On the application of two statistical approaches to establish noise exposure-response relationships from repeated binary observations

Beat Schäffer[1], Reto Pieren[2], Franco Mendolia[3], Mathias Basner[4], Mark Brink[5]

[1] Empa, Laboratory for Acoustics/Noise Control, Dübendorf, Switzerland (corresponding author)
[2] Empa, Laboratory for Acoustics/Noise Control, Dübendorf, Switzerland
[3] German Aerospace Center (DLR), Institute of Aerospace Medicine, Cologne, Germany
[4] University of Pennsylvania, Perelman School of Medicine, Philadelphia, PA
[5] Federal Office for the Environment, Noise and NIR Division, Berne, Switzerland

Corresponding author's e-mail address: beat.schaeffer@empa.ch

ABSTRACT

Noise exposure-response relationships for binary variables such as high annoyance or sleep disturbance are used to estimate the effects of noise on individuals or a population. Such relationships may be established from repeated binary observations with different statistical approaches. As the statistical approaches are inherently different and yield disparate results, it is crucial to decide which one to use. This aspect, however, was not always sufficiently considered in the past in noise effect studies. This paper gives an overview on two existing statistical approaches to establish noise exposure-response relationships from repeated binary observations, namely, a subject-specific and a population-averaged logistic regression analysis. With an example of a recent noise effect study, the potential magnitude of differences in results between the two approaches is estimated, reasons for the differences are disclosed, and possible implications for future studies are discussed.

INTRODUCTION

Noise exposure-response relationships are of importance to estimate the effects of noise on individuals or a population. In recent years, risk assessment of environmental noise, such as noise impacts on the population, became an important topic [1]. It is also required by the Environmental Noise Directive 2002/49/EC [2] to establish action plans. 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 noise effect(s). Besides risk assessment on the population level, the research focus may also be on responses of individuals to noise, for example in medical studies (e.g. [3]).

Exposure-response relationships for binary data may be derived from independent observations (non-nested and non-hierarchical data: one observation per subject) or from repeated 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 the relationships were established, their application in subsequent studies such as the estimation of noise effects on the population may be more or less straightforward and yield different results. While this critical point has been discussed in epidemiology, medicine or statistics (e.g. [4; 5; 6]), it was so far not sufficiently considered in noise effects research.

This paper gives an overview on two existing statistical approaches to establish noise exposure-response relationships from repeated binary observations, namely, a subject-specific and a population-averaged approach. With an example of a recent laboratory study [7], the potential magnitude of differences in results between the two approaches is estimated, reasons for the differences are disclosed, and possible implications for future studies are discussed.

STATISTICAL MODELLING APPROACHES: OVERVIEW

This section gives an overview of two common logistic regression modelling 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 are obtained from repeated observations of the same binary variable in subjects over different points in time, for example awakenings to noise events in the night (e.g. [3]), while independent observations are often obtained in field surveys, where each subject gives one single rating, for example on annoyance (e.g. [8]).

As long as the binary observations are independent, a “standard” binary logistic regression analysis may be applied [9], which yields population-averaged exposure-response relationships [4] that are directly applicable in risk assessment on the population level. Such a model for the jth observation can be written as

$$\operatorname{logit} p(Y_j = 1 \mid X_{jk}) = \beta_0 + \sum_k \beta_k X_{jk} \,, \tag{1}$$

where logit p = ln[p/(1–p)] is the logit link function [9] for the probability p of the response of the dependent binary variable Y_j to adapt the value of 1 given the predictor variables, β_0 is the intercept, and β_k are the regression parameters quantifying the effects of the k predictor variables X_{jk}. X_{jk} can be a categorical variable with a certain number of levels (e.g. different sleep stages [3]), a continuous variable (e.g. sound pressure level), or interactions accounting for deviations from the additive model.

Things are different for repeated binary observations. 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 (i) a generalized estimating equations (GEE) model [6; 10] or (ii) a mixed-effects logistic regression model, also referred to as random-effects logistic regression model [11; 12], which are “multi-level” or “hierarchical” models [13].

The GEE model for the jth observation of the ith subject can be written as

$$\operatorname{logit} p(Y_{ij} = 1 \mid X_{ijk}) = \beta_0^{\mathrm{PA}} + \sum_k \beta_k^{\mathrm{PA}} X_{ijk} \,, \tag{2}$$

where β_k are the regression parameters quantifying the effects of the k predictor variables X_{ijk} on the dependent variable Y_{ij}, the index PA indicates that the parameters represent population-averaged effects, and the other variables have the same notation as in Equation 1.

The mixed-effects logistic model for the *j*th observation of the *i*th subject can be written as

$$\text{logit}\ p(Y_{ij}=1\,|\,u_i, X_{ijk}) = \beta_0^{SS} + \sum_k \beta_k^{SS} X_{ijk} + u_i\ , \tag{3}$$

where the parameter u_i denotes a random effect, the index SS stands for subject-specific effects, and the other variables have the same notation as in Equation 2. For simplicity, the random effect is modelled here as a random intercept only, one for each subject. The more general case of a random intercept plus random slope(s) is not discussed here.

With Equations 1 to 3, the logistic function is defined as the inverse of the logit, as

$$p(Y=1) = 1/\left[1+\exp(-\text{logit}\ p)\right]. \tag{4}$$

Differences between the two latter approaches (Equations 2 and 3), their results and the interpretability of the results have been previously discussed in the literature about epidemiologic, medical and social sciences research [4; 5; 6; 14]. In short, the GEE model (Equation 2) is a population-averaged (PA) or marginal approach [5], while the mixed-effects logistic model (Equation 3) is a subject-specific (SS) or cluster-specific approach. The main difference is that SS describes the response of an average subject and PA the average population's response. The PA approach models a population-averaged binary variable (averaged across subjects of the population) by accounting for the correlation between the repeated observations with a certain working correlation matrix (or structure), such as exchangeable, autoregressive of order one, unstructured or independent (e.g. [15]). The SS approach, in contrast, introduces a random effect (i.e. random intercept) to account for the correlation between the repeated observations of the same subject. With the random intercept u_i, also the magnitude of variation between individuals is quantified.

The resulting estimates (i.e. absolute values) of the logistic regression model parameters of the SS approach, β^{SS}, are generally larger than those of the PA approach, β^{PA} [5], and the difference between β^{SS} and β^{PA} increase with increasing variation between individuals [6]. The SS logistic regression curves are therefore often steeper than the PA logistic regression curves, with the PA relationship 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 (cf. Figure 1). The intersection point of the SS curve of the mean subject (bold solid line in Figure 1, with a random intercept u = 0) and the PA curve (dash-dotted line) is at the value that yields a probability p = 0.5 (Figure 1), given that the random intercept follows a normal distribution [6].

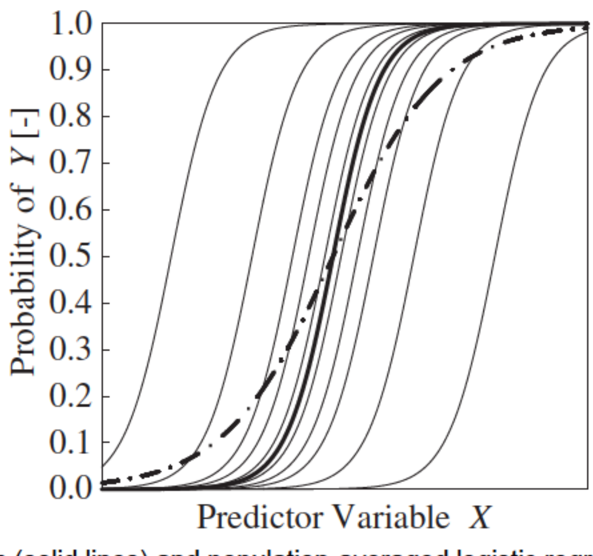

Figure 1: Subject-specific (solid lines) and population-averaged logistic regression (dash-dotted line). The bold solid line represents the mean subject with a random intercept u = 0.

Thus, the SS curve of the mean subject (with $u = 0$) usually predicts higher probabilities of an effect than the PA curve at values of the predictor variable above the intersection point of the two curves and *vice versa* (Figure 1). The differences between the two modelling approaches are due to the non-linearity of the logistic regression (e.g. [14]): The average of the non-linear SS relationship is not the same as the non-linear PA relationship of the average.

If the random intercept follows a normal distribution, the parameters of the SS model may be analytically converted (or marginalized) to those of the PA model with an exchangeable working correlation matrix, i.e. assuming the same correlation between all observations per subject (e.g. [15]), using the relationship [4; 6]

$$\beta^{\mathrm{PA}} = \beta^{\mathrm{SS}} / \sqrt{1 + \left(\frac{16 \times \sqrt{3}}{15 \times \pi} \right)^2 \times \sigma_u^2} \cong \beta^{\mathrm{SS}} / \sqrt{1 + 0.346 \times \sigma_u^2}\ , \tag{5}$$

where β^{SS} and β^{PA} are the SS and PA parameters of the logistic regression models in Equations 2 and 3 and σ_u^2 is the variance of the random intercept u. In Equation 5, the SS parameter, β^{SS}, is always larger than the PA parameter, β^{PA}, because σ_u^2 is always positive. (A similar approximation is given in [14].) The PA relationship based on Equation 5 will intersect the SS relationship at a probability $p = 0.5$. While Equation 5 is valid under the above mentioned assumption of the random intercept following a normal distribution, it is only an approximation for "real" data, and directly determining the PA parameters from the original data is recommended where possible instead of using Equation 5. Note that in contrast to the above conversion of SS to PA parameters, a conversion of PA to SS parameters is not possible [16].

Due to the inherent differences in the models, it is crucial to choose the appropriate approach for a given research question. PA models are preferred for epidemiological studies [6] or for (noise) risk assessment on the population, while SS models are suitable if the focus is on individuals [4] (e.g. patients in clinical studies), or if the target population differs from the study sample used to estimate the model, e.g. when a protection concept for a particularly sensitive population group is established [17]. While SS models, by accounting for individual responses, generally yield "stronger" effects (larger absolute β^{SS} estimates in Equation 3) than PA models (β^{PA} in Equation 2) and also higher agreement between predicted and observed individual values (e.g. in terms of classification tables summarizing the rate of correct predictions of individual observations [9]), one should not choose the SS model for these reasons, because the two approaches have different aims, as explained above.

METHODS

The above introduced modelling approaches are applied to the original data set of repeated binary observations of a study from literature [7]. The resulting logistic regression curves are compared to each other, and differences between the model results are discussed.

Overview of the study providing the data set for the re-analysis

Schäffer et al. [7] established an exposure-response relationship for the probability of the binary variable "high annoyance" to take a value of 1 (pHA) due to wind turbine and road traffic noise. The probability pHA was modelled as a function of the source type (Source: wind turbines, road traffic), A-weighted equivalent continuous sound pressure level (L_{Aeq}), amplitude modulation (AM: without, random, periodic), their interactions Source × AM, Source × L_{Aeq} and AM × L_{Aeq}, as well as the Sequence Number, i.e. the playback number with which the stimuli had been played to the participants. The PA model was chosen as the focus of the study was to estimate the mean *p*HA in the population. The underlying data set was obtained from repeated binary observations in a laboratory experiment with 60 subjects who were exposed to stimuli with an L_{Aeq} of 35–60 dB.

Statistical re-analysis of the data set

To compare possible results of the PA and SS modelling approaches, the exposure-response relationships of both approaches were established based on the original data set of the above study [7]. For the analysis, the procedures GENLIN (for PA) and GENLINMIXED (for SS) of the software IBM SPSS Statistics Version 22 was used. (Note that other software packages such as SAS or R provide similar procedures as well.) For comparative purposes, the PA relationship was also estimated from the SS relationship using Equation 5.

The PA relationship for *p*HA had already been established and was taken from Table II in [7]. 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. The SS relationship established in this study considers the same explanatory variables, but accounts 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.

RESULTS

Figure 2 shows the logistic regression models for *p*HA for wind turbine noise. For comparative purposes, also the mean observed relative frequencies of high annoyance are displayed. Note that these data do not allow to decide which relationship is “more appropriate” to represent the observations, because the appropriate model (SS or PA) is given by the research goal and not merely by the degree of agreement with observed data. Also, the regression curves adjust for all predictor variables (besides the L_{Aeq}) as well as for repeated observations, while the averaging does not. The latter aspect, however, is of minor importance here due to the full factorial design of [7]: 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 observed means, as these varied between dB classes (see [7]).

The mean observations cover a wide range of relative frequencies with values from 0.08–0.82 (Figure 2). With the model parameters set to the mean of the predictor variables AM and Sequence Number during the experiments, the PA relationship closely represents the mean values of the observations. The SS relationship, in contrast, is substantially steeper than the PA relationship, which was expected (Figure 1). Also, the SS model represents the individual ratings more closely than the PA approach: The PA model has a rate of correct annoyance predictions of 82% (classification table), while the SS approach has rate of 91%. Further, the SS approach yields mostly larger absolute parameter estimates (β^{SS} vs. β^{PA}, not shown), as was expected according to theory. Both models, however, identify the same significant effects, namely, L_{Aeq} ($p < 0.01$), Sequence Number ($p < 0.01$), interaction Source × AM ($p < 0.02$) and in tendency also Source ($p < 0.06$). The SS and PA relationships intersect at a L_{Aeq} of 48 dB and corresponding *p*HA of ~0.47, which closely corresponds to the expected value of 0.5 (Figure 1). Below this level, the PA relationship predicts larger *p*HA values than the SS relationship and *vice versa* above.

The PA model parameters established with the PA approach (Equation 2) may be reliably estimated by Equation 5 from the SS parameters (details not shown). Accordingly, the converted and directly modelled PA curves are almost identical (Figure 2).

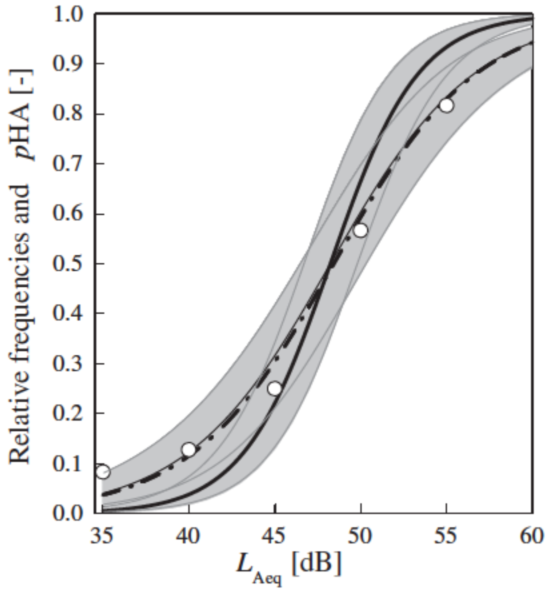

Figure 2: Mean observed relative frequencies (circles) of high annoyance and logistic exposure-response relationships for *p*HA as a function of the L_{Aeq} of wind turbine noise (re-analyzed data of [7]): SS relationship as marginal mean with $u = 0$ (bold solid line) and corresponding PA relationship (bold dash-dotted line) with 95% confidence intervals (gray areas), and approximated PA relationship using Equation 5 (thin solid line). The relationships correspond to the mean *p*HA for different situations of AM and mean Sequence Number during the experiments.

DISCUSSION

In this paper, two statistical modelling approaches (PA or SS) and their application to establish noise exposure-response relationships from repeated binary observations were discussed and exemplarily applied to an original data set of a laboratory study from literature [7].

The re-analysis revealed that, for this data set, the theoretical relation between the SS and PA approach, with characteristics such as the intersection point of the curves at a probability of 0.5 or the possibility of marginalization with Equation 5 apply, and that the differences in results between SS and PA may be large (Figure 2). However, also other relations between the PA and SS models are conceivable for other data sets, depending on (i) whether the assumption of a normal distribution of the random intercept (SS) is fulfilled, and/or (ii) the range of observed relative frequencies based on which the logistic regression analyses are performed, and/or (iii) the magnitude of variation between individuals (disparate SS and PA curves for large variations: Figure 1) and/or (iv) model complexity or lack of significance of certain model parameters.

For the present study, reason (i) was not found to apply: Visual inspection of the random intercepts with residual plots did not reveal obvious deviation from normality. Reason (ii), in contrast, is possibly the main reason for the large differences between the SS and PA curves. In fact, the data of [7] covers a wide range of observed relative frequencies of up to 0.82 (Figure 2). It was found that if the models are established on a limited L_{Aeq} range of 35–40 dB only instead of the full range of 35–55 dB, the random intercept variance and, accordingly, the model differences are distinctly smaller (details not shown). Reason (ii) thus contributes to the differences between SS and PA by influencing the random intercept variance, i.e. by contributing to reason (iii). But reason (iii) might also distinctly contribute to the differences on its own, while reason (iv) is not expected to account for the differences. This, however, cannot be veryfied without re-analysis of additional data sets from other studies.

The above results suggest that the SS and PA approaches may be expected to yield distinctly different curves if the observations cover a large range of relative frequencies and thus the SS model has a large random intercept variance, and *vice versa*. Thus, the underlying data used to establish an SS or PA model should cover a sufficiently large range of observed relative frequencies of the binary variable. The generalizability of these observations, however, still needs to be tested. In particular, re-analyses of additional data sets from other noise effect studies are desirable to estimate the nature and range of possible differences between the two modelling approaches. This is the focus of ongoing research by the authors.

The insights of the present study have practical implications for future noise effect studies:

- One needs to decide which approach, PA or SS, is more adequate for the aspired research question or application. Depending on the underlying data set, the choice may be crucial.
- The design of noise effect studies, in particular the coverage of noise and possibly other indicators and thus the range of observed relative frequencies of the binary variable, strongly influences the established exposure-response relationships.
- When applying existing exposure-response relationships to forecast either noise effects in the population or individuals' responses, one should check if the relationships were derived from repeated binary data (as opposed to independent binary data) and if yes, which statistical model was used to do so. Analogous considerations apply to meta-analyses if exposure-response relationships from repeated binary observations are included.

Regarding the last point, PA relationships, including those derived from independent observations, 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 or, if the data are unavailable, consider converting the SS into a PA relationship using Equation 5 as shown above, particularly if the SS relationship was established from data covering a large range of observed relative frequencies. If neither is possible and the SS relationship is being used, this should be discussed as a limitation and source of uncertainty, or, alternatively, adequately interpreted, namely, that the effects on a mean 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. In the past, the above aspects were not always given sufficient consideration.

CONCLUSIONS

In this paper, two statistical modelling approaches and their application to establish population-averaged or subject-specific noise exposure-response relationships from repeated binary observations were discussed. Re-analysis of an original data set from a recent noise effect study revealed that the choice of an appropriate modelling approach for the aspired research question may be crucial. It would be desirable to consider this aspect in future noise effect studies more thoroughly. The present paper is a contribution to this topic.

Acknowledgements

This study was partly funded by the Swiss Federal Office for the Environment (Assignment No. 5211.00723.100.02).

REFERENCES

[1] EAA. (2010). *Good Practice Guide on Noise Exposure and Potential Health Effects. EEA Technical report No 11/2010*. European Environment Agency (EAA), Copenhagen, Denmark. URL: http://www.eea.europa.eu/publications/good-practice-guide-on-noise.

[2] 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. URL: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2002:189:0012:0025:EN:PDF.

[3] Basner, M., Samel, A., & Isermann, U. (2006). Aircraft noise effects on sleep: application of the results of a large polysomnographic field study. *Journal of the Acoustical Society of America, 119*, 2772-2784.

[4] Hu, F. B., Goldberg, J., Hedeker, D., Flay, B. R., & Pentz, M. A. (1998). Comparison of population-averaged and subject-specific approaches for analyzing repeated binary outcomes. *American Journal of Epidemiology, 147*, 694-703.

[5] Neuhaus, J. M., Kalbfleisch, J. D., & Hauck, W. W. (1991). A comparison of cluster-specific and population-averaged approaches for analyzing correlated binary data. *International Statistical Review, 59*, 25-35.

[6] Zeger, S. L., Liang, K.-Y., & Albert, P. S. (1988). Models for longitudinal data: a generalized estimating equation approach. *Biometrics, 44*, 1049-1060.

[7] Schäffer, B., Schlittmeier, S. J., Pieren, R., Heutschi, K., Brink, M., Graf, R., et al. (2016). Short-term annoyance reactions to stationary and time-varying wind turbine and road traffic noise: a laboratory study. *Journal of the Acoustical Society of America, 139*, 2949–2963.

[8] Janssen, S. A., Vos, H., Eisses, A. R., & Pedersen, E. (2011). A comparison between exposure-response relationships for wind turbine annoyance and annoyance due to other noise sources. *Journal of the Acoustical Society of America, 130*, 3746-3753.

[9] Hosmer, D. W., & Lemeshow, S. (2000). *Applied Logistic Regression* (2nd ed.). Toronto, Canada: John Wiley & Sons, Inc.

[10] Liang, K.-Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika, 73*, 13-22.

[11] Anderson, D. A., & Aitkin, M. (1985). Variance component models with binary response: inverviewer variability. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 47*, 203-210.

[12] Stiratelli, R., Laird, N., & Ware, J. H. (1984). Random-effects model for serial observations with binary response. *Biometrics, 40*, 961-971.

[13] Goldstein, H. (2011). *Multilevel Statistical Models* (4th ed.). Chichester, UK: John Wiley & Sons, Ltd.

[14] Szmaragd, C., Clarke, P., & Steele, F. (2013). Subject specific and population average models for binary longitudinal data: a tutorial. *Longitudinal and life course studies, 4*, 147-165.

[15] Jang, M. J. (2011). *Working correlation selection in generalized estimating equations.* (Ph.D. dissertation), University of Iowa, Iowa City, IA. URL: http://ir.uiowa.edu/etd/2719.

[16] Lee, Y., & Nelder, J. A. (2004). Conditional and marginal models: another view. *Statistical Science, 19*, 219-238.

[17] Basner, M. (2009). Validity of aircraft noise induced awakening predictions. *Noise Control Engineering Journal, 57*, 524-535.