

Noise exposure-response relationships established from repeated binary observations: Modeling approaches and applications

Beat Schäffer^{a)} and Reto Pieren

Empa, Swiss Federal Laboratories for Materials Science and Technology, Laboratory for Acoustics/Noise Control, Überlandstrasse 129, 8600 Dübendorf, Switzerland

Franco Mendolia

German Aerospace Center (DLR), Institute of Aerospace Medicine, 51147 Cologne, Germany

Mathias Basner

Department of Psychiatry, Unit for Experimental Psychiatry, Division of Sleep and Chronobiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104-6021, USA

Mark Brink

Federal Office for the Environment, Noise and NIR Division, 3003 Bern, Switzerland

(Received 16 December 2016; revised 16 March 2017; accepted 20 April 2017; published online 10 May 2017)

Noise exposure-response relationships are used to estimate the effects of noise on individuals or a population. Such relationships may be derived from independent or repeated binary observations, and modeled by different statistical methods. Depending on the method by which they were established, their application in population risk assessment or estimation of individual responses may yield different results, i.e., predict “weaker” or “stronger” effects. As far as the present body of literature on noise effect studies is concerned, however, the underlying statistical methodology to establish exposure-response relationships has not always been paid sufficient attention. This paper gives an overview on two statistical approaches (subject-specific and population-averaged logistic regression analysis) to establish noise exposure-response relationships from repeated binary observations, and their appropriate applications. The considerations are illustrated with data from three noise effect studies, estimating also the magnitude of differences in results when applying exposure-response relationships derived from the two statistical approaches. Depending on the underlying data set and the probability range of the binary variable it covers, the two approaches yield similar to very different results. The adequate choice of a specific statistical approach and its application in subsequent studies, both depending on the research question, are therefore crucial.

© 2017 Acoustical Society of America. [<http://dx.doi.org/10.1121/1.4982922>]

[SF]

Pages: 3175–3185

I. INTRODUCTION

Noise exposure-response relationships are of importance to estimate the effects of noise on either individuals or a population. In recent years, risk assessment of environmental noise, i.e., assessing the health impact of noise exposure on the population, became an important topic for policy makers, authorities, and in public discussions (WHO, 2009; EAA, 2010; WHO, 2011). It is also required by the Environmental Noise Directive 2002/49/EC (European Union, 2002) to establish action plans. For risk assessment, population exposure indicators (EAA, 2010) such as the Disability-Adjusted Life Years (de Hollander *et al.*, 1999; WHO, 2011) or effect-based noise indices (Brink *et al.*, 2010) are calculated and used. To that aim, appropriate exposure-response relationships for binary data such as high annoyance (to be highly annoyed by noise or not) or awakening reactions (to awake from a noise event or not) are combined with spatial noise exposure and population data, and summed up to a single number for the considered effect dimensions (e.g., Brink *et al.*, 2010; Schäffer

et al., 2012). Besides risk assessment on the population level, the research focus may also be on the responses of individuals to noise exposure, e.g., in medical studies, or to establish noise protection concepts (Basner, 2009).

Exposure-response relationships for binary data, which are the focus of the present paper, may be derived from independent binary observations (one observation per subject), as well as from repeated binary observations (repeated observations over different points in time of the same outcome in the same subject, in a sample of multiple subjects). Depending on the method by which they were established, their application in population risk assessment or estimation of individual responses may be more or less straightforward and yield different results, i.e., predict “weaker” or “stronger” effects. However, while this critical point has been discussed in epidemiology, medicine, or statistics (e.g., Hu *et al.*, 1998; Neuhaus *et al.*, 1991; Zeger *et al.*, 1988), it was so far not sufficiently considered in noise effects research.

This paper provides an overview of two common statistical methods used to establish exposure-response relationships for repeated binary observations, namely, a subject-specific (SS) and a population-averaged (PA) approach, and their

^{a)}Electronic mail: beat.schaeffer@empa.ch

appropriate applications. The considerations are illustrated with original data from three noise effect studies of the pertinent literature, estimating also the magnitude of differences in results when applying exposure-response relationships derived from the two different statistical approaches. Based on the insights, practical implications for noise effect studies are discussed.

II. LOGISTIC REGRESSION MODELS FOR REPEATED BINARY OBSERVATIONS: BACKGROUND

This section provides an overview of two common logistic regression modeling approaches for repeated binary observations to establish exposure-response relationships for the probability of a certain noise effect (e.g., awakening probability, probability of high annoyance) in environmental noise research. Repeated binary observations, which are referred to in epidemiology as “binary longitudinal data,” are obtained from repeated measurements of the same binary variable over different points in time. Examples are awakenings to noise events in the night (Basner *et al.*, 2006; Brink *et al.*, 2011), motility (Passchier-Vermeer *et al.*, 2002; Brink *et al.*, 2008a), behavioral awakenings (Passchier-Vermeer, 2003), or annoyance reactions (Van Renterghem *et al.*, 2013; Schäffer *et al.*, 2016). Independent data (non-nested and non-hierarchical data), in contrast, are often obtained in field surveys, where each study participant gives one single rating, for example, on high annoyance (e.g., Brink *et al.*, 2008b; Janssen *et al.*, 2011; Michaud *et al.*, 2016b).

As long as the binary observations are independent, a “standard” binary logistic regression analysis may be applied (Hosmer and Lemeshow, 2000), which yields PA exposure-response relationships (Hu *et al.*, 1998) that are directly applicable in risk assessment on the population level. Things are different if data from repeated binary observations are collected. Here, one needs to account for the correlation of the data within subjects when establishing a statistical model.

Among others, possible approaches to do so are to use either a PA (Liang and Zeger, 1986; Zeger *et al.*, 1988; Hanley *et al.*, 2003) or a SS approach (Stiratelli *et al.*, 1984; Anderson and Aitkin, 1985). Differences between the two approaches have been previously discussed in the literature about epidemiologic, medical, and social sciences research (Zeger *et al.*, 1988; Neuhaus *et al.*, 1991; Hu *et al.*, 1998; Szymaragd *et al.*, 2013). Both approaches estimate mean parameters for the logit link function (e.g., Hosmer and Lemeshow, 2000), from which the probability of the dependent binary variable to adapt the value of 1 (e.g., to be highly annoyed) is determined. From a statistical point of view, the primary difference is that PA uses a certain working correlation matrix to account for the correlation between the repeated observations (see, e.g., Jang, 2011), while SS introduces a random effect, one for each subject, modeled in this study as a random intercept (u). Regarding the interpretation of results, the main difference is that PA describes the average population’s response, while SS describes the response of an average individual. Further, with the random effect of SS, individual responses and the magnitude of variation

between individuals are quantified, which can be interpreted as the propensity of subjects to react more or less strongly to the predictor variable(s).

The resulting estimates (absolute values) of the model parameters of the SS approach are generally larger than those of the PA approach (Neuhaus *et al.*, 1991), i.e., predict “stronger” effects, and the differences between the SS and PA parameters increase with increasing variation (or heterogeneity) between individuals (Zeger *et al.*, 1988). The SS logistic regression curves are therefore often steeper than the PA curves, with the latter being dominated by a few sensitive subjects at low values of a predictor variable and by a few resilient subjects at high values of a predictor variable, as illustrated in Fig. 1. The intersection point of the SS curve of the average subject (i.e., the subject with the marginal mean curve, having a random intercept $u=0$) and the PA curve is at the value that yields a probability of 0.5 (Fig. 1), given that the random intercept follows a normal distribution (Zeger *et al.*, 1988), the latter meaning that sensitive and resilient subjects are equally represented in a population. Thus, the SS curve of the average subject ($u=0$) usually predicts higher probabilities of an effect than PA curves at values of the predictor variable above the intersection point of the two curves and vice versa (Fig. 1).

Due to the inherent differences in the modeling, it is crucial to decide which approach to use. Hu *et al.* (1998) recommend PA models “when the research focus is on differences in population-averaged response,” but SS models “when the research focus is on the change in individuals’ responses.” Thus, PA models are preferred for epidemiological studies (Zeger *et al.*, 1988) or for risk assessment, e.g., to assess the impact of environmental noise on the population. SS models, in contrast, are to be used to estimate the magnitude of inter-individual variability, if the distribution of the dependent variable of the target population (i.e., specific study group in an analysis) differs from the distribution of the study sample used to estimate the model, e.g., in establishing a protection

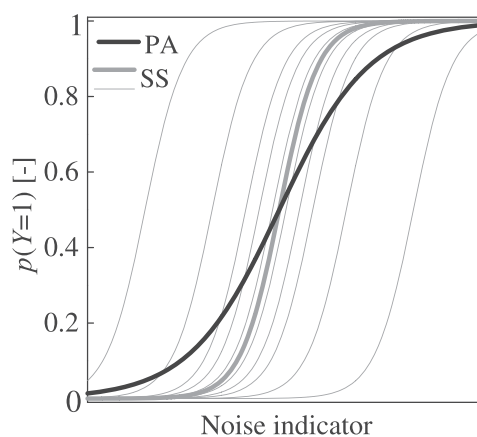


FIG. 1. Illustrative example of a random intercept SS (grey lines) and PA (black line) logistic regression curve for the probability of the dependent binary variable Y to adapt the value of 1 [$p(Y=1)$] as a function of some noise indicator. The bold grey line represents the SS curve as marginal mean, i.e., for a random intercept $u=0$, and the thin grey lines represent individual SS curves, with different random intercepts to account for SS random effects.

concept for a particularly sensitive population group (Basner, 2009), or to study responses of individuals (e.g., patients in clinical studies). While SS models, by accounting for individual responses, generally yield higher agreement between predicted and observed individual values, they should not be chosen for this reason.

III. APPLICATION OF LOGISTIC REGRESSION MODELS TO ORIGINAL DATA FROM THREE NOISE EFFECT STUDIES

In the following account, the two modeling approaches (SS and PA) are applied to the original data sets of repeated binary observations of three studies from the literature. The resulting logistic regression curves are compared to each other, and differences between the two model results are discussed.

A. Overview of the three studies used for the present re-analysis

1. Study I: High annoyance to wind turbine and road traffic noise

In study I (Schäffer *et al.*, 2016), an exposure-response relationship for the probability of the binary variable “high annoyance” to take a value of 1 (*p*HA) due to wind turbine and road traffic noise was established. The relationship was modeled as a function of the source type (source: wind turbines, road traffic), A-weighted equivalent continuous sound pressure level (L_{Aeq}) of “outdoor” situations (see Schäffer *et al.*, 2016), amplitude modulation (AM: without, random, periodic), their interactions source \times AM, source \times L_{Aeq} , and AM \times L_{Aeq} , as well as the sequence number, i.e., the playback number with which the stimuli had been played to the participants (Table II and Fig. 10 in Schäffer *et al.*, 2016). The underlying data set was obtained from repeated binary observations in a laboratory experiment of 60 participants who were exposed to stimuli with a L_{Aeq} range of 35–60 dB. A PA modeling approach was chosen as the focus of the study was to estimate the mean *p*HA in the population.

2. Study II: Awakening reactions to aircraft noise

In study II (Basner *et al.*, 2006), an exposure-response relationship for the probability of the binary variable “awakening reaction” to take a value of 1 (*p*AWR) due to aircraft noise was established. The relationship was modeled as a function of the A-weighted and SLOW-time-weighted maximum sound pressure level of an aircraft event ($L_{AS,max}$) inside the bedroom near the sleeper’s ear, background sound pressure level one minute prior to the event ($L_{Aeq,1\ min}$), their interaction $L_{AS,max} \times L_{Aeq,1\ min}$, elapsed sleep time, and sleep stage prior to an aircraft event (S2, S3, S4, or REM; Table I and Fig. 1 in Basner *et al.*, 2006). The underlying data set was obtained from repeated binary observations in a field study with 61 participants who were exposed to a mean night-time $L_{AS,max}$ indoors of 44 dB, with a $L_{AS,max}$ range of 14–73 dB (Basner and Brink, 2013). An SS modeling approach was chosen as the primary focus was on the individuals’ responses to aircraft noise events.

3. Study III: Awakening reactions to church bell noise

In study III (Brink *et al.*, 2011), an exposure-response relationship for the probability of the binary variable “awakening reaction” to take a value of 1 (*p*AWR) due to church bell noise was established. The relationship was modeled as a function of the A-weighted and FAST-time-weighted maximum sound pressure level of a church bell event ($L_{AF,max}$) inside the bedroom near the sleeper’s ear, $L_{Aeq,1\ min}$ (as in study II), the interaction $L_{AF,max} \times L_{Aeq,1\ min}$, elapsed sleep time, and sleep stage prior to a church bell event (S2, S3, S4, or REM; Table 5 and Fig. 6 in Brink *et al.*, 2011). The underlying data set was obtained from repeated binary observations in a field study with 27 participants who were exposed to a mean night-time $L_{AF,max}$ indoors of 42 dB, with a $L_{AF,max}$ range of 20–69 dB. The same SS model was established as in study II to allow for direct comparison of the effects of aircraft and church bell noise on sleep.

Note that in studies II and III, SS relationships for aircraft and church bell noise induced, (so called) *additional* awakenings, were established. These were obtained by subtracting the probability of *spontaneous* awakenings, i.e., without the triggering by a noise event, from the *observed* awakening probability (*p*AWR) introduced above (Basner *et al.*, 2006; Brink *et al.*, 2011). In the present re-analysis of the data sets, however, the focus is on the *observed* awakenings.

B. Re-analysis of the original data sets: Methodology

To compare possible results of PA and SS relationships and potential differences between them, the exposure-response relationships of both modeling approaches were established for the three studies based on their original data sets. As different statistical software packages may yield disparate results (Horton and Lipsitz, 1999; Zhang *et al.*, 2011), also the published relationships were re-established where applicable (namely, for studies II and III) to allow for comparison of the SS and PA relationships without the potential influence of the applied software packages. For the present analysis, the procedures GENLIN (for PA) and GENLINMIXED (for SS) of the software IBM SPSS Statistics Version 21 (study II) and Version 22 (studies I and III) were used.

1. Study I: High annoyance to wind turbine noise (laboratory study)

The PA relationship for *p*HA (high annoyance) is taken from Schäffer *et al.* (2016). The model considers the effects of source, L_{Aeq} , AM, their interactions, and sequence number as described above, and accounts for the repeated observations by an exchangeable working correlation matrix, which is a practical choice for small samples (Jang, 2011). The SS relationship was established in this study to consider the same explanatory variables, but accounting for the repeated observations with a random intercept. In the following account, only the *p*HA relationship as a function of wind turbine noise is presented.

2. Study II: Awakening reactions to aircraft noise (field study)

The PA and SS relationships for $pAWR$ (observed awakenings) were both newly established to account for the effects of the inside $L_{AS,max}$, $L_{Aeq,1 min}$, their interaction, elapsed sleep time, and sleep stage prior to an aircraft noise event, as in the original model described above. Repeated observations were accounted for as in study I by an exchangeable working correlation matrix (PA) and by a random intercept (SS).

3. Study III: Awakening reactions to church bell noise (field study)

The same PA and SS relationships for $pAWR$ (observed awakenings) were newly established here as for study II, with regard to the explanatory variables (except for using $L_{AF,max}$ instead of $L_{AS,max}$), as well as the repeated observations.

4. Comparison of modeling results with the observed data

Along with the PA and SS relationships, the mean values of the observed data are presented in the following account to put the models in context of the observations. To that aim, the original individual observations were “globally” averaged over all events per dB class with bin width of 5 dB, treating each observation as independent from the other observations, i.e., neglecting any correlation between observations per subject. While the clustering of observations would have been accounted for by “hierarchical averaging” (i.e., averaging first per subject and then over all subjects), the mean within some dB classes would have been strongly shifted up- or downward in the case of the data sets of studies II and III due to subjects with only one or a few observations with a relative frequency of awakenings of 1 or 0, which would not seem appropriate.

Note that the observed data do not allow deciding which relationship is “more appropriate” to represent observations for two reasons. First, the appropriate model (SS or PA) is given by the research goal and not merely by the degree of agreement with observed data (Sec. II). Second, the regression curves adjust for the other predictor variables (besides sound pressure level) as well as for repeated observations, while the averaging does not. Differences between model predictions and observations are therefore not (only) attributable to model shortcomings, but, in particular, also to differences between observed averaged and adjusted predictor variables. In particular in studies II and III, the observed averages of the predictor variables ($L_{Aeq,1 min}$, interaction $L_{AS,max} \times L_{Aeq,1 min}$, elapsed sleep time and sleep stage) may strongly differ between classes, while they are kept constant in the models. In study I with its (full factorial) laboratory design, in contrast, each dB class contains the same subjects, number of subjects, number of observations, and acoustic situations, and the model parameters were adjusted to their mean values. Only sequence number could not be exactly adjusted to the (individual) observed mean values, as these varied between dB classes (see Sec. III C 1).

C. Results: Differences in predicted effect probabilities

1. Study I: High annoyance to wind turbine noise (laboratory study)

Figure 2 shows the mean observed relative frequencies of high annoyance and the logistic regression models for pHA for wind turbine noise. The model parameters are presented in Table I. The mean observations cover a wide range of relative frequencies with values from 0.08 to 0.82 (Fig. 2). With the model parameters set to the mean sequence number and averaging over different situations of AM during the experiments, the PA relationship closely represents the mean observed relative frequencies, except for $L_{Aeq} = 35$ dB. Here, the sequence number was substantially larger than the mean, because these stimuli were presented after the other stimuli (Schäffer *et al.*, 2016). The SS relationship, in contrast, is substantially steeper than the PA relationship, which was expected (Fig. 1). This indicates a large heterogeneity between individuals (Sec. II), which is supported by the large random intercept variance of ~ 5.0 (Table I). The confidence intervals of the SS and PA curves are non-overlapping at high and low L_{Aeq} , which may indicate that the SS and PA models predict significantly different pHA values there. Both models identify the same significant effects, namely, L_{Aeq} ($p < 0.01$), sequence number ($p < 0.01$), interaction source \times AM ($p < 0.02$), and in tendency also source ($p < 0.06$). Further, the SS approach yields mostly larger absolute parameter estimates than the PA approach (Table I), as was expected according to theory.

The SS and PA relationships intersect at a L_{Aeq} of 48 dB and corresponding pHA of ~ 0.47 , which closely corresponds to the expected value of 0.5 (Fig. 1). Below this level, the PA relationship predicts larger pHA values than the SS

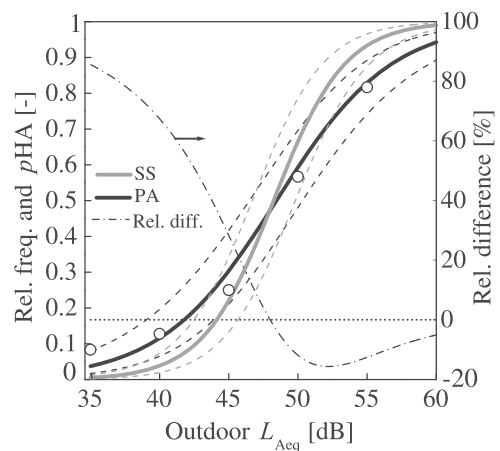


FIG. 2. Mean observed relative frequencies (Rel. freq., circles) of high annoyance, and logistic exposure-response relationships (lines) for the probability of high annoyance (pHA) as a function of the outdoor equivalent continuous sound pressure level (L_{Aeq}) of wind turbine noise (study I, Schäffer *et al.*, 2016). SS relationship as marginal mean ($u=0$) and corresponding PA relationship with 95% confidence intervals (dashed lines), and relative differences between the approaches [Rel. diff. = $(PA - SS)/PA \times 100\%$, right axis], with the dotted horizontal line indicating the difference of 0. The parameter settings of the model curves correspond to the mean pHA for different situations of amplitude modulation (no, with periodic and random AM) and mean sequence number = 15.5 of the stimuli during the experiments.

TABLE I. Model coefficients (Coeff.) with standard error (SE) and p -values (p) of the SS and PA logistic regression models for the probability of high annoyance (p HA) due to wind turbine noise (study I), accounting for the effects of amplitude modulation (AM), equivalent continuous outdoor sound pressure level (L_{Aeq}), the interaction $AM \times L_{Aeq}$, and sequence number. The parameters of the PA model are taken from Table II in Schäffer *et al.* (2016). Only the parameters necessary to calculate p HA as a function of wind turbine noise are presented here, while the original model also contains parameters to account for road traffic noise.

p HA, Wind turbine noise Parameter	SS model			PA model		
	Coeff.	SE	p	Coeff.	SE	p
Intercept	-19.5433	1.5285	0.00	-12.0779	1.3072	0.00
Without AM (yes = 1, no = 0) ^a	-1.8098	1.7092	0.29	-1.2172	1.1624	0.30
Periodic AM (yes = 1, no = 0) ^a	-0.1605	1.6264	0.92	-0.1739	0.7981	0.83
L_{Aeq}	0.3859	0.0305	0.00	0.2359	0.0254	0.00
Without AM $\times L_{Aeq}$ (yes = 1, no = 0) ^b	0.0210	0.0356	0.56	0.0150	0.0244	0.54
Periodic AM $\times L_{Aeq}$ (yes = 1, no = 0) ^b	0.0035	0.0343	0.92	0.0037	0.0164	0.82
Sequence number	0.0767	0.0115	0.00	0.0526	0.0091	0.00
Random intercept (variance)	4.9960	1.1021	0.00	—	—	—

^aBoth parameters (without AM and periodic AM) set to zero correspond to random AM.

^bBoth parameters (without AM $\times L_{Aeq}$ and periodic AM $\times L_{Aeq}$) set to zero correspond to random AM $\times L_{Aeq}$.

relationship, with relative differences of up to 86% at a $L_{Aeq} = 35$ dB and absolute differences of up to 0.10 at a $L_{Aeq} = 43$ dB. Above the intersection point ($L_{Aeq} > 48$ dB), in contrast, the PA relationship predicts smaller p HA values, with relative differences of up to -16% at a $L_{Aeq} = 52$ dB and absolute differences of up to -0.12 at a $L_{Aeq} = 53$ dB (Fig. 2). Residents in the vicinity of wind farms are usually exposed to sound pressure levels below 50 dB (e.g., Michaud *et al.*, 2016a; Pedersen and Persson Waye, 2004; Pedersen *et al.*, 2009). This means that applying the SS instead of the PA relationship to estimate the number of highly annoyed persons around wind farms would yield systematically smaller estimates of effects on the population. In this context, one further needs to consider that applying exposure-response relationships established in laboratory studies would be precarious, as the “short-term annoyance” derived in the laboratory might not be the same as the annoyance caused by long-term exposure derived from field surveys (Guski and Bosshardt, 1992), as discussed by Schäffer *et al.* (2016).

2. Study II: Awakening reactions to aircraft noise (field study)

Figure 3 shows the mean observed relative frequencies of awakening reactions and the logistic regression models for p AWR (model parameters: Table II). With values of 0.05–0.18, the relative frequencies cover a rather small probability range, and while increasing in tendency with $L_{AS,max}$, the dependence of p AWR on sound pressure level is less pronounced than in study I (Fig. 2). Accordingly, both the PA and SS curves are still in the initial “flat part” within the observed $L_{AS,max}$ values, with maximum predicted p AWR values of ~ 0.16 for SS and PA. Further, the PA and SS relationships are very similar, and the confidence intervals strongly overlap. With the model parameters corresponding to the mean values observed during the field study, both models represent the observations similarly (Fig. 3). In this context it is interesting to note that in study I (Fig. 2) the largest differences between SS and PA

were observed in the “flat part” of the model curves, which would be expected according to Fig. 1. In study II, however, these differences are very small. Again, both, SS and PA models reveal the same significant effects, namely, $L_{AS,max}$, $L_{Aeq,1 min}$, $L_{AS,max} \times L_{Aeq,1 min}$, elapsed sleep time, and sleep stage ($p < 0.01$), but contrary to study I, SS and PA yield very similar parameter estimates (Table II).

The SS and PA relationships are approximately parallel and do not intersect. Within the night-time $L_{AS,max}$ range indoors of 25–75 dB, the PA relationship yields 1.2%–2.5% larger p AWR values than the SS relationship, corresponding to absolute differences < 0.01 (Fig. 3). The differences for predicted p AWR between the two approaches are, thus, small.

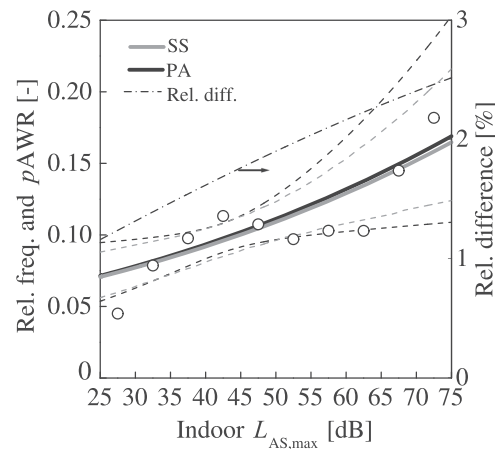


FIG. 3. Mean observed relative frequencies (Rel. freq., circles) of awakenings, and logistic exposure-response relationships (lines) for the probability of an observed awakening reaction to aircraft noise (p AWR) as a function of the maximum indoor sound pressure level ($L_{AS,max}$; study II, Basner *et al.*, 2006). SS relationship as marginal mean ($u = 0$) and corresponding PA relationship with 95% confidence intervals (dashed lines), and relative differences between the approaches [Rel. diff. = $(PA - SS)/PA \times 100\%$, right axis]. The parameter settings of the model curves correspond to the mean values observed during the field study [background noise level ($L_{Aeq,1 min}$) = 28.2 dB, different prior sleep stages (S2, S3, S4, and REM), elapsed sleep time = 417 epochs].

TABLE II. Model coefficients (Coeff.) with standard error (SE) and p -values (p) of the SS and PA logistic regression models for the probability of an observed awakening reaction (p AWR) due to aircraft noise (study II), accounting for the effects of the indoor maximum sound pressure level ($L_{AS,max}$), background sound pressure level one minute prior to the event ($L_{Aeq,1min}$), the interaction $L_{AS,max} \times L_{Aeq,1min}$, elapsed sleep time, and sleep stage prior to an aircraft event. Note that the SS model parameters differ somewhat from the parameters published in Table I of [Basner et al. \(2006\)](#) as they were established with another software package.

p AWR, Aircraft noise Parameter	SS model			PA model		
	Coeff.	SE	p	Coeff.	SE	p
Intercept	-7.0276	0.8734	0.00	-7.0367	1.1140	0.00
$L_{AS,max}$	0.0941	0.0184	0.00	0.0947	0.0215	0.00
$L_{Aeq,1min}$	0.1308	0.0327	0.00	0.1314	0.0396	0.00
$L_{AS,max} \times L_{Aeq,1min}$	-0.0027	0.0007	0.00	-0.0027	0.0008	0.00
Elapsed sleep time	0.0006	0.0002	0.00	0.0006	0.0002	0.00
Prior stages S3 and S4 (yes = 1, no = 0) ^a	-0.3209	0.1160	0.01	-0.3154	0.1567	0.04
Prior REM (yes = 1, no = 0) ^a	0.4181	0.0731	0.00	0.4114	0.0957	0.00
Random intercept (variance)	0.1180	0.0389		—		

^aBoth parameters (prior stages S3 and S4, and prior REM) set to zero correspond to prior stage S2.

3. Study III: Awakening reactions to church bell noise (field study)

Figure 4 shows the mean observed relative frequencies of awakening reactions, as well as the logistic regression models for p AWR (model parameters: Table III). Partly similar observations may be made as for study II. First, the observed relative frequencies cover a rather small probability range of 0.08–0.24 and, while tending to increase with $L_{AF,max}$, show a less clear dependence on the sound pressure level as in study I. Second, the PA and SS relationships for p AWR are similar, and their confidence intervals strongly overlap. Third, with the model parameters corresponding to the mean values observed in the field study, both models represent the observations similarly (Fig. 4). Finally, both models identify the same significant effect (namely, $p < 0.03$ for $L_{AF,max}$).

However, study III also yields disparate findings from study II. First, contrary to study II but in line with study I, SS yields larger absolute parameter estimates than PA (Table III). Second, the SS and PA curves of study III cover a relative large p AWR range of up to 0.39 (PA) and 0.44 (SS) at a $L_{AF,max}$ of 75 dB. Third, as would be expected (Fig. 1), the PA curve is somewhat flatter than the SS curve (although much less than in study I), but the curves intersect at p AWR of 0.16, which is much smaller than the theoretic value of 0.5.

Below the intersection point of the SS and PA curves at a $L_{AF,max}$ of 50 dB, the PA relationship predicts larger p AWR values than the SS relationship, with relative differences of up to 18% and absolute differences of 0.01 at a $L_{AF,max} = 25$ dB. Above the intersection point, in contrast, the PA relationship predicts smaller p AWR values, with relative differences of up to -13% and absolute differences of up to -0.05 at a $L_{AF,max} = 75$ dB (Fig. 4). For church bell noise, a large part of awakenings may be expected at distances of 100–200 m from churches (Fig. 6 in [Brink et al., 2012](#), product of awakening reactions \times number of people per distance class), corresponding to an inside $L_{AF,max}$ of ~ 35 –45 dB (Fig. 7 in [Brink et al., 2012](#)). In this $L_{AF,max}$

range, the PA relationship yields $\sim 3\%$ – 11% larger p AWR values than the SS relationship, corresponding to absolute differences of p AWR < 0.01 (Fig. 4). Thus, as for study II, the relevant differences between the two approaches are small.

4. Reasons for deviations from the theoretical relation between the model approaches

Re-analysis of the data of the above three studies shows that the differences between PA and SS vary to a lesser or larger extent, depending on the underlying data set and range of observed relative frequencies of the binary variable. In study I, where the observations cover a wide p HA range, the SS relationship is distinctly steeper than the PA relationship,

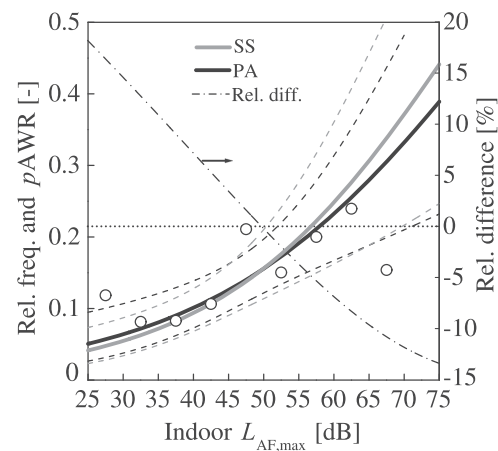


FIG. 4. Mean observed relative frequencies (Rel. freq., circles) of awakenings, and logistic exposure-response relationships (lines) for the probability of an observed awakening reaction to church bell noise (p AWR) as a function of the maximum indoor sound pressure level ($L_{AF,max}$; study III, [Brink et al., 2011](#)). SS relationship as marginal mean ($u = 0$) and corresponding PA relationship with 95% confidence intervals (dashed lines), and relative differences between the approaches [Rel. diff. = $(PA - SS)/PA \times 100\%$, right axis], with the dotted horizontal line indicating the difference of 0. The parameter settings of the model curves correspond to the mean values observed during the field study [background noise level ($L_{Aeq,1min}$) = 26.7 dB, different prior sleep stages (S2, S3, S4, and REM), elapsed sleep time = 395 epochs].

TABLE III. Model coefficients (Coeff.) with standard error (SE) and p -values (p) of the SS and PA logistic regression models for the probability of an observed awakening reaction (p AWR) due to church bell noise (study III), accounting for the effects of the indoor maximum sound pressure level ($L_{AF,max}$), background sound pressure level one minute prior to the event ($L_{Aeq,1min}$), the interaction $L_{AF,max} \times L_{Aeq,1min}$, elapsed sleep time, and sleep stage prior to a church bell event. Note that the SS model parameters differ somewhat from the parameters published in Table 5 of Brink *et al.* (2011) as they were established with another software package.

p AWR, Church bell noise Parameter	SS model			PA model		
	Coeff.	SE	p	Coeff.	SE	p
Intercept	-7.1052	2.1493	0.00	-6.3923	2.1395	0.00
$L_{AF,max}$	0.1112	0.0467	0.02	0.0972	0.0421	0.02
$L_{Aeq,1min}$	0.0963	0.0772	0.21	0.0853	0.0775	0.27
$L_{AF,max} \times L_{Aeq,1min}$	-0.0020	0.0017	0.24	-0.0018	0.0015	0.24
Elapsed sleep time	0.0004	0.0004	0.25	0.0004	0.0004	0.33
Prior stages S3 and S4 (yes = 1, no = 0) ^a	-0.2352	0.3147	0.45	-0.2132	0.3055	0.49
Prior REM (yes = 1, no = 0) ^a	-0.4637	0.2048	0.02	-0.4568	0.2287	0.05
Random intercept (variance)	0.3519	0.1844		—		

^aBoth parameters (prior stages S3 and S4, and prior REM) set to zero correspond to prior stage S2.

and the intersection point is at p HA of ~ 0.5 (Fig. 2), which is as expected according to Sec. II. The models of studies II and III, in contrast, were established based on data covering a limited p AWR range only, and here the relation between the models is not as expected according to Sec. II. In study II, the two relationships run approximately parallel to each other (Fig. 3), while in study III, SS is steeper than PA, but the intersection point is at a much lower p AWR than the theoretical value of 0.5 (Fig. 4).

Reasons for disparate relations between the PA and SS models of the three studies may be that (i) the assumption of a normal distribution of the random intercept (SS) is not fully met in all studies, and/or (ii) the logistic regression analyses are based on different ranges of observed relative frequencies of the binary variable, and/or (iii) the differences between the SS and PA parameters depend on the magnitude of variation between individuals (Fig. 1), and/or (iv) disparate model complexities or lack of significance of certain model parameters contribute to the differences.

Reason (i) will hardly contribute to the disparate relations between PA and SS observed here. Visual inspection of the random intercepts of the three studies with residual plots did not reveal any obvious deviation from normality. This is not surprising, as the SS model is fit to the assumption of normality, and the random effect is thus forced toward normality. Also, the assumption of the working correlation matrix for PA is not critical, as the PA parameter estimates are consistent even if the assumed working correlation matrix is misspecified (Zeger *et al.*, 1988; Hu *et al.*, 1998). Reason (ii) is possibly the main reason for the above observations. Study I covers a wide range of observed relative frequencies, with p HA of up to 0.82, while studies II and III are based on relatively small ranges (p AWR < 0.25). Also reason (iii) might distinctly contribute to the differences. In fact, the random intercept variance is substantially different between the three studies (cf. Tables I–III). However, the magnitude of this contribution to the observed differences is difficult to quantify, as the random intercept variance is hardly comparable between the studies, as discussed in Sec. III C 5. Reason (iv), in contrast, is not expected to account for these differences because the models of the three studies are of similar

complexity, i.e., account for a similar number of parameters, and because both, the models of studies I and III, contain non-significant parameters (Tables I–III).

To test the presumption of reason (ii) for the disparate relations between the PA and SS models of the studies, the data of study I were re-analyzed as follows: For sake of simplicity, only the subset of the annoyance to wind turbine noise was used. Accordingly, the statistical models were reduced to account for the effects of L_{Aeq} , AM, their interaction $L_{Aeq} \times AM$, as well as sequence number. Then, separate PA and SS models were established by considering the data sets for L_{Aeq} ranges of 35–55 dB (5 L_{Aeq} levels, case 1), 35–50 dB (4 levels, case 2), 35–45 dB (3 levels, case 3), and 35–40 dB (2 levels, case 4). Figure 5 shows the results, allowing for the following observations: First [Fig. 5(a)], the PA and SS curves are very different over the whole L_{Aeq} range for the complete data (case 1). However, for the reduced data set (case 4) the differences between the p HA values predicted by the two curves are distinctly smaller. (The curves established from the intermediate cases 2 and 3 allow for similar observations.) Second [Fig. 5(b)], the differences between PA and SS progressively decrease with decreasing L_{Aeq} range covered by the data set, with maximum differences between PA and SS at a $L_{Aeq} = 35$ dB decreasing from +96% (case 1) to +95% (case 2) to +88% (case 3) to +69% (case 4). Here, it is interesting to note that along with the decreasing covered L_{Aeq} range, the random intercept variance progressively decreases from 9.688 (case 1) to 8.036 (case 2) to 6.096 (case 3) to 3.675 (case 4). Apparently, at low sound pressure levels more subjects were generally, and rather “homogeneously,” not highly annoyed (“floor effect”), which results in a smaller variance between subjects for case 4. At high sound pressure levels, in contrast, the differences in the ratings between subjects were more pronounced, as some subjects were highly annoyed while others were not, which results in a large variance for case 1. The same reason holds possibly true for studies II and III, where only a small range of observed relative frequencies of awakening reactions was covered and, accordingly, small variances are observed (cf. Tables II and III). In accordance with the changes in the random intercept

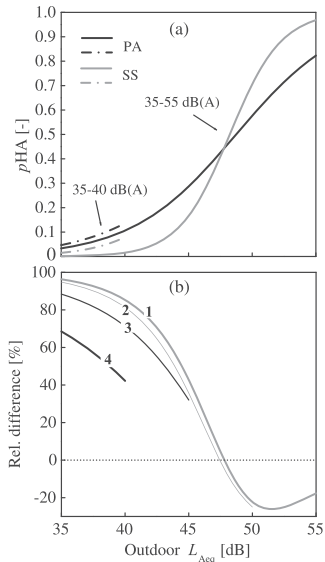


FIG. 5. (a) Logistic exposure-response relationships for the probability of high annoyance (p_{HA}) as a function of the outdoor equivalent continuous sound pressure level (L_{Aeq}) of wind turbine noise. SS relationships as marginal mean ($u=0$) and corresponding PA relationships established from a data subset of study I (Schäffer *et al.*, 2016) covering L_{Aeq} of 35–55 dB (solid lines) or 35–40 dB (dashed-dotted lines). (b) Relative differences between PA and SS $[=(PA - SS)/PA \times 100\%]$ established from the data subset covering L_{Aeq} of 35–55 dB (case 1), L_{Aeq} of 35–50 dB (case 2), L_{Aeq} of 35–45 dB (case 3), or L_{Aeq} of 35–40 dB (case 4), with the dotted horizontal line indicating the difference of 0. The parameter settings of the model curves correspond to the mean p_{HA} for different situations of amplitude modulation (no, with periodic and random AM) and mean sequence number = 15.5 of the stimuli during the experiments. The models and their differences are only displayed in the L_{Aeq} range for which they were established.

variance, the differences between the SS and PA curves are similar for cases 1 and 2 (similar variances), but smaller for cases 3 and 4 (smaller variances, thus larger PA parameters). This suggests that the two approaches only yield stable results (including a stable estimation of the random intercept variance in the case of SS) if the underlying data cover a sufficiently large range of observed relative frequencies of the binary variable. It further shows that above reason (ii) contributes to the differences between SS and PA by influencing the random intercept variance, i.e., by contributing to reason (iii). In the same context, further supporting the importance of reason (iii), it is interesting to note that the random intercept variance of the subset of the annoyance to wind turbine noise discussed here, with a value of 9.688 (above case 1) is distinctly larger than of the complete data set (4.996; Table I), and that, accordingly, the differences between the SS and PA curves are larger for the subset (Fig. 5) than for the complete data set (Fig. 2). Finally, for the above discussed data subset also the intersection point of (extrapolated) PA and SS curves decreases from $p_{HA}=0.47$ (case 1) to $p_{HA}=0.34$ (case 4). Thus, the SS and PA approaches may be expected to yield distinctly different curves if the underlying data cover a large range of observed relative frequencies of the binary variable and thus the SS model has a large random intercept variance (here, study I), and vice versa (studies II and III). Whether this conclusion is generally true, however, needs to be verified on a theoretical basis, which is beyond the scope of this paper.

5. Acoustical insights gained from the logistic regression analyses

The models established by the SS and/or PA modeling approaches (Tables I–III) reveal which parameters may influence noise reactions and how. These insights were discussed in detail in the original publications (Basner *et al.*, 2006; Brink *et al.*, 2011; Schäffer *et al.*, 2016).

In addition, the present systematic re-analysis of the original data sets with the same software yields further insights into the variation of the noise effects between individuals. The random intercept variance was found to be substantially larger in study I with ~ 5.0 (Table I) than in studies II and III with <0.4 (Tables II and III).

The random intercept variation is illustrated in more detail in Fig. 6. The random intercepts of study I [Fig. 6(a)] scatter strongly, covering values of -4.7 to $+6.0$. The extreme values correspond to p_{HA} differing by 0.49 – 0.99 in the L_{Aeq} range of 35 – 60 dB, or to a shift on the abscissa (L_{Aeq}) between the two SS curves of 27 dB. For study II [Fig. 6(b)], the random intercepts scatter distinctly less. Here, the extreme values of -0.7 to $+0.7$ correspond to moderate differences in p_{AWR} of 0.10 – 0.20 in the $L_{AS,max}$ range of 25 – 75 dB, but, due to the flat curve progression, to a large shift on the abscissa ($L_{AS,max}$) of 74 dB between the two SS curves. Finally, for study III [Fig. 6(c)], the extreme values of -0.7 to $+1.2$ correspond to differences in p_{AWR} of 0.10 – 0.43 in the $L_{AF,max}$ range of 25 – 75 dB, and to a shift on the abscissa ($L_{AF,max}$) of 32 dB between the two SS

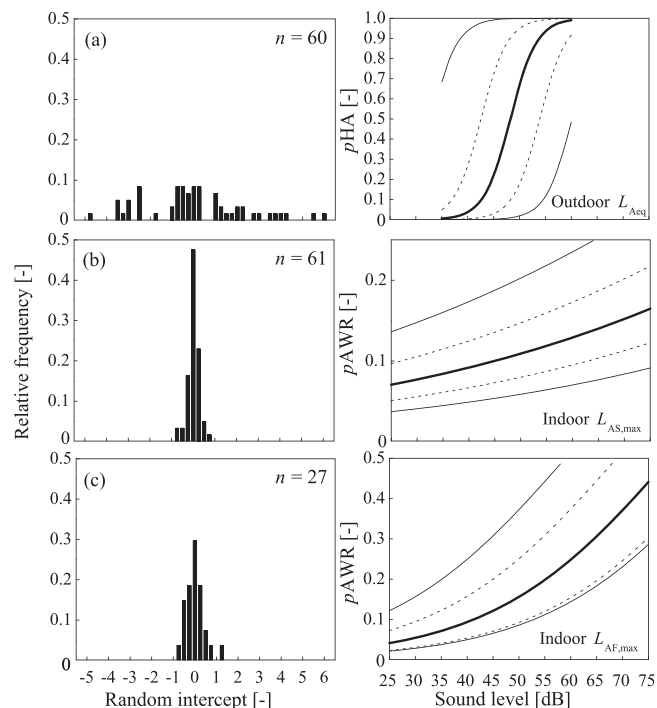


FIG. 6. Histograms (bin width = 0.25) showing the relative frequencies of the random intercepts (left) and resulting individual SS exposure-response relationships for the maximum and minimum value of the intercepts (solid thin lines), for the intercept $u=0$ (solid bold lines, marginal mean SS curves of Figs. 2–4) and for ± 1 standard deviation of the intercepts (dashed lines; right), for (a) study I, (b) study II, and (c) study III. n indicates the number of participants per study. Note the different scales of the probability ranges and the different sound level metrics between the studies in the right graphs.

curves. Thus, the random intercepts and their variance, both applied to the logit link function (cf. Sec. II), are not directly interpretable and comparable between the studies, but their meaning becomes apparent when shown in exposure-response relationships.

Figure 6 further discloses that the variation between individual SS curves is substantially larger than the differences between the SS and PA curves (Figs. 2–4). Analyzing the variation between subjects is thus reasonable. The present analysis might be refined by including a random slope, e.g., by modeling the individual dependences (curve progressions) on the sound pressure level, but this is beyond the scope of this paper.

IV. DISCUSSION

In this paper, two statistical modeling approaches and their application to establish PA or SS exposure-response relationships from repeated binary observations were discussed, with a specific focus on noise effects research. Relations between the modeling approaches and possible differences in results were estimated by systematic re-analysis of original data sets from three previously published noise effect studies (studies I–III). The authors are not aware of any other study that systematically applied and compared the two modeling approaches in the field of noise effects research.

A. Practical implications for noise effect studies

The re-analysis revealed that, depending on the underlying data set and range of observed relative frequencies based on which the models were established, the two approaches may yield very similar or very different exposure-response relationships. Thus, one needs to be aware which approach, PA or SS, is more adequate for the aspired research question or application (Sec. II). Depending on the underlying data set, the choice may be crucial, and an inappropriate approach may afflict the results by additional, though avoidable, uncertainty.

Further, the re-analysis highlights that the design of field and laboratory studies, in particular the coverage of the noise exposure and possibly other indicators and thus the range of observed relative frequencies of the binary variable, has a crucial influence on the established exposure-response relationships.

When applying existing exposure-response relationships to forecast either noise effects in the population or individuals' responses, it is therefore advisable to check if the relationships were derived from repeated binary observations (as opposed to independent binary data) and if yes, which statistical model was used to do so. As an example, PA relationships, including those derived from analysis of non-nested and non-hierarchical data, may be directly used in risk assessment on the population level, while SS relationships are less straightforward to use for this purpose. Instead of using an SS relationship, one should preferably establish a PA relationship from the original data. If the data are unavailable, one should consider converting the SS into a PA relationship as proposed by Zeger *et al.* (1988) and Hu

et al. (1998), particularly if the SS relationship was established from data covering a large range of observed relative frequencies. In fact, such conversion worked well for studies I–III, yielding results that closely correspond to the PA relationships established from the original data. However, it is still only an approximation, and directly determining the PA parameters from the original data is preferred. If neither is possible and the SS relationship is being used, this should be discussed as a limitation and source of uncertainty of the application. Alternatively, an interpretation of the results should be given, namely, that the effects on an average subject are being estimated rather than the effects on the population. Similar care must be taken if existing PA relationships are to be used to estimate effects on individuals, where SS relationships are more appropriate to use.

Analogous considerations apply to meta-analyses if exposure-response relationships from repeated binary observations are included. Generally, if available, the underlying data sets of the studies should be used to derive new exposure-response relationships. If not available, the averaging procedure (corresponding to SS or PA) of the individual relationships will depend on the research question and available relationships (see, e.g., Sec. 6.5 in Pedersen, 2007).

Finally, the SS approach has the advantage that it allows quantifying the variation of noise effects between individuals. This yields insights into individual responses to noise that remain undisclosed by the PA approach. Future research could focus on this aspect more systematically to enhance the quantitative understanding of reasons for variations among individuals and communities, e.g., by inclusion of individual and collective characteristics in the random effect analysis.

B. On past and future applications of exposure-response relationships

The relationships for *additional* awakening reactions (cf. Sec. III A 3) established in studies II and III (Basner *et al.*, 2006; Brink *et al.*, 2011) are examples of existing SS curves applied in several subsequent studies. The SS curve of study II was applied to establish night-time noise protection strategies around an airport (Basner *et al.*, 2005; Basner *et al.*, 2006). As the focus was on the effect on individuals and on preventive purposes, applying the SS approach is adequate. Besides, however, the SS relationship was also used in risk assessment studies to estimate aircraft noise effects on sleep of the population around airports (Basner *et al.*, 2010; Brink *et al.*, 2010; Schäffer *et al.*, 2012; Tétreault *et al.*, 2012). Here, the focus is on the response of the *population*, and a PA relationship would have been more appropriate to apply. The SS curve of study III was applied to estimate the impact of church bell ringing during night-time on the population (Brink *et al.*, 2012; Omlin and Brink, 2013). Also here, the application of a PA relationship would have been more appropriate. This methodological flaw in the latter risk assessment studies, however, is not very precarious in these instances. The present analysis revealed that the bias is rather small for studies II and III, as the PA and SS relationships are quite similar. Nevertheless, applying the relationships according to

their original purpose would be preferable, in particular, as in other cases (study I; Schäffer *et al.*, 2016) the PA and SS relationships may yield substantially different results.

With the above discussion in mind, the following question may be raised: *Are there instances when the application of one or the other modeling approach beyond its original purpose is justified?* The authors' opinion is that there are. While the PA approach estimates the mean response of individuals in a given population, the SS approach yields an estimate of the average individual in the population (mean SS curve with a random intercept $u = 0$) or of a particularly sensitive or resilient individual (with a certain random intercept value). Both estimates (SS and PA) can be used to represent a certain population, but the represented populations will differ. Similar to the mean, the PA estimate is affected by a few sensitive as well as resilient individuals (Fig. 1), and it hence truly reflects the variation within a population. Similar to the median (for random intercept $u = 0$), in contrast, the SS estimate is less affected by such "extreme" individuals. The median and the mean are both measures of central tendency, but differ in their statistical properties. Thus, when applying SS or PA beyond their original scope, this should be justified in light of the theory (Sec. II) and the intention of the desired application. An example is the application of SS to establish a protection concept for a particularly sensitive population group, as discussed above. In conclusion, while the PA approach should be used in the context discussed in this paper, the SS approach may be equally appropriate in specific other situations.

V. CONCLUDING REMARKS

In this paper, two statistical modeling approaches and their application to establish PA or SS noise exposure-response relationships from repeated binary observations were discussed. Systematic re-analysis of original data from three noise effect studies revealed that choosing an appropriate approach for the aspired research question may be crucial. Further, the SS approach has the advantage that the variation of noise effects between individuals is quantified, which is not possible with the PA approach. To date, many details have been discussed and accounted for in calculations to enhance the precision of noise-effect predictions, such as the effects of slope of rise, event order, event duration, or noise-free interval between noise events on sleep (e.g., Brink *et al.*, 2008a; Marks *et al.*, 2008), or the methodology of noise calculations (e.g., Boivin and Savard, 2013). The statistical models used to establish the exposure-response relationships, in contrast, have not been paid sufficient attention so far. It would be desirable to consider this aspect in future noise effect studies more thoroughly, be it in establishing new exposure-response curves or in applying existing curves in subsequent analyses. This would help avoiding one important source of uncertainty, or at least sensitize for this issue. The present paper is a contribution to this topic.

ACKNOWLEDGMENTS

This study (re-analysis of the data sets of studies I–III) was partly funded by the Swiss Federal Office for the

Environment (Assignment No. 5211.00723.100.02). The authors declare no conflicts of interest regarding the work presented in this paper.

- Anderson, D. A., and Aitkin, M. (1985). "Variance component models with binary response: Inviewer variability," *J. R. Statist. Soc. Ser. B Stat. Methodol.* **47**, 203–210.
- Basner, M. (2009). "Validity of aircraft noise induced awakening predictions," *Noise Control Eng. J.* **57**, 524–535.
- Basner, M., and Brink, M. (2013). "Sample size estimation for field studies on the effects of aircraft noise on sleep," *Appl. Acoust.* **74**, 812–817.
- Basner, M., Isermann, U., and Samel, A. (2005). "Die Umsetzung der DLR-Studie in einer lärmmedizinischen Beurteilung für ein Nachtschutzkonzept" ("The application of the DLR-study for a medical evaluation of a protective concept on adverse effects of nocturnal aircraft noise"), *Z. Lärmbekämpf.* **52**, 109–123.
- Basner, M., Müller, U., and Griefahn, B. (2010). "Practical guidance for risk assessment of traffic noise effects on sleep," *Appl. Acoust.* **71**, 518–522.
- Basner, M., Samel, A., and Isermann, U. (2006). "Aircraft noise effects on sleep: Application of the results of a large polysomnographic field study," *J. Acoust. Soc. Am.* **119**, 2772–2784.
- Boivin, D. B., and Savard, J. (2013). "Commentary on the Tétéreault *et al.* 2012 article 'Risk assessment of aircraft noise on sleep in Montreal'," *Can. J. Public Health* **104**, e275.
- Brink, M., Basner, M., Omlin, S., and Pieren, R. (2012). "Lärm von Kirchenglocken: Stören Kirchenglocken unseren Schlaf? Eine Untersuchung der Auswirkungen nächtlicher Kirchenglockenschläge auf die Bevölkerung" ("Church bell noise: Do church bells disturb sleep?—A study of the effects of nightly church bell ringing on the population"), *Lärmbekämpf.* **7**, 29–41.
- Brink, M., Lercher, P., Eisenmann, A., and Schierz, C. (2008a). "Influence of slope of rise and event order of aircraft noise events on high resolution actimetry parameters," *Somnologie* **12**, 118–128.
- Brink, M., Omlin, S., Müller, C., Pieren, R., and Basner, M. (2011). "An event-related analysis of awakening reactions due to nocturnal church bell noise," *Sci. Total Environ.* **409**, 5210–5220.
- Brink, M., Schreckenberg, D., Thomann, G., and Basner, M. (2010). "Aircraft noise indexes for effect oriented noise assessment," *Acta Acust. Acust.* **96**, 1012–1025.
- Brink, M., Wirth, K. E., Schierz, C., Thomann, G., and Bauer, G. (2008b). "Annoyance responses to stable and changing aircraft noise exposure," *J. Acoust. Soc. Am.* **124**, 2930–2941.
- de Hollander, A. E. M., Melse, J. M., Lebet, E., and Kramers, P. G. N. (1999). "An aggregate public health indicator to represent the impact of multiple environmental exposures," *Epidemiology* **10**, 606–617.
- EAA (2010). Good Practice Guide on Noise Exposure and Potential Health Effects, EEA Technical Report No. 11/2010, European Environment Agency (EAA), Copenhagen, Denmark, available at <http://www.eea.europa.eu/publications/good-practice-guide-on-noise> (Last viewed 3/13/2017).
- European Union (2002). Directive 2002/49/EC of the European Parliament and of the Council of 25 June 2002 Relating to the Assessment and Management of Environmental Noise (European Union, Bruxelles, Belgium), available <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2002:189:0012:0025:EN:PDF> (Last viewed 3/13/2017).
- Guski, R., and Bosshardt, H.-G. (1992). "Gibt es eine 'unbeeinflusste' Lästigkeit?," ("Does 'unbiased annoyance' exist?"), *Z. Lärmbekämpf.* **39**, 67–74.
- Hanley, J. A., Negassa, A., Edwardes, M. D. deB., and Forrester, J. E. (2003). "Statistical analysis of correlated data using generalized estimating equations: An orientation," *Am. J. Epidemiol.* **157**, 364–375.
- Horton, N. J., and Lipsitz, S. R. (1999). "Review of software to fit generalized estimating equation regression models," *Am. Stat.* **53**, 160–169.
- Hosmer, D. W., and Lemeshow, S. (2000). *Applied Logistic Regression*, 2nd ed. (Wiley, Toronto, Canada), pp. 1–375.
- Hu, F. B., Goldberg, J., Hedeker, D., Flay, B. R., and Pentz, M. A. (1998). "Comparison of population-averaged and subject-specific approaches for analyzing repeated binary outcomes," *Am. J. Epidemiol.* **147**, 694–703.
- Jang, M. J. (2011). "Working correlation selection in generalized estimating equations," Ph.D. dissertation, University of Iowa, Iowa City, IA, available at <http://ir.uiowa.edu/etd/2719> (Last viewed 3/13/2017).
- Janssen, S. A., Vos, H., Eisses, A. R., and Pedersen, E. (2011). "A comparison between exposure-response relationships for wind turbine annoyance and annoyance due to other noise sources," *J. Acoust. Soc. Am.* **130**, 3746–3753.

- Liang, K.-Y., and Zeger, S. L. (1986). "Longitudinal data analysis using generalized linear models," *Biometrika* **73**, 13–22.
- Marks, A., Griefahn, B., and Basner, M. (2008). "Event-related awakenings caused by nocturnal transportation noise," *Noise Control Eng. J.* **56**, 52–62.
- Michaud, D. S., Feder, K., Keith, S. E., Voicescu, S. A., Marro, L., Than, J., Guay, M., Denning, A., Bower, T., Villeneuve, P. J., Russell, E., Koren, G., and van den Berg, F. (2016a). "Self-reported and measured stress related responses associated with exposure to wind turbine noise," *J. Acoust. Soc. Am.* **139**, 1467–1479.
- Michaud, D. S., Keith, S. E., Feder, K., Voicescu, S. A., Marro, L., Than, J., Guay, M., Bower, T., Denning, A., Lavigne, E., Whelan, C., Janssen, S. A., Leroux, T., and van den Berg, F. (2016b). "Personal and situational variables associated with wind turbine noise annoyance," *J. Acoust. Soc. Am.* **139**, 1455–1466.
- Neuhaus, J. M., Kalbfleisch, J. D., and Hauck, W. W. (1991). "A comparison of cluster-specific and population-averaged approaches for analyzing correlated binary data," *Int. Stat. Rev.* **59**, 25–35.
- Omlin, S., and Brink, M. (2013). "Awakening effects of church bell noise: Geographical extrapolation of the results of a polysomnographic field study," *Noise Health* **15**, 332–341.
- Passchier-Vermeer, W. (2003). "Night-time noise events and awakening," TNO Inro Report 2003-32, TNO, Institute of Spatial Organization (Inro), Delft, The Netherlands, available at <http://repository.tudelft.nl/view/tno/uuid%3A979f4520-fd48-411a-9fa7-19fec924d25/> (Last viewed 0/13/2017).
- Passchier-Vermeer, W., Vos, H., Steenbekkers, J. H. M., van der Ploeg, F. D., and Groothuis-Oudshoorn, K. (2002). "Sleep disturbance and aircraft noise exposure. Exposure-effect relationships," TNO Report 2002.027, TNO, Division Public Health, Leiden, The Netherlands.
- Pedersen, T. H. (2007). "The 'Genlyd' noise annoyance model. Dose-response relationships modelled by logistic functions," Journal No. AV 1102/07, DELTA Danish Electronics, Light & Acoustics, Hørsholm, Denmark, available at http://assets.madebydelta.com/docs/share/Akustik/The_Genlyd_Noise_Annoyance_Model.pdf (Last viewed 3/13/2017).
- Pedersen, E., and Persson Waye, K. (2004). "Perception and annoyance due to wind turbine noise—A dose-response relationship," *J. Acoust. Soc. Am.* **116**, 3460–3470.
- Pedersen, E., van den Berg, F., Bakker, R., and Bouma, J. (2009). "Response to noise from modern wind farms in The Netherlands," *J. Acoust. Soc. Am.* **126**, 634–643.
- Schäffer, B., Schlittmeier, S. J., Pieren, R., Heutschi, K., Brink, M., Graf, R., and Hellbrück, J. (2016). "Short-term annoyance reactions to stationary and time-varying wind turbine and road traffic noise: A laboratory study," *J. Acoust. Soc. Am.* **139**, 2949–2963.
- Schäffer, B., Thomann, G., Huber, P., Brink, M., Plüss, S., and Hofmann, R. (2012). "Zurich Aircraft Noise Index: An index for the assessment and analysis of the effects of aircraft noise on the population," *Acta Acust. Acust.* **98**, 505–519.
- Stiratelli, R., Laird, N., and Ware, J. H. (1984). "Random-effects model for serial observations with binary response," *Biometrics* **40**, 961–971.
- Szmaragd, C., Clarke, P., and Steele, F. (2013). "Subject specific and population average models for binary longitudinal data: A tutorial," *Longit. Life Course Stud.* **4**, 147–165.
- Tétreault, L.-F., Plante, C., Perron, S., Goudreau, S., King, N., and Smargiassi, A. (2012). "Risk assessment of aircraft noise on sleep in Montreal," *Can. J. Public Health* **103**, e293–e296.
- Van Renterghem, T., Bockstael, A., De Weirt, V., and Botteldooren, D. (2013). "Annoyance, detection and recognition of wind turbine noise," *Sci. Total Environ.* **456-457**, 333–345.
- WHO (2009). Night Noise Guidelines for Europe, World Health Organization (WHO) Regional Office for Europe, Copenhagen Ø, Denmark, available at http://www.euro.who.int/__data/assets/pdf_file/0017/43316/E92845.pdf (Last viewed 3/13/2017).
- WHO (2011). Burden of Disease from Environmental Noise. Quantification of Healthy Life Years Lost in Europe, World Health Organization (WHO) Regional Office for Europe, Copenhagen Ø, Denmark, available at http://www.euro.who.int/__data/assets/pdf_file/0008/136466/e94888.pdf?ua=1 (Last viewed 3/13/2017).
- Zeger, S. L., Liang, K.-Y., and Albert, P. S. (1988). "Models for longitudinal data: A generalized estimating equation approach," *Biometrics* **44**, 1049–1060.
- Zhang, H., Lu, N., Feng, C., Thurston, S. W., Xia, Y., Zhu, L., and Tu, X. M. (2011). "On fitting generalized linear mixed-effects models for binary responses using different statistical packages," *Stat. Med.* **30**, 2562–2572.