Energy & Buildings 347 (2025) 116335

Contents lists available at ScienceDirect

Energy & Buildings

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

# Are we adequately accounting for human factors in analyses of thermal comfort studies?

A.K. Mishra [a,b,*], P Wargocki [b], E.J. O'Reilly [a]

[a] School of Public Health, University College Cork, Ireland
[b] DTU Sustain, Technical University of Denmark, Copenhagen, Denmark

ARTICLE INFO

*Keywords:*
Thermal comfort
Thermal sensation votes
Ordinal data
Survey response
Occupant centric

ABSTRACT

We discuss thermal comfort research data analysis, focusing on the nature of human responses, response scales, and human-built environment interactions. The goal is to highlight methods that provide an appreciation of human responses and inter-individual variability. We posit that there should be reduced focus on finding correlations and looking for "one size fits all" answers. Our position is illustrated by sample data from the ASHRAE Global Thermal Comfort Database II. Through this work, we want to promote dialogue on how data analysis can retain a strong mathematical basis while remaining true to the human responses regarding thermal comfort.

## 1. Introduction

Over the past two decades, ease of access to computational resources has allowed the use of sophisticated modelling, machine learning and AI techniques in analyses of thermal comfort research. However, two problems can undermine these advancements: (i) Applying the mathematics without consideration of the research question, the study design, and the data gathered [1–3]; and (ii) Overlooking that responses are a reflection of human experience [4,5]. We highlight here connections between them.

The nature of the questionnaire and the responses necessitate an appropriate analysis of ordered, categorical survey responses. At the same time, analysis of thermal comfort data should not be looking for a "one comfort zone fits all" kind of solution. Appropriate modelling can help us understand occupant needs and design future studies better.

Our goal is to initiate discussions regarding occupant-needs and how we may model their response considering math, physics, and the human factor. The intention is to build a dialogue on proceeding further without decrying currently used approaches based on the predicted mean vote-predicted percentage dissatisfied (PMV-PPD) or using adaptive thermal comfort (ATC) model.

## 2. The problem

Thermal comfort field studies mostly follow a cross-sectional study design. Right-now surveys using the seven-point (Cold; Cool; Slightly cool; Neutral; Slightly warm; Warm; Hot) ASHRAE thermal sensation scale are widely used. The archetype for the thermal sensation scale used today traces its origin back to the scales put forth by Houghton and Yaglou during 1920s [6] and Bedford [7]. As used by these pioneers, it was a categorical 7-point scale whose descriptors were changed by Gagge [6] to a form that reflects the current ASHRAE seven-point thermal sensation scale. The scale in this categorical form was used by Fanger [8]. This scale was later adapted by connecting the points with a line [9]. This form is often used today for analysis. However, adding the line did not change the ordered, categorical nature of the scale, even though it may be used as interval data and parametric statistics employed for analysis. While the seven-point scale may be presented as a continuum or with discrete points, it cannot be assumed that the scale categories are perceived as equidistant by respondents [4,5]. Therefore, the scale is ordinal – the magnitude of the difference in temperature and conditions that predict a two-point step change from cold to slightly cool should not be assumed the same in magnitude as the two point change from slightly warm to hot. When a survey-taker (respondent) marks the middle point between warm and slightly warm, it is only an assumption made by the research group that this response represents thermal sensation as being exactly in the middle of neutral and hot or even that the current indoor conditions are equidistant between the conditions that would elicit the respondent to choose warm and slightly warm.

The nature of the seven-point scale must be carefully considered during analysis (Fig. 1). Treating thermal sensation votes (TSV) as a continuous variable, using a linear model, assumes that responses are

* Corresponding author at: School of Public Health, University College Cork, Ireland.
*E-mail address:* AKumarMishra@ucc.ie (A.K. Mishra).

https://doi.org/10.1016/j.enbuild.2025.116335
Received 28 May 2025; Received in revised form 21 July 2025; Accepted 19 August 2025
Available online 20 August 2025

equidistant rather than ordered categories [2]. Additionally, it will promote a deterministic view of occupants, ignoring the variability between individuals who sense the thermal environment [10] and perceive thermal "comfort" differently [4,5].

Despite these limitations, linear regression models have been predominantly used to analyse TSV responses and reported in the thermal comfort research literature. However, the ordinal nature of the TSV data calls for other statistical approaches.

Recognizing these aspects inherent to the nature of scales [11], recent works have also advocated for ordinal regression approaches that respect the categorical nature of the TSV data [2]. These methods, particularly an ordered multinomial logit model (henceforth, called an ordinal model, OM), offer a valid statistical method for analysing TSV responses.

This methodological shift from linear to ordinal models is more than a statistical choice. It represents a fundamental change in how we conceptualize TSV responses to better illustrate the true nature of the occupant experience. An OM works out the cumulative probabilities of being in or below each category and do not assume that the response categories are uniformly distributed or equidistant. By moving from a deterministic neutral temperature and symmetric comfort zone (when TSV are treated as continuous variable) to a probabilistic view of occupant responses (when TSV are treated as categorical variable), the occupant preferences and needs can be more accurately reflected and modelled.

Let us consider next the question posed to respondents. The thermal sensation question is typically worded as: "How do you feel right now?" The right-now thermal sensation can be impacted by the thermal conditions experienced at the time the question is answered and over the preceding few minutes. The right-now thermal sensation can impact the hereafter thermal environment owing to occupant interactions with the environment, but only in spaces where such opportunities are available, e.g., opening a window or changing the set-point of a window-mounted air-conditioner. However, it cannot impact the right now or past thermal environment in other spaces, even in spaces where occupant-initiated adaptive actions are available. When 'Y' is caused by 'X,' 'Y' must follow 'X' [12,13] temporally. Temporality is required for causal outcomes.

Typically, we consider the thermal environment as the exposure parameter and thermal sensation as the outcome. To understand how occupant behaviour impacts the thermal environment, we could reframe this relation by considering the thermal sensation as the exposure parameter and the thermal environment behaviours (occupant adaptive behaviours to modify their thermal environment) as the outcome, predicted by the thermal environment. A time series record of TSV, and standard effective temperature (SET) would be suitable for understanding this relationship using vector autoregression (VAR) or similar models. However, time series data have not been traditionally recorded and are not available in field study databases such as the ASHRAE Global Thermal Comfort Database II (ASHRAE-II) [14].

The available data thus limit the models we can develop and the questions we can pose [2]. Given the lack of repeated measures data, flipping the relation between TSV and SET [3] does not provide a solution for causal inference because of the language of the question and temporality in the surveys. Still, it must be emphasized that understanding the limitation of the collected data can inspire better study designs in the future.

Finally, as for all surveys, the researcher should ensure that the respondent interprets the questions in the manner intended [5]. Specifically, does an occupant provide their sensation after an action they may have been taken in response to the building conditions for example, put on a sweater? Or, is the occupant responding to the sensation linked to the building conditions that prompted them to take an action to restore comfort?

## 3. The case

To illustrate our proposed approach to analysis of the thermal sensation votes as a categorical variable, we present the results of an OM for the TSV-SET relations in the data of Building ID 735 from ASHRAE-II. To create an OM for TSV based on SET, TSV values were binned together using the following approach: TSV $\geq -3$ and $<$ -2.5 were treated as

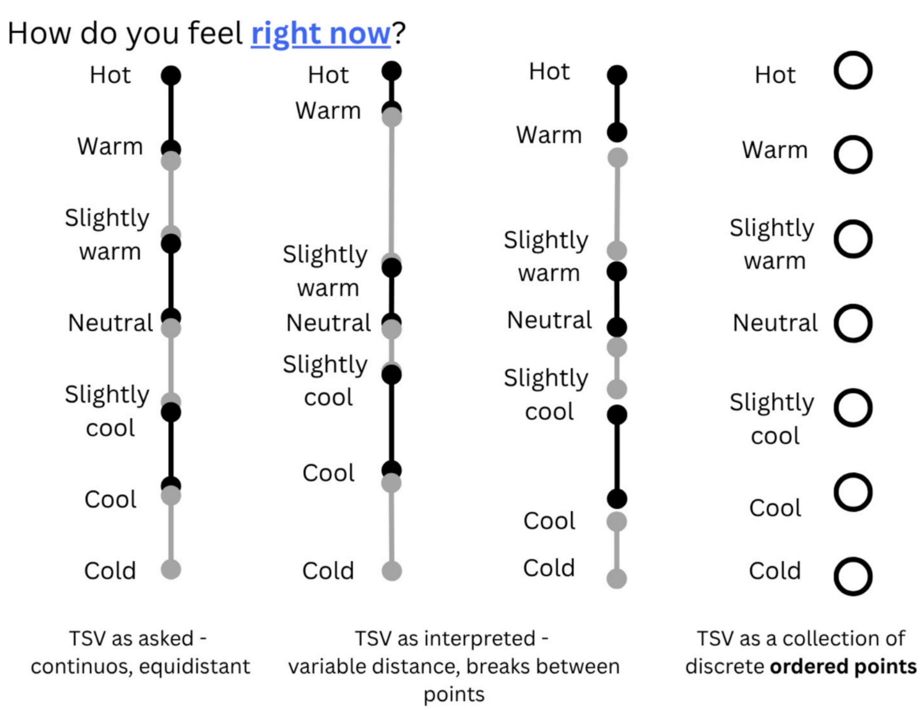


**Fig. 1.** An illustration of the ASHRAE seven-point thermal sensation vote question, with an emphasis on the "right now" sensation. Column 1 represents a continuous and equidistant interpretation as designed by the researcher. Column 2, 3 and 4 represent examples of how a respondent might interpret a scale; with variable distance between points; with breaks; and as a collection of seven ordered points, without assuming anything about the distance or continuity.

'Cold,' $\geq -2.5$ and $<$ -1.5 as 'Cool' and so on [4]. This processing was used for consistency even though the specific dataset did not use a continuous scale for TSV. The choice of cut-off points used for mapping the continuous scale back to an ordinal scale are a matter of convenience and a limitation of the data. Study designs that use and analyse TSV as ordinal data will not face this limitation. All modelling was conducted using the R software [15] (code available at: https://github.com/asitkm76/TSVandOrdinalModels).

The model output, in the form of predicted probabilities for TSV categories based on SET, is presented in Fig. 2b, along with the more traditionally used linear fit in Fig. 2a, for reference.

The peak probability for 'neutral', ~ 39 %, is predicted at 26.7°C. A combined probability of 80 % summed for the slightly cool, neutral, and slightly warm thermal votes corresponds to the range of 23.2–30.4 °C, and 85 % corresponds to the range of 25.1–28.5 °C. While the SET for peak probability of 'neutral' is similar to the neutral temperature from the linear regression model (26.5 °C), the "comfort zones" are quite different. The probability curve for neutral indicates an asymmetry around its peak for cool and warm perceptions. This is opposed to the symmetrical comfort zone around neutral temperature in linear models, which is sometimes extended beyond the measured conditions. This zone is interpreted as representing 80–90 % occupant satisfaction.

## 4. Discussion and outlook

Thermal comfort as a "condition of mind," (Fig. 3) expressing "satisfaction with the thermal environment," and being evaluated subjectively is a simple yet comprehensive definition [16]. This also complicates the work of building engineers. We need to consider the physics of the problem, appropriate choice of models, and improving model accuracy while trying to ensure the respondents are clear on the question and that we make no assumptions about human experience.

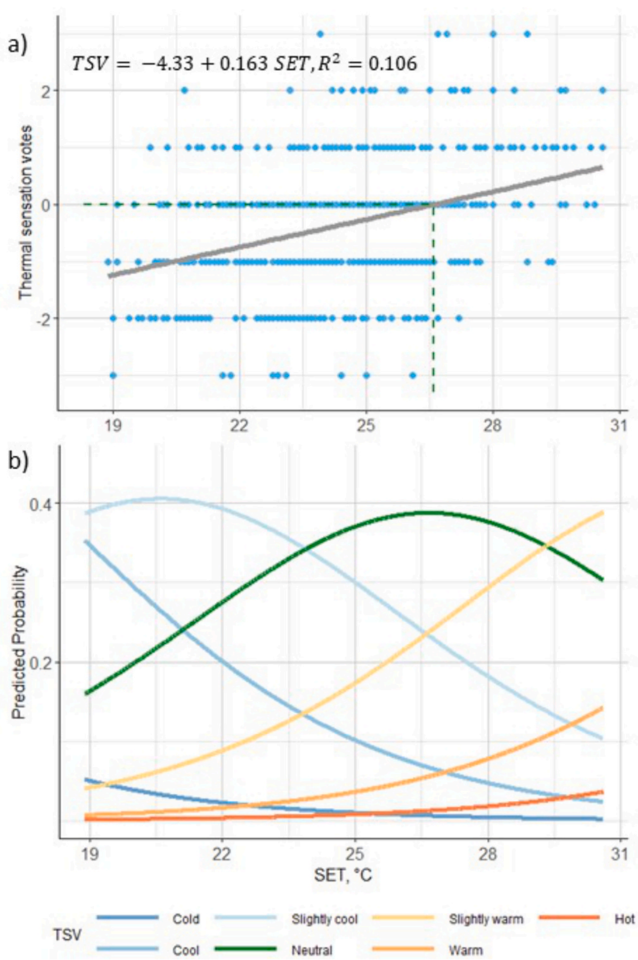


**Fig. 2.** A) linear fit of tsv ~ SET b) Predicted probability of TSV being in a certain category, as a function SET, from the ordinal model.

A major difference between the traditional PMV-PPD and the adaptive thermal comfort (ATC) models is acknowledging that occupants are not passive, i.e., they actively respond to and alter their thermal environment. But such actions also mean that we must carefully evaluate how occupants interpret the thermal comfort survey questionnaire.

At a certain point, occupants may feel they do not need any further adaptive actions, or even they can run out of adaptive options. Socio-cultural aspects, climate, past experience, availability of adaptive opportunities, co-occupants' sensation, and the ease, effectiveness, and economy of using a specific adaptive opportunity [17], will affect occupant actions. The occupant actions, in turn, affect thermal sensation and subsequently could affect their survey responses. While in many buildings and work environments, there are limited or no potential occupant actions, some occupant actions and experiences can impact both TSV and SET, or impact one through the other. Nevertheless, temporality implies that right-now TSV is much less likely to predict right-now SET or past SET, than right-now TSV predicting future SET of right-now/past SET predicting current TSV. On a timeline, cause must precede effect. Evaluation of TSV and SET as time series can help better address the TSV-SET interrelation and the impact of occupant actions.

TSV, as described by the standardized questionnaire scale, is an ordinal variable. The relationship between thermal sensation and the thermal environment is non-linear [4]. The ordinal nature of TSV suggests linear modelling is not optimal.

Occupant behaviour predicates that TSV-SET can affect each other. The right question here need not be which affects the other. Rather, knowing that the relation goes both ways, the question should be how we progress with this understanding to better ensure thermal comfort for occupants.

It is important that we should not try to align predictive and causal models (Fig. 3, Modelling). For the designer or the engineer, predictive models have their value. For example, in the ATC model, indoor comfort temperature can be predicted by representations of outdoor temperature (daily mean or running mean temperature). However, that relation should not be used to assert that it is the outdoor mean temperature which causes indoor comfort. Similarly, a cause does not have to be a good predictive parameter to be a cause. For example, wearing a sweater can be a cause of thermal comfort in a cold space. However, if we try to use "wearing sweaters" to predict overall thermal comfort levels in a building, it would not be very effective.

The TSV ~ SET ordinal model (Fig. 2b) allows better reflection of percentages of acceptance and dissatisfaction. It reinforces inter-individual variability and the fact that thermal sensation is not deterministic. This impacts the solutions we may design for benefiting the optimal thermal environment indoors in particular and indoor environmental quality in general. It takes us away from a singular set-point and zone-based control to a paradigm where we try to accommodate inter-individual differences (Fig. 3, Towards the future) because that is what the comfort model is telling us.

The previous decades of thermal research have amply demonstrated that occupant thermal sensations are much more stochastic than deterministic. Available volume of data and computational resources in thermal comfort research have grown remarkably over the past decade. Improving model prediction accuracy using complex modelling should not come at the expense of model suitability [2]. The OM, in addition to being the appropriate statistical fit, encourages researchers and building designers to think of how they may better serve the needs of building occupants. This is an important benefit of the thought process that we are promoting in this work.

To move towards future models (Fig. 3) that reflect our improved understanding of human-built environment interactions, we can take a

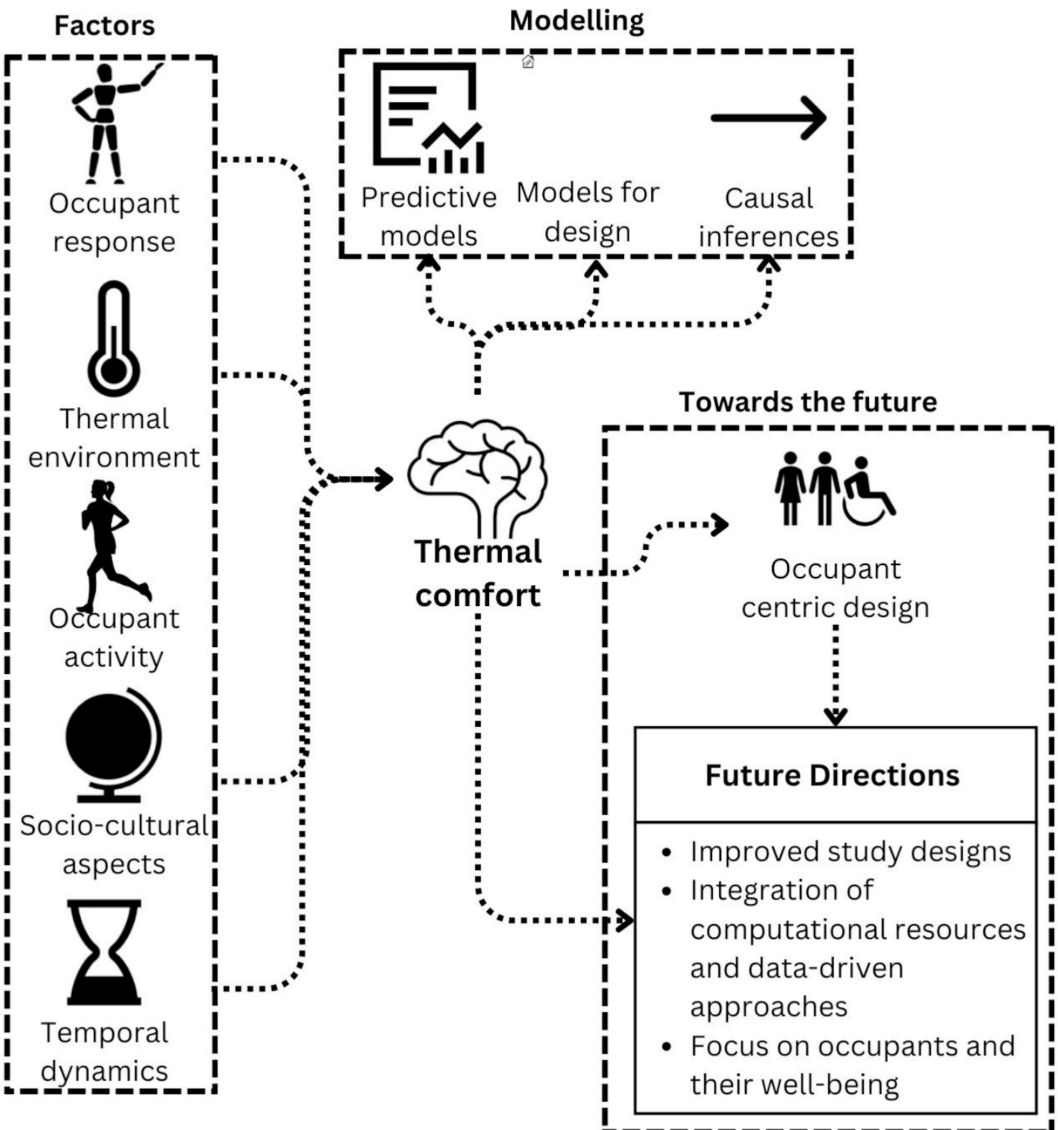


**Fig. 3.** A schematic of thermal comfort at the centre of the factors that influence it, our efforts at modelling it, and a pathway towards the future.

step back and reevaluate our study designs (to have the right kind of data) and analyses frameworks. Our goal should be to foster constructive discussions about advancing research while acknowledging the contributions of the PMV-PPD and ATC models. In survey-based research, study design must consider which effects are of interest and for whom they would be important, even before gathering data or deciding about models [18], keeping in mind the occupants we hope to benefit.

We envisage this work to initiate and support discussions regarding the interpretation and analysis of thermal comfort related data. The overarching goal is improved designs as well as modelling and prediction and thus to better serve the occupants by addressing variability in thermal sensations, perceptions and preferences.

## CRediT authorship contribution statement

**A.K. Mishra:** Writing – review & editing, Writing – original draft, Software, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **P Wargocki:** Writing – review & editing, Supervision, Methodology, Investigation, Funding acquisition, Conceptualization. **E.J. O'Reilly:** Writing – review & editing, Supervision, Methodology, Investigation, Funding acquisition, Formal analysis, 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.

## Acknowledgement

AKM is a DOROTHY-MSCA fellow, partly funded by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 101034345.

## Data availability

Data is open source and code shared

## References

[1] J. Pan, A. Mahdavi, I. Mino-Rodriguez, I. Martínez-Muñoz, C. Berger, M. Schweiker, The untapped potential of causal inference in cross-modal research, Build. Environ. 15 (248) (2024 Jan) 111074.

[2] M. Favero, S. Carlucci, G. Chinazzo, J.K. Møller, M. Schweiker, M. Vellei, et al., Ten questions concerning statistical data analysis in human-centric buildings research: a focus on thermal comfort investigations, Build. Environ. 5 (2024 Aug) 111903.
[3] R. Sun, S. Schiavon, G. Brager, E. Arens, H. Zhang, T. Parkinson, et al., Causal thinking: Uncovering hidden assumptions and interpretations of statistical analysis in building science, Build. Environ. 1 (259) (2024 Jul) 111530.
[4] M. Schweiker, X. Fuchs, S. Becker, M. Shukuya, M. Dovjak, M. Hawighorst, et al., Challenging the assumptions for thermal sensation scales, Build. Res. Inf. 45 (5) (2017 Jul 4) 572–589.
[5] M. Schweiker, M. André, F. Al-Atrash, H. Al-Khatri, R.R. Alprianti, H. Alsaad, et al., Evaluating assumptions of scales for subjective assessment of thermal environments–do laypersons perceive them the way, we researchers believe? Energy Build. 211 (2020) 109761.
[6] AP Gagge, JA Stolwijk, JD Hardy, Comfort and thermal sensations and associated physiological responses at various ambient temperatures, Environ Res 1 (1) (1967) 1–20.
[7] Bedford T. The Warmth Factor in Comfort at Work. A Physiological Study of Heating and Ventilation. [Internet]. London: Medical Research Council, Industrial Health and Research Board; 1936 [cited 2025 Aug 20]. Report No.: 76. Available from: https://www.cabidigitallibrary.org/doi/full/10.5555/19362702015.
[8] PO. Fanger, Assessment of man's thermal comfort in practice, Occup Environ Med 30 (4) (1973) 313–324.
[9] DA. McIntyre, Thermal sensation. A comparison of rating scales and cross modality matching, Int J Biometeorol 20 (4) (1976 Dec) 295–303.
[10] Z. Wang, H. Zhang, Y. He, M. Luo, Z. Li, T. Hong, et al., Revisiting individual and group differences in thermal comfort based on ASHRAE database, Energy Build. 15 (219) (2020 Jul) 110017.
[11] M. Schweiker, R.K. Andersen, S. Carlucci, G. Chinazzo, S. Hodder, A. Mahdavi, et al., Ten questions concerning the usage of subjective assessment scales in research on indoor environmental quality, Build. Environ. 113393 (2025).
[12] J. Antonakis, S. Bendahan, P. Jacquart, R. Lalive, Causality and endogeneity: Problems and solutions, Oxf Handb Leadersh Organ. 1 (6) (2014) 93–117.
[13] Č. Brukner, Quantum causality, Nat. Phys. 10 (4) (2014 Apr) 259–263.
[14] Parkinson T, Tartarini F, Földváry Ličina V, Cheung T, Zhang H, De Dear R, et al. ASHRAE global database of thermal comfort field measurements [Internet]. Dryad; 2018 [cited 2024 Aug 13]. p. 182407350 bytes. Available from: https://datadryad.org/stash/dataset/doi:10.6078/D1F671.
[15] R Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2024. Available from: https://www.R-project.org/.
[16] ANSI/ASHRAE. Standard 55. Thermal environmental conditions for human occupancy [Internet]. Georgia, USA: ASHRAE; 2023 [cited 2024 Aug 11]. Available from: https://www.ashrae.org/technical-resources/bookstore/standard-55-thermal-environmental-conditions-for-human-occupancy.
[17] A.K. Mishra, M. Ramgopal, Field studies on human thermal comfort — an overview, Build. Environ. 64 (2013 Jun) 94–106.
[18] J.B. Sande, M. Ghosh, Endogeneity in survey research, Int. J. Res. Mark. 35 (2) (2018 Jun 1) 185–204.