and Irene Martínez-Muñoz are supported by the FWF (Austrian Science Fund: "Der Wissenschaftsfonds") Project MuDoCo (Project I 5993).

The funding sources provided only financial support for this work. They had no further involvement.

References

[1] G. Chinazzo, R.K. Andersen, E. Azar, V.M. Barthelmes, C. Becchio, L. Belussi, C. Berger, S. Carlucci, S.P. Corgnati, S. Crosby, L. Danza, L. de Castro, M. Favero, S. Gauthier, R.T. Hellwig, Q. Jin, J. Kim, M. Sarey Khanie, D. Khovalyg, C. Lingua, A. Luna-Navarro, A. Mahdavi, C. Miller, I. Mino-Rodriguez, I. Pigliautile, A. L. Pisello, R.F. Rupp, A.-M. Sadick, F. Salamone, M. Schweiker, M. Syndicus, G. Spigliantini, N.G. Vasquez, D. Vakalis, M. Vellei, S. Wei, Quality criteria for multi-domain studies in the indoor environment: critical review towards research guidelines and recommendations, Build. Environ. 226 (2022), 109719, https://doi.org/10.1016/j.buildenv.2022.109719.

[2] M. Schweiker, E. Ampatzi, M.S. Andargie, R.K. Andersen, E. Azar, V.M. Barthelmes, C. Berger, L. Bourikas, S. Carlucci, G. Chinazzo, L.P. Edappilly, M. Favero, S. Gauthier, A. Jamrozik, M. Kane, A. Mahdavi, C. Piselli, A.L. Pisello, A. Roetzel, A. Rysanek, K. Sharma, S. Zhang, Review of multi-domain approaches to indoor environmental perception and behaviour, Build. Environ. 176 (2020), 106804, https://doi.org/10.1016/j.buildenv.2020.106804.

[3] A. Mahdavi, C. Berger, Critical appraisal of recent research in multi-domain indoor-environmental exposure, in: M. Schweiker, C. van Treeck, D. Müller, J. Fels, T. Kraus, H. Pallubinsky (Eds.), Proceedings of Healthy Buildings 2023 Europe, Aachen, Germany, 2023.

[4] I. Lundberg, R. Johnson, B.M. Stewart, What is your estimand? Defining the target quantity connects statistical evidence to theory, Am. Socio. Rev. 86 (2021) 532–565, https://doi.org/10.1177/00031224211004187.

[5] H. Moshontz, L. Campbell, C.R. Ebersole, H. Ijzerman, H.L. Urry, P.S. Forscher, J. E. Grahe, R.J. McCarthy, E.D. Musser, J. Antfolk, C.M. Castille, T.R. Evans, S. Fiedler, J.K. Flake, D.A. Forero, S.M.J. Janssen, J.R. Keene, J. Protzko, B. Aczel, S.Á. Solas, D. Ansari, D. Awlia, E. Baskin, C. Batres, M.L. Borras-Guevara, C. Brick, P. Chandel, A. Chatard, W.J. Chopik, D. Clarance, N.A. Coles, K.S. Corker, B.J. W. Dixson, V. Dranseika, Y. Dunham, N.W. Fox, G. Gardiner, S.M. Garrison, T. Gill, A.C. Hahn, B. Jaeger, P. Kačmár, G. Kaminski, P. Kanske, Z. Kekecs, M. Kline, M. A. Koehn, P. Kujur, C.A. Levitan, J.K. Miller, C. Okan, J. Olsen, O. Oviedo-Trespalacios, A.A. Özdoğru, B. Pande, A. Parganiha, N. Parveen, G. Pfuhl, S. Pradhan, I. Ropovik, N.O. Rule, B. Saunders, V. Schei, K. Schmidt, M.M. Singh, M. Sirota, C.N. Steltenpohl, S. Stieger, D. Storage, G.B. Sullivan, A. Szabelska, C. K. Tamnes, M.A. Vadillo, J.V. Valentova, W. Vanpaemel, M.A.C. Varella, E. Vergauwe, M. Verschoor, M. Vianello, M. Voracek, G.P. Williams, J.P. Wilson, J. H. Zickfeld, J.D. Arnal, B. Aydin, S.C. Chen, L.M. Debruine, A.M. Fernandez, K. T. Horstmann, P.M. Isager, B. Jones, A. Kapucu, H. Lin, M.C. Mensink, G. Navarrete, M.A. Silan, C.R. Chartier, The psychological science accelerator: advancing psychology through a distributed collaborative network, Adv Methods Pract Psychol Sci 1 (2018) 501–515, https://doi.org/10.1177/2515245918797607.

[6] A. Reutlinger, J. Saatsi (Eds.), Explanation beyond Causation, Oxford University Press, 2018, https://doi.org/10.1093/oso/9780198777946.001.0001.

[7] M. Parascandola, The epidemiologic transition and changing concepts of causation and causal inference, Rev Hist Sci Paris 64 (2011) 243–262, https://doi.org/10.3917/rhs.642.0243.

[8] J. Pearl, D. Mackenzie, The Book of Why, Basic Books, New York, 2018.
[9] J. Pearl, Causal inference in statistics: an overview, Stat. Surv. 3 (2009) 96–146, https://doi.org/10.1214/09-SS057.
[10] R. McElreath, Statistical Rethinking, Chapman and Hall/CRC, 2020, https://doi.org/10.1201/9780429029608.
[11] A.R. Nogueira, A. Pugnana, S. Ruggieri, D. Pedreschi, J. Gama, Methods and tools for causal discovery and causal inference, Wiley Interdiscip Rev Data Min Knowl Discov 12 (2022) e1449, https://doi.org/10.1002/widm.1449.
[12] R. Yao, S. Zhang, C. Du, M. Schweiker, S. Hodder, B.W. Olesen, J. Toftum, F. Romana d'Ambrosio, H. Gebhardt, S. Zhou, F. Yuan, B. Li, Evolution and performance analysis of adaptive thermal comfort models – a comprehensive literature review, Build. Environ. 217 (2022), 109020, https://doi.org/10.1016/j.buildenv.2022.109020.
[13] S. Schlittmeier, A. Feil, A. Liebl, J. Hellbrück, The impact of road traffic noise on cognitive performance in attention-based tasks depends on noise level even within moderate-level ranges, Noise Health 17 (2015) 148–157, https://doi.org/10.4103/1463-1741.155845.
[14] J. Pearl, M. Glymour, N.P. Jewell, Causal Inference in Statistics: A Primer, John Wiley & Sons, 2016.
[15] J.M. Hofman, D.J. Watts, S. Athey, F. Garip, T.L. Griffiths, J. Kleinberg, H. Margetts, S. Mullainathan, M.J. Salganik, S. Vazire, A. Vespignani, T. Yarkoni, Integrating explanation and prediction in computational social science, Nature 595 (2021) 181–188, https://doi.org/10.1038/s41586-021-03659-0.
[16] C.L. Ramspek, E.W. Steyerberg, R.D. Riley, F.R. Rosendaal, O.M. Dekkers, F. W. Dekker, M. van Diepen, Prediction or causality? A scoping review of their conflation within current observational research, Eur. J. Epidemiol. 36 (2021) 889–898, https://doi.org/10.1007/s10654-021-00794-w.
[17] J. Pearl, E. Bareinboim, External validity: from do-calculus to transportability across populations, Stat. Sci. 29 (2014) 579–595, https://doi.org/10.1214/14-STS486.
[18] D. Deffner, J.M. Rohrer, R. McElreath, A causal framework for cross-cultural generalizability, Adv Methods Pract Psychol Sci 5 (2022) 1–18, https://doi.org/10.1177/25152459221106366.
[19] G. Schwarz, Estimating the dimension of a model, Ann. Stat. 6 (1978) 461–464, https://doi.org/10.1214/aos/1176344136.
[20] S. Arif, M.A. MacNeil, Predictive models aren't for causal inference, Ecol. Lett. 25 (2022) 1741–1745, https://doi.org/10.1111/ele.14033.
[21] G. Shmueli, To explain or to predict? Stat. Sci. 25 (2010) 289–310, https://doi.org/10.1214/10-STS330.
[22] C. Cinelli, A. Forney, J. Pearl, A crash course in good and bad controls, Socio. Methods Res. (2022) 1–34, https://doi.org/10.1177/00491241221099552.
[23] G. Chinazzo, J. Wienold, M. Andersen, Combined effects of daylight transmitted through coloured glazing and indoor temperature on thermal responses and overall comfort, Build. Environ. 144 (2018) 583–597, https://doi.org/10.1016/j.buildenv.2018.08.045.
[24] M. Ziat, C.A. Balcer, A. Shirtz, T. Rolison, A century later, the hue-heat hypothesis: does color truly affect temperature perception? in: F. Bello, H. Kajimoto, Y. Visell (Eds.), Haptics: Perception, Devices, Control, and Applications Springer International Publishing, Cham, 2016, pp. 273–280.
[25] J.M. Rohrer, Thinking clearly about correlations and causation: graphical causal models for observational data, Adv Methods Pract Psychol Sci 1 (2018) 27–42, https://doi.org/10.1177/2515245917745629.
[26] F. Elwert, Graphical causal models, in: S.L. Morgan (Ed.), Handbook of Causal Analysis for Social Research, Springer Netherlands, Dordrecht, 2013, pp. 245–273, https://doi.org/10.1007/978-94-007-6094-3_13.
[27] J. Textor, B. van der Zander, M.S. Gilthorpe, M. Liśkiewicz, G.T.H. Ellison, Robust causal inference using directed acyclic graphs: the R package 'dagitty, Int. J. Epidemiol. (2017) 1887, https://doi.org/10.1093/ije/dyw341. –1894.
[28] R.A. Fisher, Presidential address, Sankhyā – Indian J. Stat. (1933-1960) 4 (1938) 14–17. http://www.jstor.org/stable/40383882.
[29] A.C. Wysocki, K.M. Lawson, M. Rhemtulla, Statistical control requires causal justification, Adv Methods Pract Psychol Sci 5 (2022), https://doi.org/10.1177/25152459221095823.
[30] J.G. Adair, The Hawthorne effect: a reconsideration of the methodological artifact, J. Appl. Psychol. 69 (1984) 334–345, https://doi.org/10.1037/0021-9010.69.2.334.
[31] T. Yarkoni, The generalizability crisis, Behav. Brain Sci. 45 (2022) e1, https://doi.org/10.1017/S0140525X20001685.
[32] D.J. Simons, Y. Shoda, D.S. Lindsay, Constraints on generality (cog): a proposed addition to all empirical papers, Perspect. Psychol. Sci. 12 (2017) 1123–1128, https://doi.org/10.1177/1745691617708630.
[33] E. Mamulova, M. Loomans, R. Loonen, M. Schweiker, H. Kort, Let's talk scalability: the current status of multi-domain thermal comfort models as support tools for the design of office buildings, Build. Environ. 242 (2023), 110502, https://doi.org/10.1016/j.buildenv.2023.110502.
[34] R.M. Kaplan, D.A. Chambers, R.E. Glasgow, Big data and large sample size: a cautionary note on the potential for bias, Clin Transl Sci 7 (2014) 342–346, https://doi.org/10.1111/cts.12178.
[35] D.K. Park, A. Gelman, J. Bafumi, Bayesian multilevel estimation with poststratification: state-level estimates from national polls, Polit. Anal. 12 (2004) 375–385, https://doi.org/10.1093/pan/mph024.
[36] Y. Gao, L. Kennedy, D. Simpson, A. Gelman, Improving multilevel regression and poststratification with structured priors, Bayesian Anal 16 (2021) 719–744, https://doi.org/10.1214/20-BA1223.
[37] M. Schweiker, M. André, F. Al-Atrash, H. Al-Khatri, R.R. Alprianti, H. Alsaad, R. Amin, E. Ampatzi, A.Y. Arsano, E. Azar, B. Bannazadeh, A. Batagarawa, S. Becker, C. Buonocore, B. Cao, J.-H. Choi, C. Chun, H. Daanen, S.A. Damiati, L. Daniel, R. De Vecchi, S. Dhaka, S. Domínguez-Amarillo, E. Dudkiewicz, L. P. Edappilly, J. Fernández-Agüera, M. Folkerts, A. Frijns, G. Gaona, V. Garg, S. Gauthier, S.G. Jabbari, D. Harimi, R.T. Hellwig, G.M. Huebner, Q. Jin, M. Jowkar, J. Kim, N. King, B. Kingma, M.D. Koerniawan, J. Kolarik, S. Kumar, A. Kwok, R. Lamberts, M. Laska, M.C.J. Lee, Y. Lee, V. Lindermayr, M. Mahaki, U. Marcel-Okafor, L. Marín-Restrepo, A. Marquardsen, F. Martellotta, J. Mathur, I. Mino-Rodriguez, A. Montazami, D. Mou, B. Moujalled, M. Nakajima, E. Ng, M. Okafor, M. Olweny, W. Ouyang, A.L. Papst de Abreu, A. Pérez-Fargallo, I. Rajapaksha, G. Ramos, S. Rashid, C.F. Reinhart, MaI. Rivera, M. Salmanzadeh, K. Schakib-Ekbatan, S. Schiavon, S. Shooshtarian, M. Shukuya, V. Soebarto, S. Suhendri, M. Tahsildoost, F. Tartarini, D. Teli, P. Tewari, S. Thapa, M. Trebilcock, J. Trojan, R.B. Tukur, C. Voelker, Y. Yam, L. Yang, G. Zapata-Lancaster, Y. Zhai, Y. Zhu, Z. Zomorodian, Evaluating assumptions of scales for subjective assessment of thermal environments – do laypersons perceive them the way, we researchers believe? Energy Build. 211 (2020), 109761 https://doi.org/10.1016/j.enbuild.2020.109761.
[38] D.H. Kim, K.P. Mansfield, A cross-cultural study on perceived lighting quality and occupants' well-being between UK and South Korea, Energy Build. 119 (2016) 211–217, https://doi.org/10.1016/j.enbuild.2016.03.033.
[39] J. Westfall, T. Yarkoni, Statistically controlling for confounding constructs is harder than you think, PLoS One 11 (2016), e0152719, https://doi.org/10.1371/journal.pone.0152719.
[40] E. Loken, A. Gelman, Measurement error and the replication crisis, Science 355 (2017) 584–585, https://doi.org/10.1126/science.aal3618.
[41] T.B. Brakenhoff, M. Mitroiu, R.H. Keogh, K.G.M. Moons, R.H.H. Groenwold, M. van Smeden, Measurement error is often neglected in medical literature: a systematic review, J. Clin. Epidemiol. 98 (2018) 89–97, https://doi.org/10.1016/j.jclinepi.2018.02.023.
[42] M.A. Hernán, S.R. Cole, Invited commentary: causal diagrams and measurement bias, Am. J. Epidemiol. 170 (2009) 959–962, https://doi.org/10.1093/aje/kwp293.
[43] M. Van Smeden, T.L. Lash, R.H.H. Groenwold, Reflection on modern methods: five myths about measurement error in epidemiological research, Int. J. Epidemiol. 49 (2020) 338–347, https://doi.org/10.1093/ije/dyz251.
[44] E. Saccenti, M.H.W.B. Hendriks, A.K. Smilde, Corruption of the Pearson correlation coefficient by measurement error and its estimation, bias, and correction under different error models, Sci. Rep. 10 (2020) 438, https://doi.org/10.1038/s41598-019-57247-4.
[45] L.J. Cronbach, P.E. Meehl, Construct validity in psychological tests, Psychol. Bull. 52 (1955) 281–302, https://doi.org/10.1037/h0040957.
[46] S. Vesely, C.A. Klöckner, Social desirability in environmental psychology research: three meta-analyses, Front. Psychol. 11 (2020) 1395, https://doi.org/10.3389/fpsyg.2020.01395.
[47] D.P. Crowne, D. Marlowe, A new scale of social desirability independent of psychopathology, J. Consult. Psychol. 24 (1960) 349–354, https://doi.org/10.1037/h0047358.
[48] M.I. Eronen, L.F. Bringmann, The theory crisis in psychology: how to move forward, Perspect. Psychol. Sci. 16 (2021) 779–788, https://doi.org/10.1177/1745691620970586.
[49] J. Uttley, Power analysis, sample size, and assessment of statistical assumptions—improving the evidential value of lighting research, Leukos 15 (2019) 143–162, https://doi.org/10.1080/15502724.2018.1533851.
[50] M.S. Thiese, Z.C. Arnold, S.D. Walker, The misuse and abuse of statistics in biomedical research, Biochem. Med. 25 (2015) 5–11, https://doi.org/10.11613/BM.2015.001.
[51] T.Z. Keith, Multiple Regression and beyond, third ed., Routledge, New York, 2019 https://doi.org/10.4324/9781315162348.
[52] M. Milinski, How to Avoid Seven Deadly Sins in the Study of Behavior, Adv Study Behav, 1997, pp. 159–180, https://doi.org/10.1016/S0065-3454(08)60379-4.
[53] W. Forstmeier, E.J. Wagenmakers, T.H. Parker, Detecting and avoiding likely false-positive findings – a practical guide, Biol. Rev. 92 (2017) 1941–1968, https://doi.org/10.1111/brv.12315.
[54] A. Gelman, J. Hill, Data Analysis Using Regression and Multilevel/Hierarchical Models, Cambridge University Press, 2006, https://doi.org/10.1017/CBO9780511790942.
[55] A.V. Diez Roux, A glossary for multilevel analysis, J. Epidemiol. Community Health 56 (2002) 588–594, https://doi.org/10.1136/jech.56.8.588.
[56] T.M. Liddell, J.K. Kruschke, Analyzing ordinal data with metric models: what could possibly go wrong? J. Exp. Soc. Psychol. 79 (2018) 328–348, https://doi.org/10.1016/j.jesp.2018.08.009.
[57] S.S. Stevens, On the theory of scales of measurement, Science 103 (1946) 677–680, https://doi.org/10.1126/science.103.2684.677.
[58] X. Fuchs, S. Becker, K. Schakib-Ekbatan, M. Schweiker, Subgroups holding different conceptions of scales rate room temperatures differently, Build. Environ. 128 (2018) 236–247, https://doi.org/10.1016/j.buildenv.2017.11.034.
[59] M. Schweiker, X. Fuchs, S. Becker, M. Shukuya, M. Dovjak, M. Hawighorst, J. Kolarik, Challenging the assumptions for thermal sensation scales, Build. Res. Inf. 45 (2017) 572–589, https://doi.org/10.1080/09613218.2016.1183185.
[60] M. Smithson, Y. Shou, Generalized Linear Models for Bounded and Limited Quantitative Variables, SAGE Publications, 2020, https://doi.org/10.4135/9781544318523.
[61] P. Macdonald, Power, type I, and type III error rates of parametric and nonparametric statistical tests, J. Exp. Educ. 67 (1999) 367–379, https://doi.org/10.1080/00220979909598489.

[62] P.C. Bürkner, M. Vuorre, Ordinal regression models in psychology: a tutorial, Adv Methods Pract Psychol Sci 2 (2019) 77–101, https://doi.org/10.1177/2515245918823199.
[63] P. Bürkner, E. Charpentier, Modelling monotonic effects of ordinal predictors in Bayesian regression models, Br. J. Math. Stat. Psychol. 73 (2020) 420–451, https://doi.org/10.1111/bmsp.12195.
[64] R.E. Barlow, H.D. Brunk, The isotonic regression problem and its dual, J. Am. Stat. Assoc. 67 (1972) 140–147, https://doi.org/10.1080/01621459.1972.10481216.
[65] J.M. Rohrer, R.C. Arslan, Precise answers to vague questions: issues with interactions, Adv Methods Pract Psychol Sci 4 (2021), https://doi.org/10.1177/25152459211007368.
[66] J.M. Rohrer, P. Hünermund, R.C. Arslan, M. Elson, That's a lot to process! Pitfalls of popular path models, Adv Methods Pract Psychol Sci 5 (2022), https://doi.org/10.1177/25152459221095827.
[67] R. Spake, D.E. Bowler, C.T. Callaghan, S.A. Blowes, C.P. Doncaster, L.H. Antão, S. Nakagawa, R. McElreath, J.M. Chase, Understanding 'it depends' in ecology: a guide to hypothesising, visualising and interpreting statistical interactions, Biol. Rev. (2023), https://doi.org/10.1111/brv.12939.
[68] E.J. Wagenmakers, A.M. Krypotos, A.H. Criss, G. Iverson, On the interpretation of removable interactions: a survey of the field 33 years after Loftus, Mem. Cognit. 40 (2012) 145–160, https://doi.org/10.3758/s13421-011-0158-0.
[69] C.H. Achen, Let's put garbage-can regressions and garbage-can probits where they belong, Conflict Manag. Peace Sci. 22 (2005) 327–339, https://doi.org/10.1080/07388940500339167.
[70] F.J.R. van de Vijver, K. Leung, Methods and Data Analysis for Cross-Cultural Research, Cambridge University Press, 2021, https://doi.org/10.1017/9781107415188.
[71] G.Z. Heller, M. Manuguerra, R. Chow, How to analyze the visual analogue scale: myths, truths and clinical relevance, Scand J Pain 13 (2016) 67–75, https://doi.org/10.1016/j.sjpain.2016.06.012.