Determining the Effects of Sound Levels on Physiological Wellbeing in the Workplace – A Field Study Using Wearable Devices

Karthik Srinivasan[1], Faiz Currim[1], Sudha Ram[1], Casey M. Lindberg[2], Javad Razjouyan[3], Brian Gilligan[4], Hyoki Lee[3], Kelli J Canada[4], Nicole Goebel[5], Matthias R Mehl[6], Melissa M Lunden[5], Judith Heerwagen[4], Bijan Najafi[3], Esther M. Sternberg[2], Kevin Kampschroer[4]

Keywords: Physiological wellbeing, wearables, sound-wellbeing, environment-wellbeing, curvilinear effects, simultaneous modeling of multiple outcomes, individual heterogeneity effects, statistical machine learning.

[1] INSITE: Center for Business Intelligence and Analytics, Department of Management Information Systems, Eller College of Management, University of Arizona, Tucson, AZ.
[2] Institute on Place and Wellbeing, University of Arizona, Tucson, AZ.
[3] Interdisciplinary Consortium on Advanced Motion Performance (iCAMP), Michael E. DeBakey Department of Surgery, Baylor College of Medicine, Houston, TX.
[4] US General Services Administration, Office of Federal High-Performance Washington, DC.
[5] Aclima, Inc. San Francisco, CA.
[6] Department of Psychology, University of Arizona, Tucson, AZ.

Abstract

In today's world, people spend a significant proportion of their active hours in enclosed workplaces. Workplace environment is closely tied to an individual's wellbeing. We conduct a field study using wearable devices to determine the impact of indoor environment on individual wellbeing. In our study, participants carry out their day to day activities while wearing sensors that continuously record their physiological wellbeing state and ambient environmental conditions including sound level. We observed that the relationship between sound level and two physiological wellbeing measures (i.e., SDNN, normalized-HF) is curvilinear and varies across individuals. For modeling the dynamics of the sound-wellbeing relationship, we propose new methods for representing curvilinear effects, simultaneous modeling of multiple outcomes, and identifying factors contributing to between-individual heterogeneity. We show that our methods have better model fit as well as predictive performance than existing methods for each of the three modeling problems. We find that an individual's physiological wellbeing is optimal when sound level in the workplace is 50 dBA. For sound amplitudes lower than 50 dBA, a 10 dBA increase in sound level is related to a 3.6% increase in physiological wellbeing; whereas for amplitude above 50 dBA, a 10 dBA increase in sound level is related to decrease in physiological wellbeing by 1.3%. Age, body-mass-index, high blood pressure, anxiety and computer use intensive work are person level factors contributing to heterogeneity in effects of sound level on physiological wellbeing across individuals. Workers with higher blood pressure are more negatively affected by increase in sound levels than others. Workers with computer intensive work are more negatively affected by sound level extremities of low sound (i.e., quietude) or high sound (i.e. noise) than others. Our study informs policies and practices that affect the health and wellbeing of office workers worldwide. It proposes new quantitative methods to address key challenges in modeling digital data generated using wearable devices.

1. Introduction

Nearly 50 million workers in the United States spend over one-fifth of their time at their workplace (Bureau of Labor Statistics 2017). On an average, four out of ten US workers report their job and workplace to be stressful and affecting their wellbeing (Harvard School of Public Health 2016). Workplace-related stress and absenteeism cost up to $225 billion, or more than 10% of office workers' contribution to the US GDP (CDC Foundation 2018). Research shows that our workplace environment is closely tied to our mental state, productivity, and physiological wellbeing (Heerwagen and Zagreus 2005; MacNaughton et al. 2016; Thayer et al. 2010).

Workplace sound level[7] has been identified as a significant stressor in the past (Frontczak et al. 2012; Seidman and Standring 2010). The psychosocial effects of high sound levels (i.e., loud noise) including its effects on satisfaction, environmental control, social interaction, social support, and perceived insensitivity to social cues in a workplace setting have been studied previously (Rashid and Zimring 2008). However, the underlying mechanism of sound effects on wellbeing at workplace is not yet fully understood (Kraus et al. 2013). Existing sound-wellbeing studies have employed controlled experiments with few subjects, limited set of treatments and controls. They suffer from low external validity, as the real office environment ecosystem is more complex than simulated environments. Current understanding of sound level effects on physiological wellbeing (i.e., sound-wellbeing) at the workplace is also limited by the ability of standard methods (e.g., analysis of variance, linear regression, etc.) to capture different aspects of the sound-wellbeing relationship.

We conducted a field study using multiple wearable devices to determine the impact of indoor environment on individual wellbeing. In our study, participants carry out their day to day activities while wearing sensors that continuously record their (short-term) physiological wellbeing state and ambient

[7] Section 3 of supplementary materials presents a reference for subjective reference of different sound levels in real world

environmental conditions including sound level. We observed that the relationship between sound level and two physiological wellbeing measures (i.e., SDNN, normalized-HF) is curvilinear and varies across individuals. We propose new methods to address challenges in representing curvilinear effects, simultaneous modeling of multiple outcomes, and identifying factors contributing to individual heterogeneity in effects. The first method is a semi-automated method for change point determination in multilevel segmented regression models for representing curvilinear relationships. The second method is a Bayesian latent variable modeling method for simultaneous modeling of multiple outcomes. The third method tackles the problem of modeling individual heterogeneity in the sound-wellbeing relationship using a two-step approach; modeling the heterogeneity using a Bayesian varying coefficients models followed by identification of person-level factors contributing to the heterogeneity using regularized models. We show that our proposed methods have better predictive performance than existing methods and are elemental in developing critical insights about the sound-wellbeing relationship in the workplace.

Our major findings on the workplace sound-wellbeing relationship are as follows. A sound level threshold value of 50 dBA (averaged at 5-minute interval) is optimal for physiological wellbeing. A 10 dBA increase in sound levels below 50 dBA is related to a 3.6% increase in physiological wellbeing whereas a 10 dBA increase in sound levels above 50 dBA is related to decrease in physiological wellbeing by 1.3%. Blood pressure level and computer intensive work are two individual personality factors that moderate the relationship between sound levels and physiological wellbeing. People with higher blood pressure are negatively affected by increase in lower as well as higher sound levels. Computer intensive work is related to an amplification of positive effect of lower sound levels as well as negative effect of higher sound levels. Our findings show that quiet workstations are optimal for the physiological wellbeing of office-workers with high blood pressure, whereas workspaces with moderate sound levels (~50dBA) are suitable for workers with computer intensive work.

In this study, we make three major contributions. First, to our knowledge, our paper is the first to implement a field study using multiple wearable devices to model the environment-wellbeing phenomenon.

Second, our study proposes new quantitative methods addressing three key challenges in modeling digital data generated by wearable devices (i.e., curvilinear modeling, simultaneous modeling, individual heterogeneity effects modeling). Lastly, it unravels aspects of the sound-wellbeing relationship that were unknown earlier, thus informing workplace planning policies and practices that affect the health and wellbeing of office workers worldwide.

The rest of this paper is organized as follows. In Section 2, we introduce the study background and related literature. In Section 3 and Section 4, we describe our field study using wearable devices and the need for new modeling methods. In Section 5, we expound on the three new methods for sound-wellbeing modeling. Section 6 contains the evaluation of our methods and key-learning based on their application on our data, followed by the discussion and conclusions in Section 7 and Section 8 respectively.

2. Background and Related Literature

2.1 What is Physiological Wellbeing?

Psychological well-being consists of positive relationships with others, personal mastery, autonomy, a feeling of purpose and meaning in life, and personal growth and development (Ryff 1989). On the other hand, physiological wellbeing is associated with a dynamic, ever-adapting balance in the human physiological system conditioned by momentary demands (Malik et al. 1996). When we are in good health or at a higher physiological wellbeing state, we experience flexibility and resilience in relation to our environment and experiences.

Stress is a major factor that impacts physiological wellbeing (Boron and Boulpaep 2012). The physiological stress response or physiological response to the demands put upon the body has a direct relationship with the two components of the autonomic nervous system (ANS) - Sympathetic nervous system (SNS) and the parasympathetic nervous system (PNS). When the body is stressed, the SNS generates the "fight or flight" response where the body shifts all of its energy resources toward fighting off a life threat. Whereas the PNS indicates the "rest and digest" and is involved with restoration and repair, nourishment and detoxification processes in the body. Heart rate variability (HRV) reflects the modulation in the normal rhythm of the heart and is known to assess overall cardiac health and the state of the ANS. HRV is considered as a proxy measure for the physiological wellbeing of a person, i.e., higher its value, higher the wellbeing (Xhyheri et al. 2012). HRV is more widely used when compared to other physiological stress response measures such as salivary cortisol and skin conductance as it can be measured at short intervals using commercially available heart rate monitors and is least intrusive (Acharya et al. 2006; Xhyheri et al. 2012). While many measures of HRV exist, each serves as a slightly different lens in terms of viewing the body's physiological stress response (Shaffer and Ginsberg 2017). The mean of standard deviation for all successive R-R intervals (SDNN) is a global index of HRV and reflects longer term circulation differences or the overall activity in the ANS. Normalized high frequency component (normalized-HF) is the ratio between absolute value of the High Frequency and difference between Total Power and Very Low Frequency bands in the power spectrum of frequency domain of heart rate that emphasizes changes in parasympathetic regulation. High values of SDNN and normalized-HF have consistently been found to indicate better health and wellbeing (Soares-Miranda et al. 2014).

2.2 Workplace and Wellbeing

White-collared office workers in USA spend a majority of their active hours in a day in enclosed workplaces (Bureau of Labor Statistics 2017). The workplace environment not only affects people at work (Heerwagen and Zagreus 2005), but is also known to have a carry-over effect on our personal lives outside office (Lindberg et al. 2018). Workplace characteristics consist of elements such as workstation design (i.e., workstation type, workstation area, furniture, nature view, etc.), indoor environment quality (i.e., ambient sound levels, temperature, humidity, air quality, etc.), social influence (i.e., interaction with colleagues in close vicinity, proximity with team members, tele-working facilities, etc.) and amenities (i.e., proximate breakout areas, availability of quiet spaces, control over thermostat and window-blinds, etc.). Indoor environment quality of the workplace is closely tied to mood, productivity, activity and longevity of office workers (Backé et al. 2012; Kivimäki et al. 2012). Air-quality factors such as carbon dioxide are shown to have impair cognitive performance and degrade physiological wellbeing at workplace (MacNaughton et al. 2016). Workstation type (e.g., open bench seating, cubicles, private space, etc.) and structure (e.g., length of passage, furniture, workstation area, etc.) also has an effect on worker's stress and physiological wellbeing (Lindberg et al. 2018; Thayer et al. 2010). Intervention based controlled experiments and analysis of variance have been the traditional approaches in analyzing the indoor environment-wellbeing phenomenon (MacNaughton et al. 2016; Rashid and Zimring 2008).

2.3 Workplace sound levels and Wellbeing

Sources of sounds in offices include other people's conversations, telephone-calls, and mechanical equipment. High ambient sound levels or noise in the work environment is reported to be one of the highest stressors in workplaces in the US (Heerwagen and Zagreus 2005). There is substantial literature on the effects of sound in office settings on different aspects of wellbeing in workplace such as social, psychological, physiological and performance (Jahncke et al. 2011; Kjellberg et al. 1996; Kraus et al. 2013; Rashid and Zimring 2008). Table 1 lists literature on effects of sound levels on physiological wellbeing in chronological order.

Table 1: Literature on effect of workplace sound levels on social, psychological and physiological wellbeing

Study	Input	Outcome(s)	Study design	Findings
(Lusk et al. 2002)	Areas with sound levels averaged across a 5 years interval	Blood pressure and heart rate	N=374; Correlating person-level noise exposure with physiological wellbeing; Method: Linear regression	Areas with high sound levels are predictive of increase in blood pressure
(Lee et al. 2010)	Discrete sound levels	HRV (LF, LF/HF), Mean blood pressure, Mean heart rate	N=16; Treatment = Sound level exposure of No noise, 50 dBA, 60 dBA, 70 dBA and 80 dBA for 5 minutes with 2 minutes interval; Method: Repeated measures ANOVA; Spearman's Rho	HRV decreases with higher sound level exposures, but no change in blood pressure and mean heart rate

(Jahncke et al. 2011)	Noisy background, river sounds, nature movie	Cortisol, Catecholamines, self-rating of tiredness, mood	N=47; Treatment = Completed tasks for 2 hours each in a low and high noise conditions; Repeated measures ANOVA	Though noisy background and river sounds have an effect on psychological outcomes, they had not significant effect on physiological outcomes
(Kraus et al. 2013)	Sound levels	HRV (LF/HF, SDNN, RMSSD)	N=110; Prospective panel study with participants spending upto 7.5 hours in a room; Method = Additive mixed models	Sound levels have a positive effect below 65 dBA on SDNN, but is not significantly related to any of the other outcomes
(Sim et al. 2015)	Sound types, sound levels	HRV (SDNN, HF, LF/HF)	N=40; Treatment: 45 dBA exposure for 5 minutes; Method: Linear regression	Sound types do not have a significant effect on physiological outcomes
(Walker et al. 2016)	Noise exposure at 75 dBA at low frequency and high-frequency	HRV (SDNN, LF, RMSSD), blood pressure, salivary cortisol	N=10; Treatment = 40 minutes noise exposure; Method=Multivariate multilevel regression	High sound levels at low-frequencies and high-frequencies negatively affect physiological wellbeing
(Park and Lee 2017)	Floor impact noises ranging from 31.5 dBA to 63 dBA	Noticeability, Annoyance, Heart rate, electrodermal activity, respiration rate	N=21; Treatment = 5 sessions of 15 minutes of different floor impact noises; Method=Repeated measures ANOVA	Annoyance, noticeability, electrodermal activity and respiration rate increases with sound level, but no significant change in heart rate
(Cvijanović et al. 2017)	Sound levels	Mental effort, HRV (LF, LF/HF) and skin conductance	N=40; Treatment = 6 dBA background noise added while participants completed collaborative tasks; Method=Multilevel regression	Though mental effort required increases with sound levels, effect on physiological wellbeing was not significant

Early sound-wellbeing studies shown that higher sound frequencies as well as higher sound levels are detrimental to a worker's psychological and physiological wellbeing (Kjellberg et al. 1996; Landström et al. 1995; Lusk et al. 2002). However, the effects are not dependent on underlying sound type (i.e., conversation, mechanical noise, traffic sounds, etc.) (Sim et al. 2015). Recent studies identified that effect of sound levels on physiological wellbeing is not monotonic (Kraus et al. 2013), and are instantaneous

(Srinivasan et al. 2017). Previous studies employ controlled experiments with few subjects and standard modeling methods to determine the sound-wellbeing relationship, and are either inconclusive or contradicting each other (Cvijanović et al. 2017; Park and Lee 2017; Sun Sim et al. 2015). Even though, sound-wellbeing relationship is not linear (Kraus et al. 2013), the nature of its relationship (i.e., where the function is optimal, how wellbeing varies as a function of sound level) is still not clear. Further, reporting results from multiple models corresponding to different measures of physiological wellbeing (Cvijanović et al. 2017; Park and Lee 2017; Sim et al. 2015) makes it difficult for interpretation and decision-making based on findings. Finally, even though previous studies account for the heterogeneity in sound effects on wellbeing by fitting multilevel regression model, the heterogeneity has not been modeled explicitly.

3. Field study using wearable devices

We conducted a multi-phase field study between 5th May 2015 and 25th August 2016 as part of the US General Services Administration's Wellbuilt-for-Wellbeing (WB2) research program to understand the impact of indoor environment on wellbeing of office-workers (Sternberg et al. 2016). In the study, self-described healthy adult workers involved in a variety of office-based roles for the US government were recruited across four federal office buildings in the Mid-Atlantic and Southern regions of the USA. Buildings were selected for their representation of common office workstation types across the US General Services Administration's portfolio of over 370 million square feet of office space that houses over 1 million employees. Staff in sections of each office building from organizations with leadership approval were offered the opportunity to participate. After giving written informed consent, participants completed an intake survey consisting of demographic questions. Participants wore two sensors for three days while carrying out their day-to-day activities, (a) a heart and physical activity monitor, and (b) a personal environment quality sensor-based device. The study also included experience sampling mobile surveys to collect perceived (psychological) responses of individuals at periodic intervals of 1-2 hours, stationary environmental sensors mounted on multiple walls in the study areas. Figure 1 shows a visual representation of the data collection mechanism used in the study using wearable devices, mobile-based surveys and wall-mounted sensors. The heart and physical activity monitor is a chest-worn wearable device named *EcgMove* 3 developed by *movisens* (Verkuil et al. 2016). The personal-environment quality sensor-based device is a multi-modal sensing neckwear device developed by *Aclima, Inc.* that measures ambient environment conditions including sound levels. Sound level exposure was measured as A-weighted continuous sound pressure levels reported in units of A-weighted decibels (dBA); a measurement of the relative loudness of sounds relative to absolute silence as perceived by the human ear.

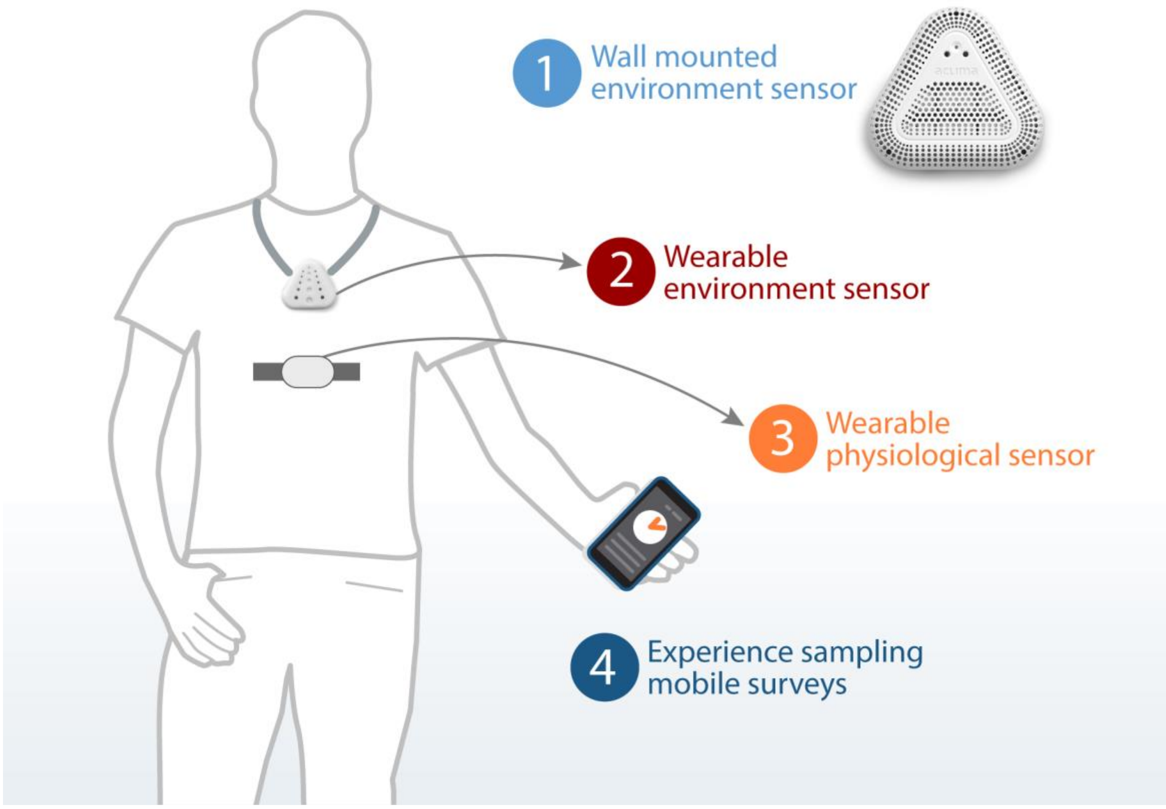


Figure 1: WB2 study data collection mechanism consisting of two wearable sensors, a mobile survey application and wall mounted sensors

Our study tracks effects of environment factors on physiological wellbeing of workers in the naturalistic setting of an office workplace without doing any interventions. Our comprehensive data collection mechanism consists of multiple wearable sensors, mobile survey prompts, wall-mounted sensors and other offline information (i.e., calendar data, workstation characteristics, work type, etc.). Sensors are worn by multiple participants for a long period of time with participation from different office buildings over an entire year.

4. Preliminary analysis

We trained a (two-level) multilevel regression model for our data for different physiological outcomes as shown below:

$$y_{ij} = \beta_0 + \gamma_{0j} + \sum_{k=1}^{K} \beta_k x_{kij} + \sum_{m=1}^{M} \gamma_{mj} z_{mij} + \epsilon_{ij} \tag{1}$$

In Equation (1), y_{ij} is the physiological wellbeing measure (i.e., SDNN, RMSSD, normalized-HF) for the i^{th} observation and j^{th}individual, β_0 is the fixed intercept, $\mathrm{B} = \{\beta_1, \dots, \beta_K\}$ are coefficients for K fixed effects $\{x_1, x_2, \dots, x_K\}$, $\Gamma_0 = \{\gamma_{01}, \gamma_{02}, \dots, \gamma_{0j}, \dots, \gamma_{0J}\}$ are J random intercepts for each individual, $\Gamma = \{\gamma_{11}, \dots, \gamma_{1j}, \dots, \gamma_{MJ}\}$ are coefficients for $M\ X\ J$ random effects $\{z_1, z_2, \dots, z_M\}$, and ϵ_{ij} is the residual error. We assume a variance component structure for the covariance matrix of the random effects coefficients, since it makes the least assumptions (Raudenbush and Bryk 2002). Sound level was included as a fixed effect as well as random effect in the model. We considered including higher-levels in the multilevel model (i.e., organization type, buildings, participant cohorts, work type, etc.) but the model fit did not improve significantly and hence we restrict the model to vary at two levels, i.e., for variables varying at within-individual level (level-1), and for variables varying at between-individual level (level-2) in the data.

Consistent with previous studies (Kraus et al. 2013; Lee et al. 2010), we observed a curvilinear relationship between sound levels and physiological wellbeing measures (i.e., SDNN, normalized-HF) in two dimensional scatter plots. We found that the fixed-effect of sound level variable in the two univariate multilevel regression models for SDNN and normalized-HF as outcomes were significant, i.e., $\beta_{Sound,SDNN} = 0.1038\ (p = 0.0000)$, $\beta_{Sound^2,SDNN} = -0.0075\ (p = 0.0000)$, $\beta_{Sound,normalized-HF} =$

-0.0979 ($p = 0.0000$), $\beta_{Sound^2, normalized-HF} = 0.0013$ ($p = 0.015$). Thus, we can infer that sound level has a curvilinear effect on SDNN and normalized-HF.

Secondly, we found that including sound level as a random-effect improve the quality of fit of the multilevel model. This implied that the effect of sound level on the two physiological wellbeing measures varies across individuals. To study the sound-wellbeing relationship further, we therefore required methods to (a) Model the curvilinear relationship of sound level on two physiological wellbeing measures (i.e., SDNN and normalized-HF), (b) Simultaneously model the effects of sound level on two different measures of physiological wellbeing (i.e., SDNN and normalized-HF), and (c) Model the heterogeneity in the effects of sound on physiological wellbeing. In the next section, we consider each modeling problem, review corresponding modeling methods literature, identify the research gaps, and present new methods to address the corresponding modeling challenges.

5. Modeling Methods

5.1 Modeling curvilinear effects

Research question: How can we effectively model curvilinear relationships between sound level and the two wellbeing measures (i.e., SDNN and normalized-HF)?

5.1.1 Existing methods

A curvilinear relationship between an input and an outcome is commonly observed in IS (Liu and Goodhue 2012; Pant and Srinivasan 2010; Xue et al. 2011) and other disciplines (Geng et al. 2017). Polynomial regression models account for higher order relationships but they are not directly interpretable (Durban et al. 2005). Segmented regression is an optimal approach for modeling curvilinear relationships as it is robust, has fewer underlying assumptions and is easier to interpret (Jirschitzka et al. 2016). The primary challenge in using a segmented regression approach is the determination of change points linking the input segments (Shuai et al. 2003). Common procedures to determine change points in segmented regression models for simple data (Shuai et al. 2003) cannot be used for modeling the sound-wellbeing relationship as determination of the likelihood function is not straightforward for the multilevel data structure in our study. A recent method was proposed based on maximum-likelihood estimation of a continuous functional approximation of the piece-wise linear function (Muggeo et al. 2014) as an alternative to subjective assignment based on visualization of pair-wise plots (Kraus et al. 2013). However, this method estimates multiple change points automatically with no scope for user inputs into the estimation process (e.g., including or dropping change points if they are at extremities of the input distribution, etc.). Therefore, existing procedures for determining change points in studies employing segmented models for sound-wellbeing are either ad-hoc or analytically complex, leading to problems such as low external validity and overfitting respectively. To summarize, there is a need for a validated method to determine the change points in segmented multilevel models that is robust, efficient and transparent. With such a method, we can accurately determine change points that create piece-wise linear functions of the sound-wellbeing relationship facilitating direct interpretation from the linear model.

5.1.2 Semi-automated change point determination method

Consider the multilevel model described in Section 4. Sound level as an input variable varying at level-1 (i.e., fixed-effect) and having a curvilinear relationship with the outcome can be expressed as sum of segmented variables as follows:

$$x_r = x_{1r} \cdot I(x_r < \eta_1) + x_{2r} \cdot I(\eta_1 \leq x_r < \eta_2) + \cdots + x_{kr} \cdot I(\eta_k \leq x_r) \tag{2}$$

In Equation (2), $\mathrm{H} = \{\eta_1, \eta_2, \ldots, \eta_k\}$ is a set of k change points defined for the input variable x_r. $I(\varphi)$ is an indicator function equal to 1 if condition φ is true else 0. As can be seen, the problem here is to estimate each change point η_i as well as to determine the total number of change points k. We propose a three step semi-automated method to estimate the change points and determine k as follows:

Step I: Fit a Generalized additive mixed model and visualize component smooth functions

Fit input x_r as a non-parametric spline in a Generalized Additive Mixed Model (GAMM) and visualize its component smooth function (Faraway 2006). Identify the order of the curve by inspecting the number of extrema (i.e., minima and maxima) and set value of k. Note the value of the maxima and minima to be used as starting points in a linear search algorithm in the next step. This step also is used to determine whether or not to opt for a segmented model, over a linear model, by inspecting the curvilinear nature of component smooth function.

Step II: Perform a linear search for change points using optimization with box constraints.

Consider a model fit metric such as Akaike information criteria (AIC), Bayesian information criteria (BIC), Deviation or Mean-squared error as an optimizing function. Select a suitable range around each starting point selected in step I and run the optimization algorithm with the given range as a box constraint (Brent 2013). Select set of change points H that maximize model fit.

Step III: Fit the segmented multilevel regression model as shown below:

$$y_{ij} = \beta_0 + \gamma_{0j} + \sum_{s \in S} \beta_{rs} x_{rij} I\left(x_{rij} \in s\right) + \sum_{k=1,\ k \neq r}^{K} \beta_k x_{kij} + \sum_{m=1}^{M} \gamma_{mj} z_{mij} + \epsilon_{ij} \tag{3}$$

In Equation (3), S is a set of segments constructed using change points H identified in Step II for input x_r. Significance of the effect of input variable x_r at each segment s can be determined by inspecting the corresponding fixed effects coefficient β_{rs}, under regular conditions.

5.2 Simultaneous modeling of multiple outcomes

Research question: How can we simultaneously model the relationship between sound level and two wellbeing measures (i.e., SDNN and normalized-HF)?

5.2.1 Existing methods

Existing studies analyzing effect of sound level on multiple physiological wellbeing measures fit a different model for each outcome and report coefficients for each of the models separately (Cvijanović et al. 2017; Kraus et al. 2013; Park and Lee 2017; Sim et al. 2015). Interpretation and communication of results from multiple models for decision-making can be challenging. A statistical model with a single set of coefficients for multiple outcomes is suitable for this purpose and known as simultaneous modeling (Baldwin et al. 2014; Das et al. 2004; Pituch and Stevens 2016). Simultaneous modeling differs from multivariate modeling where coefficients are estimated for each outcome along with cross-correlation parameters (Lin et al. 2017; Ritz et al. 2017). For example, for three outcome and three inputs, a simultaneous multiple regression model will contain three coefficients (excluding the intercept) whereas a multivariate regression modeling procedure will estimate nine coefficients (excluding the intercepts for outcomes) and corresponding covariance between the coefficients. Simultaneous modeling can be done by carrying out a univariate transformation of the outcomes after accounting for heterogeneity in error variances (Baldwin et al. 2014; Faraway 2016; Pituch and Stevens 2016). In the univariate transformation method, even though different outcomes have different error variances in the model, the effects of input variables are assumed to be uniform across outcomes. For example, for a model measuring stress using breathing rate and heart rate as two health indicators, one can expect that the effects of inputs on each of the outcomes are different in scale. Latent variable modeling is an alternative approach for simultaneous modeling of multiple outcomes (Muthén 2002). However, classical latent variable modeling approaches (e.g. structural equation modeling) traditionally require individual items of a latent construct to be theoretically related and have construct validity (Kline 2012). Secondly, for multilevel modeling, diagnostic checking for structural equation modeling is more challenging to satisfy all the assumptions required in the classical approach (Hox 2013). The estimation procedure becomes more complex with a large number of random effects. Hence, there is a need for a new method that can overcome challenges of existing methods. Such a method can be used for making inferences over effects of sound level on the two physiological wellbeing measures, SDNN and normalized-HF in our study.

5.2.2 Bayesian latent variable modeling method

Consider a Bayesian latent variable model (Merkle and Wang 2016) for outcomes $Y = \{y_1, y_2, \ldots, y_h, \ldots, y_H\}$ as follows:

$$y_{ih} | \theta_i, \gamma_h, \lambda_h, \sigma_{\mathrm{ih}} \sim N(\mu_{ih}, \sigma_{ih}^2) \tag{4}$$

$$\mu_{ih} = \gamma_h + \sum_{k=1}^{m} \lambda_{hk}\theta_{ik} \tag{5}$$

$$\theta_{ik} \sim N_m(0, \Phi) \tag{6}$$

For simultaneous modeling, we set $m = 1$ in the above equation and express the latent variable as an outcome of a multilevel regression model as shown below:

$$\theta_{ij} = \beta_0 + \gamma_{0j} + \sum_{k=1}^{K} \beta_k x_{kij} + \sum_{m=1}^{M} \gamma_{mj} z_{mij} + \xi_{ij} \tag{7}$$

$$\gamma_{0j} \sim N(0, \sigma_{\gamma_0}^2), \gamma_{mj} \sim N(0, \sigma_{\gamma_m}^2), \xi_{ij} \sim N(0, \sigma_\theta^2) \tag{8}$$

Upon centering the outcomes and dropping the outcome intercept parameter γ_h, we can combine the within-individual level error variances (i.e., σ_{ih}^2 and σ_θ^2). The resultant Bayesian latent variable model is represented as follows:

$$y_{hij} - \overline{y_{hij}} = \left(\beta_0 + \gamma_{0j} + \sum_{k=1}^{K} \beta_k x_{kij} + \sum_{m=1}^{M} \gamma_{mj} z_{mij}\right) \cdot \lambda_h + \epsilon_{ij}^{(h)} \tag{9}$$

$$\gamma_{0j} \sim N(0, \sigma_{\gamma_0}^2), \gamma_{mj} \sim N(0, \sigma_{\gamma_m}^2), \epsilon_{ij}^{(h)} \sim N(0, \sigma_h^2) \tag{10}$$

The above Bayesian latent variable model (i.e., Equations (9) and (10)), can be used for simultaneously modeling the effect of sound level on the two physiological wellbeing measures, SDNN and normalized-HF. The factor loadings λ_h automatically assigns different weights to each outcome (i.e., λ_1 and λ_2), thus overcoming the limitation in the existing univariate transformation-based modeling methods. The univariate transformation-based modeling method is a special case of latent variable modeling method, where we set the factor loadings of all outcomes λ_h to 1 (Refer to section 2 of supplementary materials for model representation in univariate transformation-based modeling method).

A corresponding latent variable model can be developed using a classical approach. In the classical approach, the outcomes can be considered as reflective measures for a latent construct, followed by fitting a two-level structural equation model (SEM) consisting of the input variables and the latent construct. Software such as Mplus, LISREL, EQS, lavaan, Mplus, OpenMx can fit two-level SEM with random intercepts (Skrondal and Rabe-Hesketh 2004). In the multilevel SEM model, each outcome y_{ijh} is split into a within and a between component as follows:

$$y_{ij} = \left(y_{ij} - \overline{y_J}\right) + \overline{y_J} = y_W + y_B \tag{11}$$

In Equation (11), both the within and between covariance components are treated as orthogonal and additive latent variables (Heck and Thomas 2015). The maximum likelihood estimate for parameters is derived by minimizing the overall loglikelihood which is the sum of likelihood of data from J individuals. The latent variable model using the classical approach offers much lesser flexibility than its Bayesian counterpart, as it solicits more data-related assumptions and does not account for random effects of sound level (Heck and Thomas 2015; Kline 2012).

5.3 Modeling individual heterogeneity effects

Research question: What are the factors contributing to the individual heterogeneity in effect of sound level on wellbeing?

5.3.1 Existing methods

While it is of interest to understand the overall effect of an input on an outcome in a population, insights regarding how and why effects differ across individuals can be valuable (Abrams and Hens 2015; Dingemanse and Dochtermann 2013; Gimenez et al. 2018). The random-effects indicate the presence of individual heterogeneity in effects of an input on the outcome in a multilevel model (Raudenbush and Bryk

2002). A naïve approach to identify factors contributing to between-individual heterogeneity is to introduce each factor in an interaction term with the input variable and test its significance; an approach known as slopes-as-outcomes modeling (Becker et al. 2013; Raudenbush and Bryk 2002). However, this hypothesis testing-based approach is sensitive to noise in data, increases the chances of a Type II error (i.e., even if person-level factor contributes to individual heterogeneity, it is insignificant as a moderator in the model), and becomes cumbersome as the number of potential factors increases. There are no existing validated methods for the identification of factors contributing to individual heterogeneity in effects measured by random effects in multilevel models. There is a need for a new method to identify person-level factors associated with individual heterogeneity in effects of sound level on wellbeing.

5.3.2 Varying-coefficients modeling method

We propose the varying-coefficients modeling method as a two-step procedure: (a) Step1: Quantifying heterogeneity, and (b) Step 2: Identifying factors contributing to heterogeneity. The method can be used to identify person-level variables explaining the heterogeneity in effects of sound level on physiological wellbeing across individuals.

In the first step, we fit a Bayesian hierarchical linear model with all input variables with varying coefficients having normal priors with non-zero means. Person-level variables (e.g., age, BMI, gender, etc.) are not included in the model as their value is constant for each individual (i.e., varying coefficients for person-level variables have distribution with zero variance). The hierarchical Bayesian linear model for in step 1 is given in Equations (12) and (13).

$$y_{hij} = \left(\gamma_{0j} + \sum_{m=1}^{M} \gamma_{mj} z_{mij}\right) \cdot \lambda_h + \epsilon_{ij}^{(h)} \quad (12)$$

$$\gamma_{0j} \sim N\left(\mu_{\gamma_0}, \sigma_{\gamma_0}^2\right), \gamma_{mj} \sim N\left(\mu_{\gamma_m}, \sigma_{\gamma_m}^2\right), m \in \mathbb{Z}_M, \epsilon_{ij}^{(h)} \sim N(0, \sigma_h^2) \quad (13)$$

Note that the mean values $\mu_{\gamma_0}$and μ_{γ_m}in Equation (13) are analogous to the model intercept and the corresponding fixed effect coefficients of the m^{th} variable in the Bayesian latent variable model (i.e., Equations (9) and (10)).

In the second step, we formulate the varying coefficients of sound level as an outcome of a linear model with person-level variables as the input variables as given in Equation (14).

$$\gamma_{rj} = \beta_0 + \sum_{p=1}^{P} \beta_p x_{pj} + \epsilon_j,\ \epsilon_j \sim N(0, \sigma_r^2) \tag{14}$$

In Equation (14), $\Gamma_r = \{\gamma_{r1}, \gamma_{r2}, \dots, \gamma_{rJ}\}$ are the varying coefficients for sound level in the Bayesian hierarchical linear model in step 1. $\{x_{1\cdot}, x_{2\cdot}, \dots, x_{P\cdot}\}$ are P person-level variables and ϵ_j is a normally distributed residual error varying across J individuals.

The problem of identifying person-level factors contributing to individual heterogeneity effects becomes a variable selection problem for the above linear model. Traditional stepwise feature selection methods for regression models are ridden with challenges such as sensitivity to changes in data and low external validity (Hastie et al. 2009). These challenges are particularly relevant in our problem where there are multiple person-level variables that could be factors contributing to heterogeneity in sound effects on wellbeing across individuals. We therefore choose three regularization based methods, *lasso* (Tibshirani 1996), *elasticnet* (Zou and Hastie 2005) and *adaptive lasso* (Zou 2006) to determine significant inputs in the linear model. The *elasticnet* and *adaptive lasso* methods are improvements over the *lasso* feature selection method accounting for correlated features and possessing oracle properties respectively (Hastie et al. 2009). For each of these methods, the problem of feature selection is represented in the regularization modeling framework as an optimization problem, as $\underset{\beta}{\operatorname{argmin}} f_{Loss}(\beta) + f_{Penalty}(\beta)$. The loss function for linear regression is the sum of squared errors given by $\sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{p=1}^{P} x_{ip}\,\beta_p\right)^2$ and the penalty function for regularized models is given in Equation

$$f_{Lasso-Penalty}(\beta) = \lambda \sum_{p=1}^{P} |\beta_p| \tag{15}$$

$$f_{Elasticnet-Penalty}(\beta) = \lambda[(1-\alpha)\sum_{p=1}^{P}|\beta_p|^2 /2 + \alpha \sum_{p=1}^{P} |\beta_p|] \tag{16}$$

$$f_{Adaptive\ lasso-Penalty}(\beta) = \lambda \sum_{p=1}^{P} w_p |\beta_P| \tag{17}$$

The hyperparameter λ and α are determined using grid-search procedure (Hastie et al. 2009). The initial adaptive weights are set as $\frac{1}{|\beta_p^{OLS}|}$, or inversely proportional to the absolute values of naïve regression coefficients of inputs as proposed by Zou (2006). We choose the person-level variables that have non-zero coefficients in all three regularized models as the final set of factors contributing to individual heterogeneity effects.

Table 2 summarizes existing and proposed methods for addressing the three challenges in sound-wellbeing modeling.

Table 2: Methods addressing challenges in sound-wellbeing modeling

Existing/Proposed methods	**Modeling challenges**		
	Modeling curvilinear effects (using segmented models)	**Simultaneous modeling of multiple outcomes**	**Modeling individual heterogeneity effects**
Existing methods	• Heuristic approach (Kraus et al. 2013) • Maximum likelihood based method (Muggeo et al. 2014)	• Classical approach – Univariate method (Baldwin et al. 2014)	• Slopes-as-outcomes modeling method (Raudenbush and Bryk 2002)
Proposed methods	• Semi-automated change point determination method	• Bayesian latent variable modeling method	• Varying-coefficients modeling method

6. Analysis

6.1 Data pre-processing

A total of 248 office workers expressed interest in participating in our study (described in section 3), representing approximately 12% of the workers located in areas of the office buildings where recruitment took place. Pregnant women and those wearing pacemakers or insulin pumps were excluded. Participants taking medication known to affect cardiac activity were noted but not excluded. Due to scheduling problems, sickness and exclusionary criteria, 17 office workers did not participate, resulting in a total enrolment of 231 participants. Due to unexpected changes in work schedules, 8 of the 231 participants were only observed for two workdays, rather than the full 3 days.

The heart rate variability measures SDNN and normalized-HF were calculated using cardiac activity measured by EcgMove3, according to the guidelines of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology (Malik et al. 1996). Physical activity levels were assessed in mG from the EcgMove3's triaxial accelerometer sensor (Razjouyan et al. 2018). Sound levels were aggregated at 5-minute intervals to be integrated with physiological wellbeing measures SDNN and normalized-HF, assuming no lagged effects (Srinivasan et al. 2017). Only observations with both outcome values present were considered in the analysis. Observations with outcome values above the 99.5th percentile were discarded. Age and BMI were discretized to five and four levels respectively for ease of interpretation. Data of participants with less than one hour of recorded data were excluded from analysis. Missing values in input variables were imputed using mean values. Our final dataset contained 31,557 observations aggregated at five-minute intervals and processing approximately 200,000 minutes of wearable data streams from the 231 participants. Apart from sound level as the input variable and SDNN and normalized-HF as the outcomes, person-level variables (e.g., age, gender, etc.), temporal indicators (time of day, day of the week) and physical activity levels were included as covariates in the statistical models. We used the entire dataset training the models for making inference. Observations from day 1 and day 2 of participation of all participants were considered as the training dataset, and day 3 observations were used as the holdout sample (i.e., test dataset) for evaluating the predictive performance of models. Summary statistics of relevant intrapersonal variables (i.e., wearable device based repeated measures and temporal information) and interpersonal variables (i.e., person-level information) in this study are shown in Table 3.

Table 3: Summary statistics of our data

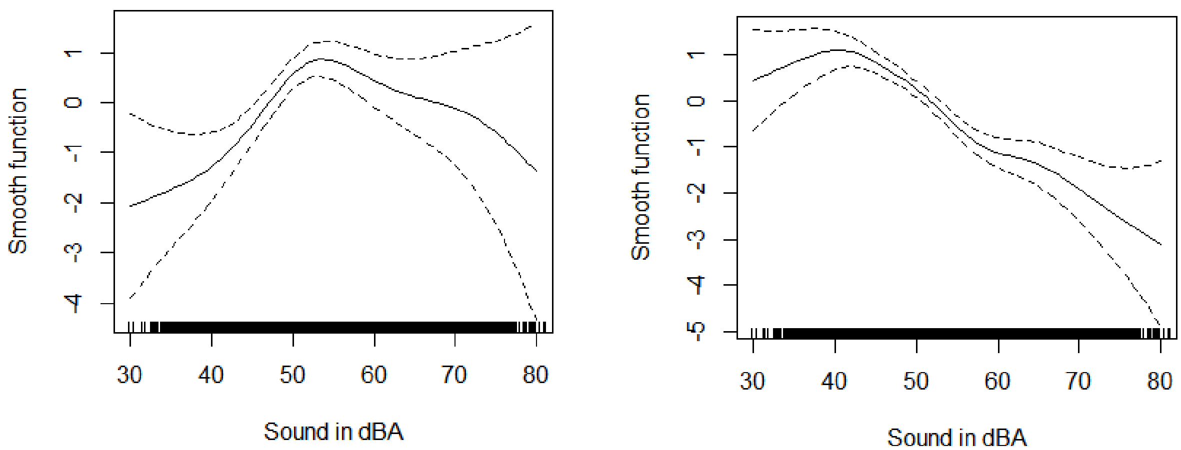
Variable	**Summary**			
INTRAPERSONAL				
Numerical	**Mean**	**SD**	**Units**	**% missing**
SDNN	53.08	23.33	ms	-
Normalized-HF	19.81	12.70	%	-
Sound level	51.85	8.79	dBA	4.29
Physical activity level	0.1738	0.3164	G	0.07
Categorical	**Category**	**Hours:Mins**	**Proportion**	**% missing**
Time of Day				0
	Morning	1224:10	45.76	
	Afternoon	1039:30	38.85	
	Evening	411:15	15.37	
Day of week				0
	Monday	449:25	16.80	
	Tuesday	860:50	32.18	
	Wednesday	916:55	34.28	
	Thursday	431:50	16.14	
	Friday	15:45	0.59	
INTERPERSONAL				
Numerical	**Mean**	**SD**	**Units**	**% missing**
Neuroticism	3.21	0.97	Scale 1-7	10.38
Noise sensitivity	4.05	1.17	Scale 1-7	9.52
Average Sound exposure	51.99	4.89	dBA	4.33
Categorical	**Category**	**No. of participants**	**Proportion**	**% missing**
Age				9.95
	Less than 30 years	30	12.98	
	30 - 39 years	62	26.83	
	40 - 49 years	43	18.61	
	50 - 59 years	56	24.24	

	60 years or above	17	7.36	
Gender				12.12
	Male	88	38.09	
	Female	115	49.78	
BMI				10.39
	18.5 - 25	76	32.9	
	25.1 - 30	81	35.06	
	30.1 - 35	30	12.98	
	Above 35.1	20	8.66	
Computer-dominant work				8.66
	Yes	93	40.26	
	No	118	51.08	
Management work				8.66
	Yes	69	29.87	
	No	142	61.47	
Technical work				8.66
	Yes	90	38.96	
	No	121	52.38	
Meeting heavy work				8.66
	Yes	42	18.18	
	No	169	73.16	
Sleep problems				9.09
	Yes	42	18.18	
	No	168	72.73	
High blood pressure				9.09
	Yes	42	18.18	
	No	168	72.73	
Anxiety				9.09
	Yes	38	16.45	
	No	172	74.46	

6.2 Modeling curvilinear effects of sound levels on physiological wellbeing

We applied and evaluated our semi-automated method to determine the change points for fitting segmented multilevel models associating sound level with SDNN and normalized-HF as physiological wellbeing outcomes.

The first step of the semi-automated method is to fit a Generalized additive mixed model (GAMM) and visualize its sound level component smooth function. The component smooth functions of sound level on the outcomes, SDNN and normalized-HF are shown in Figure 2(a) and Figure 2(b) respectively. Smooth functions in both models were observed to be curvilinear with a single maximum across the range of sound level in the dataset.



(a) (b)

Figure 2. Component smooth function of sound level in GAMM for (a) SDNN as outcome and (b) Normalized-HF as outcome

The second step of the semi-automated method is to perform a linear search for change points using optimization with box constraints. For models corresponding to outcomes SDNN and normalized-HF, we chose starting values of 55 dBA and 45 dBA across search ranges [40, 60] and [30, 50] respectively for running linear search of change points. Brent's optimization algorithm with box constraints (Brent 2013) was used as the linear search method. We identified 51 dBA and 39 dBA as change points for two models with SDNN and normalized-HF as outcomes respectively. Finally, we fitted segmented multilevel regression models using these change points over the training data. Fixed-effects coefficients of sound level segments in models with outcomes SDNN and normalized-HF are as shown in Table 4.

Table 4. Fixed-effects coefficients of sound level in segmented multilevel models

Outcome	Segment	Coefficient (SE)
SDNN	Sound level < 51 dBA	0.1425 (0.06)**
	Sound level ≥ 51 dBA	-0.0682 (0.03) **
Normalized-HF	Sound level < 39 dBA	Insignificant
	Sound level ≥ 39 dBA	-0.0998 (0.02)***

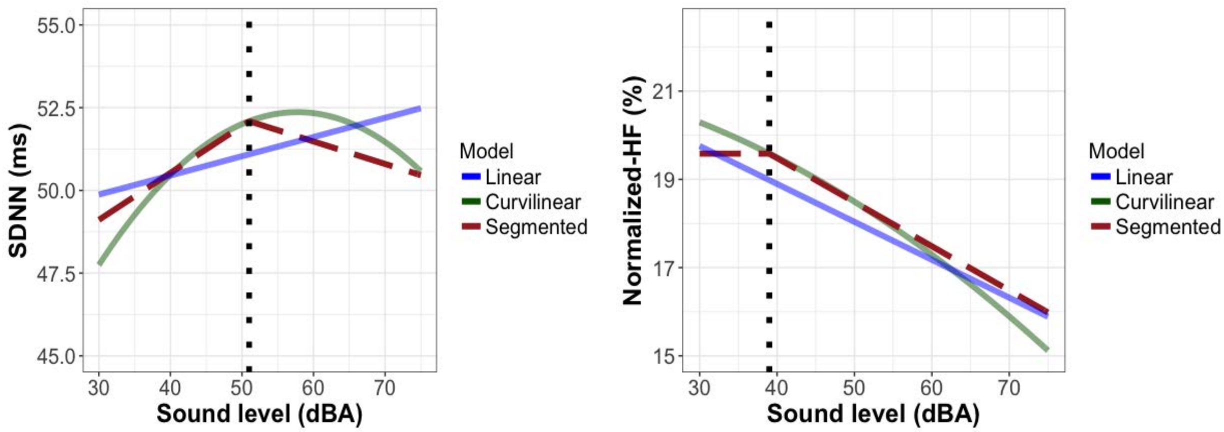
** = p < .05, *** = p < .01

We tested the robustness of the change point estimates by varying search ranges and starting points around the maxima identified in previous step and got estimates within +/- 1 dBA tolerance of the previous estimates. Further, we found that other sophisticated optimization algorithms such as BFGS and L-BFGS (Fletcher 2013) gave similar estimates for change points. We compared the performance of the segmented multilevel model with change points determined using our proposed method with the performance of multilevel models with: (a) Sound level as linear input (b) Sound level as curvilinear input (i.e., first order and second order effects) (c) Sound level segmented using the maximum likelihood approach (Muggeo et al. 2014), and (d) Sound level segmented using ad-hoc approach (Kraus et al. 2013). Using the maximum likelihood method (Muggeo et al. 2014), the change points determined for sound level were 53 dBA and 76 dBA for models with SDNN and normalized HF as outcomes respectively. For the ad-hoc method, we inspect the component smooth curves of the GAMM models and set the change points as 55 dBA and 45 dBA for models with SDNN and normalized HF as outcomes respectively. The fixed effects model was used as a baseline, denoting the performance when only fixed effects of sound level are considered in the multilevel model. Sound level was included in fixed as well as random effects components of other models. Model fit was checked using pseudo R-Squared (Nakagawa and Schielzeth 2013). Predictive performance was compared using Root Mean Squared error (RMSE) and Mean Absolute Prediction Error (MAPE) on the test dataset. The model fit and prediction accuracy comparisons across models described in previous section are shown in Table 5. The model fit and error estimates for best performing models are highlighted for reader convenience. Better model fit, and predictive performance corresponds to a higher value of R-squared and lower error values. The models with segmented inputs perform better than models with linear inputs (Table 5), but equivalent to models with curvilinear inputs in terms of fit and predictive performance. Table 5 shows that segmented models with change points determined using our method are better than segmented models with change points using inspection alone (ad-hoc method) or using the maximum likelihood method.

Table 5. Model fit and predictive performance comparison of segmented multilevel models

Model	**SDNN**			**Normalized-HF**		
	R-sq.	**RMSE (ms)**	**MAPE (%)**	**R-sq.**	**RMSE (ms)**	**MAPE (%)**
Fixed effects only (baseline)	0.5098	17.84	26.19	0.5026	9.17	46.12
Linear inputs	0.5555	17.44	25.18	0.5202	9.08	44.71
Curvilinear inputs	0.5815	17.21	24.73	**0.5329**	8.97	**44.17**
Segmented inputs using ad-hoc method (Kraus et al. 2013)	0.5832	17.21	24.71	0.5316	8.97	44.19
Segmented inputs using Maximum Likelihood method (Muggeo et al. 2014)	0.5837	**17.20**	24.71	0.5319	8.98	44.18
Segmented inputs using our semi-automated method	**0.5838**	**17.20**	**24.69**	0.5323	**8.96**	44.18

The fixed effects coefficient of sound level in the linear, curvilinear and segmented model are visually represented in Figure 3. Figure 3 shows that segmented models represent the curvilinear relationship better than a linear model and are easier to interpret than the curvilinear model in terms of unit change in outcome as a function of unit change in the sound level.



(a) (b)

Figure 3. Trajectory of linear, curvilinear and segmented fixed-effects coefficients for (a) SDNN as outcome, and (b) Normalized-HF as outcome

6.3 Simultaneous modeling the effects of sound level on SDNN and normalized-HF

We evaluated and applied the Bayesian latent variable modeling method for simultaneous modeling effects of sound level on SDNN and normalized-HF. In the model, fixed-effects were introduced for variables sound level, physical activity level, time of day, day of week, age group, BMI group and gender, and random-effects were introduced for variables sound level and physical activity level. We repeated the change point determination procedure for segmented multilevel models to get a single optimal sound level for outcomes SDNN and normalized-HF. The component smooth function for the GAMM model for Sound level is shown in Figure 4. We selected starting values of 45 dBA and search range [40, 60] for running linear search of change point and identified 50 dBA as an optimal change point for the segmented regression dual-outcome model. We used the classical univariate transformation method for simultaneous modeling of outcomes in the change point determination procedure since the linear search procedure with Bayesian variable modeling method took more than a day to converge.

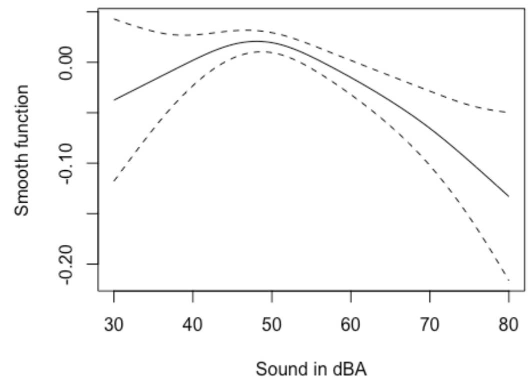


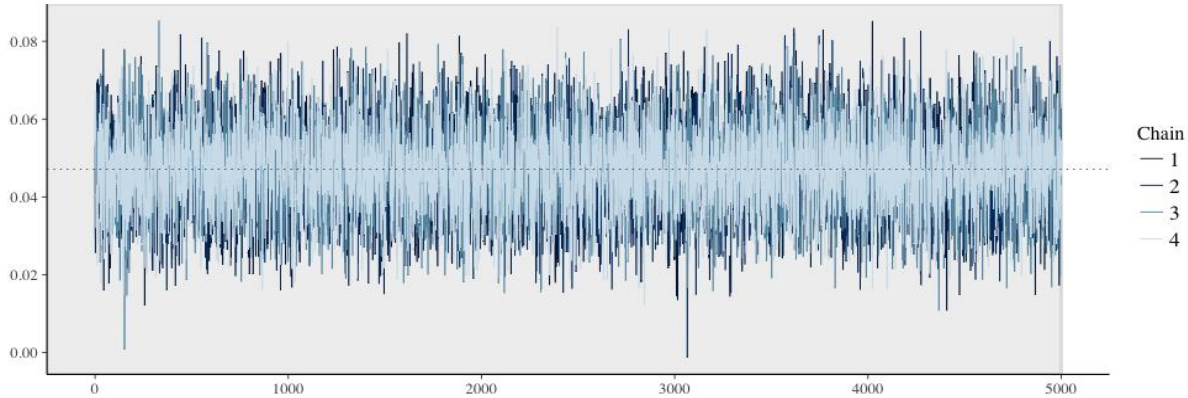
Figure 4. Component smooth function of sound level in GAMM for physiological wellbeing as a bivariate function of SDNN and Normalized-HF

We standardized the input (Sound level) as well as the outcomes (SDNN and normalized-HF) to remove sensitivity and challenges in posterior estimation convergence due to scale differences in the units. For models fit using the Bayesian approach, parameters were assigned a diffused Normal prior and the error variances were assigned a diffused half-Cauchy prior (Gelman and Hill 2007). The Hamiltonian Monte Carlo algorithm was used for sampling four parallel chains (Carpenter et al. 2017). The R-hat statistic cutoff < 1.1 and zero divergence check were used as validation tests for posterior estimates of parameters and assessing quality of fit (Carpenter et al. 2017).

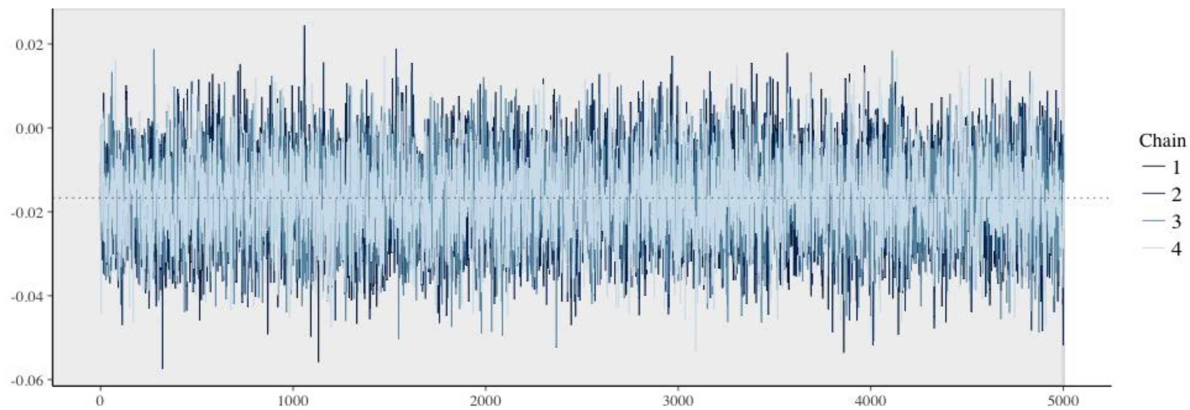
The mean posterior distribution estimates and the 90% credible intervals (values between 5th and 95th percentile of the posterior distribution) of the fixed effects coefficients for models fitted using the Bayesian latent variable modeling method is given in Table 6. The posterior estimates of the fixed-effects of Sound level, Time of day, Day of week, Physical activity level, Age and BMI indicate that they are interpersonal and intrapersonal factors related to an individual's physiological wellbeing at workplace. The trace plots of four chains of MCMC draws for the coefficient of Sound level for the two conditions, Sound level < 50 dBA and Sound level >= 50 dBA are shown in Figure 5, indicating good convergence. Posterior estimates of all the parameters in the model (fixed-effects as well as varying effects) had R-hat values less than 1.1, and the model convergence report indicated zero divergence check, indicating an acceptable model fit.

Table 6 – Fixed effects of models using different simultaneous outcomes modeling methods

Coefficients	Posterior estimate (mean)	90% Credible interval
Sound level$_{Normalized}$ (< 50 dBA)	0.0471	(0.0199 – 0.0648)
Sound level$_{Normalized}$ (>= 50 dBA)	-0.0167	(-0.0337 – -0.0042)
Physical activity level$_{Normalized}$	0.2756	(0.2316 – 0.2932)
Time of day – Morning	*Baseline*	
Time of day – Afternoon	-0.1479	(-0.1675 – -0.1277)
Time of day – Evening	-0.0939	(-0.1206 – -0.0690)
Day of week – Monday	*Baseline*	
Day of week – Tuesday	-0.1301	(-0.2670 – 0.0092)
Day of week – Wednesday	-0.0571	(-0.0888 – -0.0108)
Day of week – Thursday	-0.0588	(-0.0886 – -0.0287)
Day of week –Friday	-0.0430	(-0.0836 – 0.0277)
Age group – Below 30	*Baseline*	
Age group – 30-40	0.1361	(-0.1439 – 0.4115)
Age group – 40-50	-0.1468	(-0.4495 – 0.1641)
Age group – 50-60	-0.3119	(-0.6235 – -0.0038)
Age group – Above 60	-0.4413	(-0.7475 – -0.0132)
BMI group – Below 25	*Baseline*	
BMI group – 25-30	-0.2278	(-0.4281 – -0.0165)
BMI group – 30-35	-0.3619	(-0.6751 – -0.0896)
BMI group – Above 35	-0.6169	(-0.9768 – -0.2363)
Gender – Male	*Baseline*	
Gender – Female	-0.0439	(-0.2278 – 0.1435)



(a)



(b)

Figure 5: MCMC trace plots for coefficient of sound levels in following Bayesian latent variable model for: (a) Sound levels < 50 dBA, and (b) Sound levels >= 50 dBA.

The fixed effect of sound level in the Bayesian latent variable model represents the global effects of sound on individual wellbeing after accounting for individual heterogeneity effects through the varying effects coefficients. The coefficient for sound level in Table 6 indicates change in physiological wellbeing by a standard deviation (SD) related to a unit standard deviation (SD) change in sound level as both input and outcomes are standardized. Knowing that 68.3% of variability is explained by 1 SD of a normal distribution and the SD of Sound level in the dataset is 8.79 dBA (refer to Table 3), we can make the following inferences. For sound amplitudes lower than 50 dBA, a 10 dBA increase in sound level is related to a 3.6% increase in physiological wellbeing. For sound amplitudes higher than 50 dBA, a 10 dBA increase in sound level is related to decrease in physiological wellbeing by 1.3%.

We compared the predictive performance of the Bayesian latent variable modeling method with the following three alternative methods for simultaneous modeling of multiple outcomes: (a) the classical univariate transformation method (Baldwin et al. 2014), (b) the univariate transformation method trained using a Bayesian approach, and (c) the classical multilevel structural equation modeling method (Kline 2011). Models using the classical approach are trained using the R packages lavaan (Rosseel 2012), nlme (Pinheiro et al. 2007) in a 16 GB RAM, 2.7 GHz processor PC, whereas models using the Bayesian approach are written and executed using Stan program through the RStan interface (Carpenter et al. 2017), in a high performance computer cluster with 28 nodes (192 GB RAM per node, Intel Haswell V3 28 core processors). The RMSE and MAPE of the models trained using the four methods are given in Table 7.

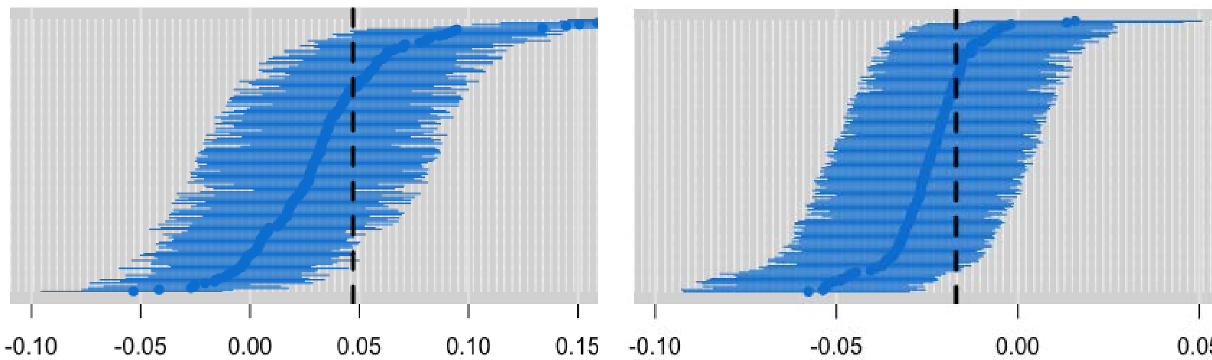
Table 7: Comparing predictive performance of methods for simultaneous modeling of multiple outcomes

Model		**SDNN**		**Normalized HF**	
		RMSE	**MAPE**	**RMSE**	**MAPE**
Classical	*Univariate*	20.12	34.13	10.98	54.36
	Latent	23.71	44.78	11.22	57.10
Bayesian	*Univariate*	21.50	37.39	10.04	52.64
	Latent	**17.06**	**26.56**	**8.90**	**44.36**

Table 7 shows that the model trained using the Bayesian latent variable modeling method has the lowest prediction errors RMSE and MAPE, indicating that our method is superior to other methods for simultaneous modeling of multiple outcomes.

6.4 Individual heterogeneity in effects of sound level on physiological wellbeing

The heterogeneity in the effect of sound level on physiological wellbeing across individuals is accounted for by the varying coefficients of sound level input in the Bayesian latent variable model. Figure 6 shows a caterpillar plot visualization of posterior estimates of varying coefficients of sound level and their 60% credible interval (values between 20th percentile and 80th percentile of the posterior distribution) in the Bayesian latent variable model. The spread of mean values of posterior estimates of the varying coefficients indicate substantial individual heterogeneity effects. We applied the varying-coefficients method to identify person-level variables contributing to individual heterogeneity in sound level effects on physiological wellbeing.



(a) (b)

Figure 6: Caterpillar plots of posterior estimates of varying coefficients of sound level and their 60% credible interval in the Bayesian latent variable model for (a) Sound level < 50 dBA, and (b) Sound level >= 50 dBA. The vertical lines show the corresponding fixed effects coefficients of sound level.

We considered two subsets of the data, one with sound levels less than 50 dBA and other with sound levels greater than or equal to 50 dBA to fit two independent models. By fitting two independent models for instances with high sound levels (>= 50dbA) and instances with low sound levels (< 50 dBA), we were able to make independent inferences about individual heterogeneity effects for each scenario.

The varying-coefficients modeling method uses a two-step procedure for modeling heterogeneity effects and for identifying factors contributing to the heterogeneity effects. We fitted a Bayesian hierarchical model with input variables Sound level, Physical activity level, Time of day, and Day of week as variables with varying coefficients with normal prior having non-zero means. In step 2, we used lasso, elasticnet and adaptive-lasso regularized models to identify person-level variables that have a significant relation with the varying coefficients of Sound level. The coefficients for the regularized feature selection models are shown in Table 8.

Table 8: Coefficients of person-level input variables in regularized models in varying coefficients modeling method

Predictors	Below 50 dBA			Above 50 dBA		
	Lasso	**Elastic-net**	**Adaptive lasso**	**Lasso**	**Elastic-net**	**Adaptive lasso**
Neuroticism						
Noise sensitivity						
Age group - Below 30	*baseline*	*baseline*	*baseline*	*baseline*	*baseline*	*baseline*
Age group - 30-40				-0.0011	-0.0076	-0.0002
Age group - 40-50						
Age group - 50-60	-0.0026	-0.0141	-0.0003			
Age group - Above 60	-0.0047	-0.0224	-0.0007	0.0084	0.0160	0.0010
BMI group - Below 25	*baseline*	*Baseline*	*baseline*	*baseline*	*baseline*	*baseline*
BMI group - 25-30	-0.0009	-0.0044	-0.0001		0.0004	
BMI group - 30-35				-0.0001	-0.0096	
BMI group - Above 35				0.0076	0.0123	0.0011
HighBP - Yes	-0.0133	-0.0764	-0.0021	-0.0207	-0.0203	-0.0042
Anxiety - Yes	-0.0015	-0.0013	-0.0002	0.0060	0.0148	0.0007
Sleep problems - Yes						
Computer use intensive work - Yes	0.0187	0.0881	0.0036			
Managerial work - Yes						
Meeting intensive work – Yes						
Technical work - Yes						
Average Sound exposure						

Table 8 shows that Age groups, BMI groups, High BP, Anxiety, Computer use intensive worktype are the person level factors that are related to the variability in coefficients of sound level in the physiological wellbeing models. The blank cells show that coefficients of corresponding variables have been shrunk to zero in the corresponding feature selection method (i.e., lasso, adaptive lasso, elastic net). To evaluate the performance of the varying coefficient modeling method, we compared the predictive performance of multilevel models with three set of input variables: (a) inputs including no person-level variables as moderators, (b) inputs including all person-level variables as moderators, and (c) inputs including person-level variables identified by varying-coefficients modeling method as moderators. Moderators were included as two-way interactions with fixed effect of sound level in the multilevel models. Table 9 shows the prediction errors of all three models with respect to outcomes SDNN and normalized HF (Univariate outcomes were reconstructed using latent factor estimated in Bayesian latent variable modeling method). Table 9 shows that the model including the person-level variables identified using the varying-coefficients modeling method as moderators have least RMSE and MAPE as compared to other models.

Table 9: Performance comparison of multilevel models with different set of moderators

Moderators of sound level in multilevel model	**SDNN**		**Normalized HF**	
	RMSE	**MAPE**	**RMSE**	**MAPE**
No moderators	17.06	26.56	8.90	44.36
All person-level variables	19.66	31.38	11.13	47.73
Person-level variables identified by varying-coefficients modeling method	**16.65**	**24.97**	**8.41**	**43.23**

High BP and Computer use intensive worktype are person level factors that contribute most to the between-individual heterogeneity in sound level effects on physiological wellbeing (see Table 8). Figure 7(a) and Figure 7(b) are plots showing the change in outcome due to introducing interaction effects of High BP and Computer use intensive worktype variables with sound level fixed effects in multilevel model respectively. The fixed-effect coefficient of Sound level in multilevel models with stratified datasets participants belonging to categories Normal BP, High BP, Computer use intensive worktype, and Not computer use intensive worktype are given in Table 10.

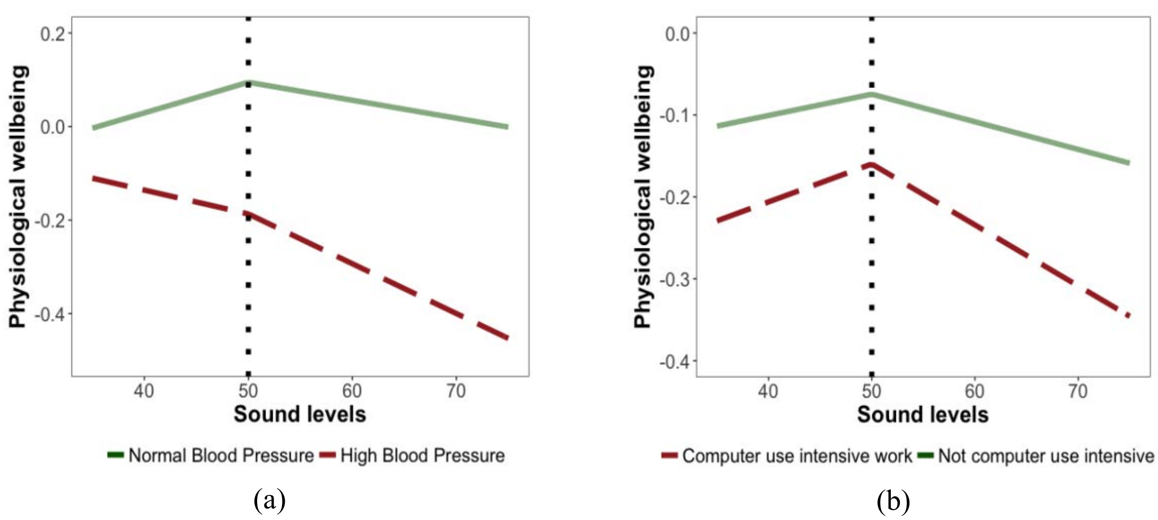


Figure 7: Interaction plots of the top two person-level variables moderating the sound-wellbeing

relationship

Figure 7(a) and Table 10 show that office-workers with blood pressure are more negatively affected than participants with normal blood pressure. Figure 7(b) and Table 10 show that office-workers involved in computer intensive work have higher positive effects of sound levels on physiological wellbeing at amplitudes less than 50 dBA, but have higher negative effects of sound levels on physiological wellbeing at amplitudes more than 50 dBA, as compared to office-workers involved in work that is not computer use intensive.

Table 10: Coefficients of sound level in stratified datasets

Stratification		**Standardized coefficients of Sound level**	
		Less than 50 dBA	**Greater than 50 dBA**
Blood pressure	Normal BP	0.0232	-0.0239
	High BP	-0.0181	-0.0665
Type of work	Not computer use intensive	0.0092	-0.0211
	Computer use intensive	0.0165	-0.0461

6.5 Empirical evidence

In order to further validate the presence of optimal Sound level for physiological wellbeing at 50 dBA and the influence of blood pressure and work involving intensive computer use in moderating the sound-wellbeing relationship, we conduct post-hoc comparison of wellbeing across different stratified populations for three sound level conditions (i.e., Sound level is lesser than 45 dBA, Sound level is between 45 dBA and 55 dBA and Sound level is greater than 55 dBA). Table 11 shows the post-hoc comparisons of mean wellbeing score adjusted for random effects for three sound level ranges for different sub-populations in our data. In support of our finding that 50 dBA is an optimal sound level at workplace, we find that Sound level range 45-55 dBA has the highest mean adjusted wellbeing score across the complete population, when compared to low and high sound level ranges. But for individuals with High blood pressure, the lowest sound level range (i.e., Sound level <= 45 dBA) is optimal, unlike individuals with normal blood pressure. Finally, individuals with computer use intensive work have a lower mean adjusted wellbeing score for low as well as high Sound level ranges (i.e., Sound level <=45 dBA and Sound level > 55 dBA), when compared to individuals with regular computer use at work. These empirical findings are concurrent with person-level moderator pattern identified and characterized using our varying co-efficient modeling method for modeling individual heterogeneity effects.

Table 11: Post-hoc group comparisons across sound level ranges

Sub-population	Mean adjusted wellbeing score††		
	Sound <= 45 dBA	45 dBA < Sound <= 55 dBA	Sound > 55 dBA
Complete dataset	0.0054	0.0174	-0.0203
High BP†	-0.1395	-0.1600	-0.1884
Normal BP	0.0120	0.0302	-0.0081
Intensive computer use	-0.0407	-0.0186	-0.0985
Regular computer use	0.0229	0.0333	0.0144

† Repeated measures MANOVA shows significant differences across sound level ranges for each sub-population except high BP
†† Mean value of wellbeing is adjusted for random effects using estimated marginal means procedure (Searle 1980)

7. Discussion

In this study, we introduced a novel study design using wearable devices and proposed three new statistical modeling methods to determine the effects of workplace sound level on an individual's physiological wellbeing. Traditional study methodologies such as controlled experiments and survey-based methods are not suitable for sound-wellbeing modeling as the real office environment ecosystem is more complex than simulated environments with a limited set of treatments and controls. We overcome the above challenge of low external validity by using wearable sensors to collect continuous measurements of environmental and physiological conditions of multiple persons simultaneously. We conducted a preliminary analysis of the collected data and discovered that sound level has a curvilinear effect on two physiological wellbeing measures (i.e., SDNN and Normalized-HF), and this effect varies across individuals. We proposed three new methods for addressing the gap in statistical modeling methods for representing curvilinear effects, simultaneous modeling of multiple outcomes, and identifying factors contributing to individual heterogeneity in order to model the sound-wellbeing relationship.

Our methods can be naturally extended to other similar studies employing multiple wearable sensors and multiple participants. We propose the following guidelines for applying the statistical modeling methods to future applications:

Modeling curvilinear relationships: We have focused on the problem of identification of change points for input variables to be segmented, assuming that a curvilinear relationship has already been ascertained in advance. Curvilinear relationships can be based on prior literature or hypothesized through theoretical deductions. It can be explored using scatter plot visualizations and validated using second-order coefficients in models. If curvilinear effects are absent, the segmented modeling approach should be avoided to prevent over-fitting. The proposed semi-automated method for change point determination in segmented modeling of the curvilinear sound-wellbeing relationship requires manual inspection and supply of initial values to optimization algorithms. In future applications with multiple inputs and multiple distinctive extrema observed in GAMM smooth functions, a linear search procedure for each change point can be tedious. A few ways to avoid this problem are – making variable transformations, treating extreme values and discretizing inputs.

Simultaneous modeling of multiple outcomes: The Bayesian latent variable modeling method is suitable for simultaneous modeling of outcomes belonging to the same family of distributions but cannot handle outcomes from different families of distributions in its current formulation. For example, our method cannot a combination of discrete and continuous outcomes. An interim solution to this problem will be to carry out transformations of outcomes (e.g., Box-Cox, log-linear, quantization, etc.) to convert them into a singular family of distributions. Future studies can extend the Bayesian latent variable modeling method to implicitly handle mixed-type of outcomes.

Modeling individual heterogeneity effects: The varying-coefficient modeling method identifies person level factors contributing to between individual heterogeneity in effects, given that there is sufficient person-level information available in data. Therefore, one should collect as much person-level information as possible, since our method works better with more variables available for feature selection step using the three regularized models. Secondly, for the varying-coefficients model in our application, we have assumed that SDNN and normalized-HF can be represented using a single latent variable. In future applications, there can be theoretical justification to model multiple outcomes simultaneously through two or more latent variable constructs. Future studies can extend our method using parameter expansion (Merkle and Wang 2016) or other suitable approaches for conceptualizing multiple latent variables in the hierarchical Bayesian modeling paradigm.

We have developed statistical modeling methods by assuming a two-level structure, with variables varying at within-individual (level-1) and for variables varying at individual level (level-2) in the data. But our methods can be applied to data with more than two levels of grouping structures without loss of generality. Ours is the first study that attempts to model individual heterogeneity effects and therefore we need future research to validate its application in other problem domains with individual heterogeneity effects. Even though curvilinear effects, multiple outcomes and individual heterogeneity effects have been presented in the context of sound-wellbeing modeling, they are modeling challenges encountered in a wider set of applications employing multilevel models. Our proposed methods can be used for applications such as patient monitoring systems, military fitness management programs, smart diet applications, etc. which have multilevel streaming data.

Relevance to IS:

IT artifacts are broadly defined as constructs (vocabulary and symbols), models (abstractions and representations), methods (algorithms and practices), and instantiations (implemented and prototype systems) (Hevner et al. 2004). In this study, we introduce the problem of determining the effects of workplace sound level on wellbeing of physiological wellbeing of office workers and introduce three new methods for modeling the sound-wellbeing relationship. Our major contributions include introduction of a unique study design using wearable devices for analyzing a complex relationship, new quantitative methods for modeling digital data generated by wearables and informing workplace planning policies and practices that affect the health and wellbeing of office workers worldwide.

Predictive modeling and statistical modeling in analytics go side-by-side as one predicts the future using existing data, focusing on informing us on the question “What will be”, while the other explicates hidden patterns and tells us about “What is” with respect to a phenomenon. Both of them are important for creating value out of data generated from digital sources such as wearable devices. As the number of wearable technology-based applications increases in future, the quantum of available data to analyze will exponentially increase and warrant for more and more advancements in statistical modeling for meaningful interpretations of patterns. Our method contributions in statistical modeling of wearables generated data are timely in IS research, as the discipline is widening its scope in design science using novel data sources including wearable devices. Wearable data analytics is a promising area that solicits attention from IS researchers owing to the ubiquitous nature of wearables in today’s lifestyle, and the promise of wearables to generate rich, personalized, temporal and highly-grained information content. In future, IS research employing wearable technology may encounter other challenges related to cost of sensors, ethical constraints on human research, participant privacy concerns, etc.

Study limitations:

Our study has a few limitations. In this study, we have focused on modeling the effects of workplace sound level on physiological wellbeing of office workers, but we have not collected information about the sound types (e.g., conversation, mechanical background noise, etc.) and frequencies (e.g., low, speech, high-tone, etc.) due to individual privacy concerns and sensor technology limitations. Since sound type and sound frequency do not moderate the effects of sound level on physiological wellbeing outcomes (Sun Sim et al. 2015; Walker et al. 2016), we believe our findings will still hold when controlling for the type and frequency of ambient sounds. Future studies can focus on effects of sound type and sound frequency and use our study design and modeling methods. Secondly, we have aggregated the sound level and other within individual variables at 5-minute intervals to match the grain of short-term physiological wellbeing measures SDNN and normalized-HF; since the latter cannot be determined meaningfully for grains finer than 5 minutes. Therefore, the lasting effects of spikes in sound level due to sudden events (e.g., shrieking sound, breaking glass, etc.) have not been investigated. However, effects of events repeated multiple times as well as background noises consistent across the 5-minute interval are accounted for, in our models. The Bayesian approach for simultaneous modeling of outcomes and modeling heterogeneity effects leads to better performance when compared to the classical (i.e., frequentist) approaches, at the cost of computing power and time. Posterior estimation of parameters in Bayesian models take much more time than estimation of coefficients in classical models. This limitation can be partially overcome by using parallel processing and high-performance computing clusters to estimate parameters of the Bayesian models.

8. Conclusion

The majority of the working population in today's world spend a significant part of their active hours in closed office workspaces. Research shows that not only our work and social interactions at workplace, but also the workplace environment has an impact on our social, psychological, and physiological wellbeing. Among all the indoor environment quality factors, the ambient sound level is reported to be one of the highest stressors at the workplace in US. We conducted a field study using wearable devices where participants carried out their day-to-day activities while wearing sensors that continuously recorded their physiological wellbeing state and ambient sound level. We fitted multilevel regression models to the resultant data and observed evidences of a relationship between sound level and two physiological wellbeing measures (i.e., SDNN, normalized-HF). To better understand the mechanism of the sound-wellbeing relationship, we have proposed three new statistical modeling methods for representing curvilinear effects, simultaneous modeling of multiple outcomes, and identifying factors contributing to individual heterogeneity effects. Our methods have better predictive performance than existing methods for each of the three modeling problems. Using our methods, we infer that an individual's physiological wellbeing is optimal when sound level in the workplace is 50 dBA. For sound amplitudes lower than 50 dBA, a 10 dBA increase in sound level is related to a 3.6% increase in physiological wellbeing; whereas for amplitude above 50 dBA, a 10 dBA increase in sound level is related to decrease in physiological wellbeing by 1.3%. Age, body-mass-index, high blood pressure, anxiety and computer use intensive work are person level factors contributing to heterogeneity in sound level effects on physiological wellbeing across our study population. Our modeling method shows that workers with higher blood pressure are more negatively affected by increase in sound levels while workers with computer intensive work are more negatively affected by sound level extremities (i.e., quietude and loud noise). Our study informs policies and practices for designing workplaces with optimal sound levels in general and customized for certain subsets of the office worker population. It contributes to statistical modeling in wearable data analytics to facilitate the advancement of IS in BI&A 3.0 through meaningful use of sensor-based content.

Moving forward, we plan to extend our analysis to determine the effects of other indoor environment quality factors including temperature, CO_2, relative humidity and light intensity on physiological wellbeing of office workers. We also plan to validate our findings using additional (causal) experiments for specific target groups and discrete values of sound levels. Finally, we plan to investigate other factors contributing to the heterogeneity in sound level effects on physiological wellbeing such as extent of social interaction interactions between office-workers and average sound level exposure during leisure time.

Funding

This study was funded by a United States General Services Administration contract # GS-00-H-14-AA-C-0094.

Competing Interests

Authors declare no competing interests.

References

Abrams, S., and Hens, N. 2015. "Modeling Individual Heterogeneity in the Acquisition of Recurrent Infections: An Application to Parvovirus B19," *Biostatistics*.

Acharya, U. R., Joseph, K. P., Kannathal, N., Lim, C. M., and Suri, J. S. 2006. "Heart Rate Variability: A Review," *Medical and Biological Engineering and Computing*, pp. 1031–1051.

Backé, E. M., Seidler, A., Latza, U., Rossnagel, K., and Schumann, B. 2012. "The Role of Psychosocial Stress at Work for the Development of Cardiovascular Diseases: A Systematic Review," *International Archives of Occupational and Environmental Health*.

Baldwin, S. A., Imel, Z. E., Braithwaite, S. R., and Atkins, D. C. 2014. "Analyzing Multiple Outcomes in Clinical Research Using Multivariate Multilevel Models," *Journal of Consulting and Clinical Psychology* (82:5), pp. 920–930.

Becker, J.-M., Rai, A., Ringle, C. M., and Völckner, F. 2013. "Discovering Unobserved Heterogeneity in Structural Equation Models to Avert Validity Threats," *MIS Quarterly* (37:3), Society for Information Management and The Management Information Systems Research Center, pp. 665–694.

Boron, W. F., and Boulpaep, E. L. 2012. "Medical Physiology," *Physiology*.

Brent, R. P. 2013. *Algorithms for Minimization without Derivatives*, Courier Corporation.

Bureau of Labor Statistics. 2017. "Average Weekly Hours and Overtime of All Employees on Private Nonfarm Payrolls by Industry Sector, Seasonally Adjusted." (https://www.bls.gov/news.release/pdf/atus.pdf).

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., and Riddell, A. 2017. "Stan: A Probabilistic Programming Language," *Journal of Statistical Software*.

CDC Foundation. 2018. "Business Pulse: Healthy Workforce | CDC Foundation." (https://www.cdcfoundation.org/businesspulse/healthy-workforce, accessed September 11, 2018).

Cvijanović, N., Kechichian, P., Janse, K., and Kohlrausch, A. 2017. "Effects of Noise on Arousal in a

Speech Communication Setting,” *Speech Communication* (88), pp. 127–136.

Das, A., Poole, W. K., and Bada, H. S. 2004. “A Repeated Measures Approach for Simultaneous Modeling of Multiple Neurobehavioral Outcomes in Newborns Exposed to Cocaine in Utero,” *American Journal of Epidemiology*.

Dingemanse, N. J., and Dochtermann, N. A. 2013. “Quantifying Individual Variation in Behaviour: Mixed-Effect Modelling Approaches,” *Journal of Animal Ecology*.

Durban, M., Harezlak, J., Wand, M. P., and Carroll, R. J. 2005. “Simple Fitting of Subject-Specific Curves for Longitudinal Data,” *Statistics in Medicine* (24:8), pp. 1153–1167.

Faraway, J. J. 2006. “Extending the Linear Model With R: Generalized Linear, Mixed Effects and Nonparametric Regression Models,” *Journal of the American Statistical Association*, pp. 1–28.

Faraway, J. J. 2016. “Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models,” *CRC Press*.

Fletcher, R. 2013. “Practical Methods of Optimization,” *John Wiley & Sons* (53), p. 456.

Frontczak, M., Schiavon, S., Goins, J., Arens, E., Zhang, H., and Wargocki, P. 2012. “Quantitative Relationships between Occupant Satisfaction and Satisfaction Aspects of Indoor Environmental Quality and Building Design,” *Indoor Air* (22:2), pp. 119–31.

Gelman, A., and Hill, J. 2007. “Data Analysis Using Regression and Multilevel/Hierarchical Models,” *Cambridge*.

Geng, Y., Ji, W., Lin, B., and Zhu, Y. 2017. “The Impact of Thermal Environment on Occupant IEQ Perception and Productivity,” *Building and Environment*.

Gimenez, O., Cam, E., and Gaillard, J. M. 2018. “Individual Heterogeneity and Capture–Recapture Models: What, Why and How?,” *Oikos* (127), pp. 664–686.

Harvard School of Public Health. 2016. “The Workplace and Health.” (https://news.harvard.edu/wp-content/uploads/2016/07/npr-rwjf-harvard-workplace-and-health-poll-report.pdf).

Hastie, T., Tibshirani, R., and Friedman, J. 2009. “The Elements of Statistical Learning: Data Mining, Inference, and Prediction," *Springer Series in Statistics*, The Mathematical Intelligencer.

Heck, R. H., and Thomas, S. L. 2015. "An Introduction to Multilevel Modeling Techniques: MLM and SEM Approaches Using Mplus," *Hodder Arnold*, Routledge.

Heerwagen, J., and Zagreus, L. 2005. "The Human Factors of Sustainable Building Design: Post Occupancy Evaluation of the Philip Merrill Environmental Center," *Indoor Environmental Quality*.

Hevner, A. R., March, S. T., Park, J., and Ram, S. 2004. "Design Science in Information Systems Research," *MIS Quarterly* (28:1), pp. 75–105.

Hox, J. J. 2013. "Multilevel Regression and Multilevel Structural Equation Modeling," *The Oxford Handbook of Quantitative Methods* (2), pp. 281–294.

Jahncke, H., Hygge, S., Halin, N., Green, A. M., and Dimberg, K. 2011. "Open-Plan Office Noise: Cognitive Performance and Restoration," *Journal of Environmental Psychology* (31:4), pp. 373–382.

Jirschitzka, J., Kimmerle, J., and Cress, U. 2016. "A New Method for Re-Analyzing Evaluation Bias: Piecewise Growth Curve Modeling Reveals an Asymmetry in the Evaluation of pro and Con Arguments," *PLoS ONE* (11:2).

Kivimäki, M., Nyberg, S. T., Batty, G. D., Fransson, E. I., Heikkilä, K., Alfredsson, L., Bjorner, J. B., Borritz, M., Burr, H., Casini, A., Clays, E., De Bacquer, D., Dragano, N., Ferrie, J. E., Geuskens, G. A., Goldberg, M., Hamer, M., Hooftman, W. E., Houtman, I. L., Joensuu, M., Jokela, M., Kittel, F., Knutsson, A., Koskenvuo, M., Koskinen, A., Kouvonen, A., Kumari, M., Madsen, I. E. H., Marmot, M. G., Nielsen, M. L., Nordin, M., Oksanen, T., Pentti, J., Rugulies, R., Salo, P., Siegrist, J., Singh-Manoux, A., Suominen, S. B., Väänänen, A., Vahtera, J., Virtanen, M., Westerholm, P. J. M., Westerlund, H., Zins, M., Steptoe, A., and Theorell, T. 2012. "Job Strain as a Risk Factor for Coronary Heart Disease: A Collaborative Meta-Analysis of Individual Participant Data," *The Lancet*.

Kjellberg, A., Landström, U., Tesarz, M., Söderberg, L., and Åkerlund, E. 1996. "The Effects of Nonphysical Noise Characteristics, Ongoing Task and Noise Sensitivity on Annoyance and Distraction Due to Noise at Work," *Journal of Environmental Psychology* (16:2), pp. 123–136.

Kline, R. B. 2011. “Principles and Practice of Structural Equation Modeling,” *Guilford Publication* (Third Edit., Vol. 156).

Kline, R. B. 2012. “Assumptions in Structural Equation Modeling,” *Handbook of Structural Equation Modeling*.

Kraus, U., Schneider, A., Breitner, S., Hampel, R., Rückerl, R., Pitz, M., Geruschkat, U., Belcredi, P., Radon, K., and Peters, A. 2013. “Individual Daytime Noise Exposure during Routine Activities and Heart Rate Variability in Adults: A Repeated Measures Study,” *Environmental Health Perspectives* (121:5), pp. 607–612.

Landström, U., Åkerlund, E., Kjellberg, A., and Tesarz, M. 1995. “Exposure Levels, Tonal Components, and Noise Annoyance in Working Environments,” *Environment International* (21:3), pp. 265–275.

Lee, G.-S., Chen, M.-L., and Wang, G.-Y. 2010. “Evoked Response of Heart Rate Variability Using Short-Duration White Noise,” *Autonomic Neuroscience : Basic & Clinical*, pp. 94–97.

Lin, Y.-K., Chen, H., Brown, R. A., Li, S.-H., and Yang, H.-J. 2017. “Healthcare Predictive Analytics for Risk Profiling in Chronic Care: A Bayesian Multitask Learning Approach,” *MIS Quarterly* (41:2), pp. 473–495.

Lindberg, C. M., Srinivasan, K., Gilligan, B., Razjouyan, J., Lee, H., Najafi, B., Canada, K. J., Mehl, M. R., Currim, F., Ram, S., Lunden, M. M., Heerwagen, J. H., Kampschroer, K., and Sternberg, E. M. 2018. “Effects of Office Workstation Type on Physical Activity and Stress,” *Occupational and Environmental Medicine* (10:75), pp. 689–695.

Liu, B. Q., and Goodhue, D. L. 2012. “Two Worlds of Trust for Potential E-Commerce Users: Humans as Cognitive Misers,” *Information Systems Research* (23:4), pp. 1246–1262.

Lusk, S. L., Hagerty, B. M., Gillespie, B., and Caruso, C. C. 2002. “Chronic Effects of Workplace Noise on Blood Pressure and Heart Rate,” *Archives of Environmental Health* (57:4), pp. 273–281.

MacNaughton, P., Spengler, J., Vallarino, J., Santanam, S., Satish, U., and Allen, J. 2016. “Environmental Perceptions and Health before and after Relocation to a Green Building,” *Building and Environment*

(104), pp. 138–144.

Malik, M., Bigger, J. T., Camm, A. J., Kleiger, R. E., Malliani, A., Moss, A. J., and Schwartz, P. J. 1996. "Heart Rate Variability: Standards of Measurement, Physiological Interpretation, and Clinical Use," *European Heart Journal*.

Merkle, E. C., and Wang, T. 2016. "Bayesian Latent Variable Models for the Analysis of Experimental Psychology Data," *Psychonomic Bulletin & Review*.

Muggeo, V. M., Atkins, D. C., Gallop, R. J., and Dimidjian, S. 2014. "Segmented Mixed Models with Random Changepoints: A Maximum Likelihood Approach with Application to Treatment for Depression Study," *Statistical Modelling* (14), pp. 293–313.

Muthén, B. O. 2002. "Beyond SEM : General Latent Variable Modelling," *Behaviormetrika*.

Nakagawa, S., and Schielzeth, H. 2013. "A General and Simple Method for Obtaining R2 from Generalized Linear Mixed-Effects Models," *Methods in Ecology and Evolution* (4:2), pp. 133–142.

Pant, G., and Srinivasan, P. 2010. "Predicting Web Page Status," *Information Systems Research* (21:2), pp. 345–364.

Park, S. H., and Lee, P. J. 2017. "Effects of Floor Impact Noise on Psychophysiological Responses," *Building and Environment* (116), Pergamon, pp. 173–181.

Pinheiro, J., Bates, D., DebRoy, S., and Sarkar, D. 2007. "Nlme: Linear and Nonlinear Mixed Effects Models," *R Package Version 3*.

Pituch, K. A., and Stevens, J. P. 2016. "Applied Multivariate Statistics for the Social Sciences," *Routledge*.

Rashid, M., and Zimring, C. 2008. "A Review of the Empirical Literature on the Relationships Between Indoor Environment and Stress in Health Care and Office Settings: Problems and Prospects of Sharing Evidence," *Environment and Behavior* (40:2), pp. 151–190.

Raudenbush, S. W., and Bryk, A. S. 2002. "Hierarchical Linear Models: Applications and Data Analysis Methods," *Advanced Quantitative Techniques in the Social Sciences 1* (Vol. 2nd).

Razjouyan, J., Naik, A. D., Horstman, M. J., Kunik, M. E., Amirmazaheri, M., Zhou, H., Sharafkhaneh, A., and Najafi, B. 2018. "Wearable Sensors and the Assessment of Frailty among Vulnerable Older Adults: An Observational Cohort Study," *Sensors*.

Ritz, C., Pilmann Laursen, R., and Trab Damsgaard, C. 2017. "Simultaneous Inference for Multilevel Linear Mixed Models—with an Application to a Large-Scale School Meal Study," *Journal of the Royal Statistical Society. Series C: Applied Statistics* (66:2), pp. 295–311.

Rosseel, Y. 2012. "Lavaan: An R Package for Structural Equation Modeling," *Journal of Statistical Software* (48:2), pp. 1–36.

Ryff, C. D. 1989. "Happiness Is Everything, or Is It? Explorations on the Meaning of Psychological Well-Being.," *Journal of Personality and Social Psychology*.

Seidman, M. D., and Standring, R. T. 2010. "Noise and Quality of Life," *International Journal of Environmental Research and Public Health* (7:7), pp. 3730–3738.

Shaffer, F., and Ginsberg, J. P. 2017. "An Overview of Heart Rate Variability Metrics and Norms," *Frontiers in Public Health* (5).

Shuai, X., Zhou, Z., and Yost, R. S. 2003. "Using Segmented Regression Models to Fit Soil Nutrient and Soybean Grain Yield Changes Due to Liming," *Journal of Agricultural, Biological, and Environmental Statistics* (8:2), pp. 240–252.

Sim, C. S., Sung, J. H., Cheon, S. H., Lee, J. M., Lee, J. W., and Lee, J. 2015. "The Effects of Different Noise Types on Heart Rate Variability in Men," *Yonsei Medical Journal* (56:1), pp. 235–243.

Skrondal, A., and Rabe-Hesketh, S. 2004. "Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models," *CRC Press*.

Soares-Miranda, L., Sattelmair, J., Chaves, P., Duncan, G. E., Siscovick, D. S., Stein, P. K., and Mozaffarian, D. 2014. "Physical Activity and Heart Rate Variability in Older Adults: The Cardiovascular Health Study," *Circulation*.

Srinivasan, K., Currim, F., Ram, S., Mehl, M. R., Lindberg, C., Sternberg, E., Skeath, P., Najafi, B., Razjouyan, J., Lee, H.-K., Lunden, M., Goebel, N., Andrews, S., Herzl, D., Herzl, R., Gilligan, B., Heerwagen, J., Kampschroer, K., and Canada, K. 2017. "A Regularization Approach for Identifying Cumulative Lagged Effects in Smart Health Applications," in *Proceedings of the 7th International Conference on Digital Health*, London, United Kingdom: ACM Press, pp. 99–103.

Sternberg, E., Gilligan, B., and Lindberg, C. 2016. "Health and Wellbeing in GSA Office Buildings and Beyond," Philadelphia.

Sun Sim, C., Hyun Sung, J., Hyeon Cheon, S., Myung Lee, J., Won Lee, J., and Lee, J. 2015. "The Effects of Different Noise Types on Heart Rate Variability in Men," *Yonsei Medical Journal* (56:1).

Thayer, J. F., Verkuil, B., Brosschot, J. F., Kampschroer, K., West, A., Sterling, C., Christie, I. C., Abernethy, D. R., Sollers, J. J., Cizza, G., Marques, A. H., and Sternberg, E. M. 2010. "Effects of the Physical Work Environment on Physiological Measures of Stress," *European Journal of Cardiovascular Prevention and Rehabilitation* (17:4), pp. 431–9.

Tibshirani, R. 1996. "Regression Selection and Shrinkage via the Lasso," *Journal of the Royal Statistical Society B*, pp. 267–288.

Verkuil, B., Brosschot, J. F., Tollenaar, M. S., Lane, R. D., and Thayer, J. F. 2016. "Prolonged Non-Metabolic Heart Rate Variability Reduction as a Physiological Marker of Psychological Stress in Daily Life," *Annals of Behavioral Medicine*.

Walker, E. D., Brammer, A., Cherniack, M. G., Laden, F., and Cavallari, J. M. 2016. "Cardiovascular and Stress Responses to Short-Term Noise Exposures—A Panel Study in Healthy Males," *Environmental Research* (150:October), pp. 391–397.

Xhyheri, B., Manfrini, O., Mazzolini, M., Pizzi, C., and Bugiardini, R. 2012. "Heart Rate Variability Today," *Progress in Cardiovascular Diseases* (55:3), pp. 321–331.

Xue, L., Ray, G., and Gu, B. 2011. "Environmental Uncertainty and IT Infrastructure Governance: A Curvilinear Relationship," *Information Systems Research* (22:2), pp. 389–399.

Zou, H. 2006. "The Adaptive Lasso and Its Oracle Properties," *Journal of the American Statistical Association* (101:476), Taylor & Francis, pp. 1418–1429.

Zou, H., and Hastie, T. 2005. “Regularization and Variable Selection via the Elastic-Net,” *Journal of the Royal Statistical Society* (67:2), pp. 301–320.

Supplementary materials

1. Multilevel model inference using Classical and Bayesian approaches

Multilevel or hierarchical levels of grouped data are a commonly occurring phenomenon (Raudenbush and Bryk 2002). For example, in organizational studies, information about firms as well as workers are available such that there exists a hierarchical structured data of individual workers nested within multiple firms. Multilevel models (also called as hierarchical linear models, random coefficients models, mixed-effects models) are statistical models with parameters that capture variability across multiple levels of data.

In the classical or frequentist approach, multilevel models can be considered as an extension of ordinary least squares (OLS) regression model that is used to analyze variance in the outcome variables when the predictor variables are at varying hierarchical levels. A two-level hierarchical linear model can be mathematically expressed as follows:

$$\text{Level 1:} \qquad Y_{ij} = \beta_{0j} + \textstyle\sum_{k=1}^{K} \beta_{kj} V_{kij} + r_{ij} \tag{18}$$

$$\text{Level 2:} \qquad \beta_{kj} = \gamma_{k0} + \textstyle\sum_{m=1}^{M} \gamma_{km} W_{mj} + u_{kj} \tag{19}$$

Where Y_{ij} is the outcome, $\beta_{kj}$are the level-1 coefficients, V_{kij} are level-1 input variables, r_{ij} are level-1 residuals, γ_{km} are level-2 coefficients, W_{mj} are level-2 input variables and u_{kj} are level-2 variables for i^{th}observation of j^{th}individual for $k \in \mathbb{Z}_K$ and $m \in \mathbb{Z}_M$. The assumptions for the model are as follows:

$$E(r_{ij}) = 0; var(r_{ij}) = \sigma^2;\ E(u_{kj}) = 0; cov(u_{kj}, r_{ij}) = 0\ \forall i, j, k;\ \begin{bmatrix} u_{11} & \cdots \\ \cdots & u_{kj} \end{bmatrix} = T \tag{20}$$

Where T is the level-2 variance covariance component that model the inter-relationship between level-2 errors. Combining equations (1) and (2), we can represent hierarchical linear models as follows:

$$y_{ij} = \beta_0 + \gamma_{0j} + \sum_{k=1}^{K} \beta_k x_{kij} + \sum_{m=1}^{M} \gamma_{mj} z_{mij} + \epsilon_{ij} \tag{21}$$

Where $\beta = \{\beta_0, \beta_1, \dots, \beta_K\}$ are fixed effects coefficients, $\gamma = \{\gamma_{0j}, \gamma_{1j}, \dots, \gamma_{Mj}\}$ are random-effects coefficients for J groups $j \in \mathbb{Z}_J$ and ϵ_{ij} is the sum of fixed-effects error and random-effects error components. In matrix notation, the above equation is represented as follows:

$$Y = \alpha + X\beta + Z\gamma + \epsilon \tag{22}$$

Where X is a matrix of fixed effects and Z is a matrix of random effects. Conditional to the above assumptions, the parameters in the model can be estimation by maximizing the likelihood function y as shown below:

$$y \sim N(\alpha + X\beta, \sigma^2 I + Z'TZ) \tag{23}$$

The significance of the fixed effects and random effects are tested using Wald test, Likelihood Ratio Test, F-test, parametric bootstrap or MCMC methods (Raudenbush and Bryk 2002). Model fit can be compared using AIC, deviance and R-squared approximations (Nakagawa and Schielzeth 2013).

Bayesians on the other in-hand describe their beliefs about the unknows in a hierarchical linear model before observing data with prior distributions and the following likelihood function:

$$y \sim N(\alpha + X\beta + Zb, \sigma^2 I) \tag{24}$$

A single level regression disregards between-group heterogeneity is called model with complete pooling and can yield parameter estimates that are wrong if there is between-group heterogeneity. On the other hand, regression models for each group of the level-2 data independently is called modeling with no pooling and results imprecise parameter estimates as it ignores common variance across groups. Hierarchical linear models are considered as a subset of Hierarchical Bayesian models that are models with partial pooling (Gelman and Hill 2007). Parameters are allowed to vary by group at lower levels of the hierarchy while estimating common parameters at higher levels. Note that the level-2 and higher effects are not part of the error variance as in the classical/frequentist approach but modeled as parameters themselves (also called varying coefficients). The varying parameters have hyper-parameters that are estimated based on level-2 and higher order grouping in the data. The estimated posterior distribution of parameters for a hierarchical linear model with normally distributed error and identity link function has the following form:

$$p\left(\alpha, \beta, \gamma, \sigma_Y, \sigma_\gamma \middle| Y, X, Z, U\right) \propto \tag{25}$$

$$\prod_{j=1}^{J} \prod_{i=1}^{n_j} N(Y| \ \beta_0 + \gamma_{0j} + \beta X + \gamma_j Z, \sigma_Y^2) \prod_{j=1}^{J} N\left(\gamma_{0j}, \gamma_j \middle| \alpha_0 + \alpha U, \ \sigma_\gamma^2\right)$$

MCMC estimation approaches such as Metropolis Hastings, Gibbs Sampling, Hamiltonian Monte Carlo families of methods are used to estimate the posterior probability given the prior distribution of all parameters and likelihood of given data (Gelman et al. 2014). Comparison of implementations and general purpose software packages for classical and Bayesian multilevel modeling is done in West and Galecki 2011 (West and Galecki 2011), Mai and Zhang 2018 (Mai and Zhang 2018) respectively.

2. Univariate transformation-based modeling method

In the univariate transformation-based modeling method, we first convert the data into a long format as shown in Figure 8.

Row no.	Person ID	Outcome 1	Outcome 2
1	1	36.48	10.05
2	1	37.12	7.33
3	1	60.83	13.68
4	2	38.99	7.11
5	2	73.80	10.75
6	3	48.02	12.22
…	…	…	…

Row no.	Person ID	Outcome label	Value
1	1	Outcome 1	36.48
2	1	Outcome 1	37.12
3	1	Outcome 1	60.83
4	2	Outcome 1	38.99
5	2	Outcome 1	73.80
6	3	Outcome 1	48.02
7	1	Outcome 2	10.05
8	1	Outcome 2	7.33
9	1	Outcome 2	13.68
10	2	Outcome 2	7.11
11	2	Outcome 2	10.75
12	3	Outcome 2	12.22
…	…	…	…

Figure 8: Converting data in wide format to long format for univariate representation of multivariate outcomes

Equation resembles the equation for multilevel model for univariate outcome, except that each variable is subscripted with h which indicates the value for the h^{th} outcome.

$$y_{hij} = \beta_0 + \gamma_{0j} + \sum_{k=1}^{K} \beta_k x_{hkij} + \sum_{m=1}^{M} \gamma_{mj} z_{hmij} + \epsilon_{ij}^{(h)} \tag{26}$$

Here, residual errors $\epsilon_{ij}^{(h)}$ are defined as $N(0, \sigma_h^2)$ with estimated independent error variances for each outcome h. Marginal contributions of any outcome on input variables can be derived by including an indicator function in an interaction effect. For example, to derive marginal effects of level-1 variables, model is shown as follows:

$$y_{hij} = \beta_0 + \gamma_{0j} + \sum_{k=1}^{K} \beta_k x_{hkij} \cdot I(h) + \sum_{m=1}^{M} \gamma_{mj} z_{hmij} + \epsilon_{ij}^{(h)} \tag{27}$$

Where $I(\varphi)$ is an indicator function equal to 1 if condition φ is true else 0.

3. Noise sources and their sound levels[8]

Table 1: Noise sources and sound levels

Noise Source	Decibel Level (dBA)	Decibel Effect
Jet take-off (at 25 meters)	150	Eardrum rupture
Aircraft carrier deck	140	
Military jet aircraft take-off from aircraft carrier with afterburner at 50 ft (130 dBA).	130	
Thunderclap, chain saw. Oxygen torch (121 dBA).	120	Painful. 32 times as loud as 70 dBA.
Steel mill, auto horn at 1 meter. Turbo-fan aircraft at takeoff power at 200 ft (118 dBA). Riveting machine (110 dBA); live rock music (108 - 114 dBA).	110	Average human pain threshold. 16 times as loud as 70 dBA.
Jet take-off (at 305 meters), use of outboard motor, power lawn mower, motorcycle, farm tractor, jackhammer, garbage truck. Boeing 707 or DC-8 aircraft at one nautical mile (6080 ft) before landing (106 dBA); jet flyover at 1000 feet (103 dBA); Bell J-2A helicopter at 100 ft (100 dBA).	100	8 times as loud as 70 dBA. Serious damage possible in 8 hr exposure.
Boeing 737 or DC-9 aircraft at one nautical mile (6080 ft) before landing (97 dBA); power mower (96 dBA); motorcycle at 25 ft (90 dBA). Newspaper press (97 dBA).	90	4 times as loud as 70 dBA. Likely damage in 8 hour exposure.
Garbage disposal, dishwasher, average factory, freight train (at 15 meters). Car wash at 20 ft (89 dBA); propeller plane flyover at 1000 ft (88 dBA); diesel truck 40 mph at 50 ft (84 dBA); diesel train at 45 mph at 100 ft (83 dBA). Food blender (88 dBA); milling machine (85 dBA); garbage disposal (80 dBA).	80	2 times as loud as 70 dBA. Possible damage in 8 hour exposure.
Passenger car at 65 mph at 25 ft (77 dBA); freeway at 50 ft from pavement edge 10 a.m. (76 dBA). Living	70	Arbitrary base of comparison. Upper 70s are

[8] Source: IAC Acoustic website: http://www.industrialnoisecontrol.com/comparative-noise-examples.htm

room music (76 dBA); radio or TV-audio, vacuum cleaner (70 dBA).		annoyingly loud to some people.
Conversation in restaurant, office, background music, Air conditioning unit at 100 feet.	60	Half as loud as 70 dBA. Fairly quiet.
Quiet suburb, conversation at home. Large electrical transformers at 100 feet.	50	One-fourth as loud as 70 dBA.
Library, bird calls (44 dBA); lowest limit of urban ambient sound	40	One-eighth as loud as 70 dBA.
Quiet rural area.	30	One-sixteenth as loud as 70 dBA. Very Quiet.
Whisper, rustling leaves	20	
Breathing	10	Barely audible

References

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. 2014. "Bayesian Data Analysis," *Bayesian Data Analysis*, CRC press.

Gelman, A., and Hill, J. 2007. "Data Analysis Using Regression and Multilevel/Hierarchical Models," *Cambridge*.

Mai, Y., and Zhang, Z. 2018. "Software Packages for Bayesian Multilevel Modeling," *Structural Equation Modeling: A Multidisciplinary Journal* (25:4), Routledge, pp. 650–658.

Nakagawa, S., and Schielzeth, H. 2013. "A General and Simple Method for Obtaining R2 from Generalized Linear Mixed-Effects Models," *Methods in Ecology and Evolution* (4:2), pp. 133–142.

Raudenbush, S. W., and Bryk, A. S. 2002. "Hierarchical Linear Models: Applications and Data Analysis Methods," *Advanced Quantitative Techniques in the Social Sciences 1* (Vol. 2nd).

West, B. T., and Galecki, A. T. 2011. "An Overview of Current Software Procedures for Fitting Linear Mixed Models," *The American Statistician* (65:4), pp. 274–282.