A review of explained variance in exposure-annoyance relationships in noise annoyance surveys

Mark Brink[1]

[1] Federal Office for the Environment, Noise/NIR Division, CH-3003 Bern, SWITZERLAND
mark.brink@bafu.admin.ch

ABSTRACT

The troubles with small degrees of variance explained by noise exposure in noise annoyance surveys was recognized long ago. Typically, in bivariate models, not more than about 10-20 percent of the variation of individual annoyance ratings can be explained by (usually Leq-based) exposure measures. This paper aims at taking inventory of the level of achieved explained variance of the exposure-annoyance relationships as derived from original datasets from a handful of recent noise annoyance surveys. The probability distributions of R-squared are estimated from subsamples of the original survey data by applying a simple and straightforward bootstrap approach. Results from this exercise revealed that explained variance of bivariate exposure-effect models amounts to not more than R-squared values between 0.05 and 0.25, if individual responses are used. The further analysis aimed at elucidating whether R-squared is associated with the noise source, type of survey (postal questionnaire, telephone interview, face-to-face), or the characteristics of the annoyance scale that was used to measure annoyance responses.

INTRODUCTION

In the August 2011 issue of the Journal of the Acoustical Society of America, Meyer expressed his concern, that current noise annoyance research applies rather lenient criteria regarding the required level of explained variance in exposure-effect models that aim at informing policy decisions, particularly, in the scope of defining exposure limit values (Meyer, 2011). As pertinent information is missing about what can practically be expected in an empirical noise annoyance survey in terms of explained variance, the present exercise aimed at taking inventory of the level of R-squared of the exposure-annoyance relationships as derived from a handful of datasets of recent noise annoyance surveys.

R-squared in the context of socio-acoustic surveys

In their reviews on the role of nonacoustical variables, both Fields (1993) and Job(1988), mostly considering studies from the 70ies and 80ies, came to the conclusion, that noise exposure measures typically explain not more than about 20 percent of the individual variation of annoyance. In the linear regression context, explained variance is usually expressed in the R-squared statistic, i.e. the squared correlation

between observed and predicted values in a given sample. R-squared is a common goodness of fit measure and can take values between 0 and 1. Job in 1988 concluded from his review (Job, 1988) of 39 empirical studies of noise annoyance, that the average R-squared between noise exposure and noise annoyance for group data is 0.67 and for individual data 0.18, which is - given common standards - very low.

In the statistical model building process, high values of R-squared are desired, as R-squared is a relatively straightforward estimate of how good the model will predict the level of the dependent variable at different levels of the independent variable, in other terms: how good the exposure to noise predicts annoyance or other effects that are in the focus of a particular study. For studies into exposure-annoyance relationships of noise, the achieved levels of R-squared are of practical concern for most researchers (and policy makers alike) – the lower R-squared turns out, the higher the risk of coincidental and, therefore, unjustified conclusions.

R-squared usually grows with the adjunction of predictors, and accounting for more that just exposure measures (e.g. Leq) as predictors, will almost certainly result in better predictions. However, in socio-acoustic annoyance studies the researcher usually tries to explain as much of the variance of annoyance by characteristics of the exposure alone. As political demands in the context of noise regulation call for parsimony, the "choice" of usable predictor variables is often restricted to those that refer to acoustical properties of a noise source. This can be illustrated as follows: It is well known fact, that e.g. house owners show greater noise annoyance than renters, and that noise-sensitive persons are more easily annoyed by a given level of exposure than the insensitives. While any multivariate model that, in addition to the exposure, accounts for such predictors, will enhance our understanding of the genesis of annoyance reactions and certainly reduce the fraction of unexplained variance, it is difficult to use these predictors in forecasts, e.g. in environmental impact assessments. This is simply the case because most usually, exposure measures are available on the level of spatial grid points, whereas data on the level of individuals are not. Thus, policy makers usually demand estimates of effects that rely on measured or calculated exposure alone.

The level of achievable explained variance in a socio-acoustic survey is depending on various factors: It is a general property of regression, that model fit characteristics are strongly determined by (1) the range and variability of observations in the sample and (2) the accuracy of the measurement of both the independent and the dependent variable. These simple facts may already explain a great deal of differences of explained variance that are encountered between different studies. E.g., correlation coefficients are larger, the larger the range of noise exposure is. Hence, a study that e.g. investigates annoyance reactions in the 40-50 dB exposure range, will most probably find a weaker association with annoyance (a smaller R-squared) than a study whose exposure range is between 30 and 70 dB. Accuracy of exposure and annoyance measurements play an important role too for achieving high explained variance, as misclassification biases 'blur' the true exposure-response relationship, i.e. lead to an underestimation of the strength of the relationship. In addition to these more general effects, a range of context and scale as well as psychometric effects such as type of interview (postal, telephone, face-to-face), survey context, season, type of annoyance questions, question placement, and order of presentation of answer alternatives as well as the measurement scale to assess annoyance are proba-

bly relevant factors that account for the variability of explained variance one typically encounters in socio-acoustic annoyance studies (Brink, 2013).

Aims and goals

In the present exercise we aim at, in a manner of speaking, taking inventory of achieved levels of explained variance (with individual response data) in recent noise annoyance surveys. Furthermore, we try to elucidate whether R-squared of exposure-annoyance relationships is associated with (1) noise source, (2) exposure measure (LDay, Lden, or Ldn), (3) type of interview (postal questionnaire, telephone interview, face-to-face) and (4) the annoyance scale (4/5-point verbal vs. 11-point numerical) that was used to measure annoyance responses.

METHODS

Selection of surveys

Potential surveys/studies to include in the exercise were identified by directly asking colleagues in the research community to make available their original exposure-annoyance data. In addition to these studies (that all appeared in the peer-reviewed literature), we included a few of our own (not yet published) surveys and datasets. Hence as can be seen, the present selection does not represent any systematic approach to data collection, but well serves the explorative purpose of the exercise. A minimum sample size of 100 respondents was defined as a prerequisite, as well as the application of either the 5-pt or 11-pt ICBEN/ISO (Fields et al., 2001) or similar annoyance scales in the survey. Furthermore, only studies having made the effort of rigid exposure assessment either by individual measurements according to established protocols, or exposure calculations were included. Most of the datasets we obtained included several exposure measures for sometimes more than one source and/or responses on two annoyance scales (e.g. numerical and verbal scale). The bootstrap resamples were based upon combinations of noise source, available exposure measures (either Ldn, Lden, or LDay, but omitting LNight), and type of interview. In longitudinal studies with two or more waves, each wave was treated as a stand-alone case. Table 1 lists the surveys/studies (=41 cases) that were included in the exercise (for the sake of space saving, the respective references are omitted).

Table 1: Studies in the current exercise

Study (Country)	Sub-study, Wave, or Year	N	Source	Interview Type	Scale	Exposure Measure	Median Exposure [dB(A)]
ALPNAP (AT)	Road	1641	Railway	telephone	5-point verbal	Lden	59
ALPNAP (AT)	Road	1641	Road (High-way)	telephone	5-point verbal	Lden	54
ALPNAP (AT)	Road	1641	Road (Main roads)	telephone	5-point verbal	Lden	37
BBT (AT)	Railway	1449	Railway	facetoface	11-point numerical	Lden	62
BBT (AT)	Railway	1081	Railway	telephone	5-point verbal	Lden	55
BBT (AT)	Road	1327	Road	telephone	5-point verbal	Lden	60

BBT (AT)	Road	1762	Road (High-way)	facetoface	11-point numerical	Lden	54
BBT (AT)	Road	1991	Road (Main Roads)	facetoface	11-point numerical	Lden	53
Exasca (CH)	1st wave	1175	Road	postal	11-point numerical	LDay	62
Exasca (CH)	1st wave	1188	Road	postal	5-point verbal	LDay	62
Exasca (CH)	2nd wave	1186	Road	postal	11-point numerical	LDay	62
Exasca (CH)	2nd wave	1190	Road	postal	5-point verbal	LDay	62
Exasca (CH)	1st wave	1175	Road	postal	11-point numerical	Ldn	63
Exasca (CH)	1st wave	1188	Road	postal	5-point verbal	Ldn	63
Exasca (CH)	2nd wave	1186	Road	postal	11-point numerical	Ldn	63
Exasca (CH)	2nd wave	1190	Road	postal	5-point verbal	Ldn	63
Noise Study 2000 (CH)	postal	2585	Aircraft	postal	11-point numerical	LDay	49
Noise Study 2000 (CH)	postal	2585	Aircraft	postal	11-point numerical	Ldn	52
Noise Study 2000 (CH)	telephone	509	Aircraft	telephone	11-point numerical	LDay	49
Noise Study 2000 (CH)	telephone	509	Aircraft	telephone	11-point numerical	Ldn	52
Military Shooting (CH)	Military Shooting	1000	other	telephone	5-point verbal	Lae (1 year)	108
Military Shooting (CH)	Military Shooting	1000	other	telephone	11-point numerical	Lae (1 year)	108
MOSAiCH (CH)	Rail	746	Railway	facetoface	5-point verbal	LDay	37
MOSAiCH (CH)	Rail	746	Railway	facetoface	5-point verbal	Ldn	41
MOSAiCH (CH)	Road	1086	Road	facetoface	5-point verbal	LDay	55
MOSAiCH (CH)	Road	1086	Road	facetoface	5-point verbal	Ldn	56
Oslo Road Traffic (NO)	Road	3081	Road	postal	5-point verbal	Lden	58
QUALIFEX (CH)	Wave 2008	1337	Road (+Tramway)	postal	4-point verbal	LDay	47
QUALIFEX (CH)	Wave 2009	1088	Road (+Tramway)	postal	4-point verbal	LDay	46
RDF (DE)	RDF	2269	Aircraft	facetoface	5-point verbal	LDay	52
RDF (DE)	RDF	2269	Aircraft	facetoface	11-point numerical	LDay	52
RDF (DE)	RDF	2269	Aircraft	facetoface	5-point verbal	Ldn	54
RDF (DE)	RDF	2269	Aircraft	facetoface	11-point numerical	Ldn	54
Japan Railway (JP)	Railway	467	Railway	postal	5-point verbal	Ldn	57
Japan Railway (JP)	Railway	467	Railway	postal	11-point numerical	Ldn	57
SAPALDIA (CH)	SAPALDIA2	5282	Road	facetoface	11-point numerical	LDay	53
SAPALDIA (CH)	SAPALDIA2	5282	Road	facetoface	11-point numerical	Ldn	53
Schiphol 2002 (NL)	2002	5753	Aircraft	postal	11-point numerical	Lden	52
Wind Turbines Nether-lands (NL)	2007	708	other	postal	5-point verbal	Lden	34
Wind Turbines Sweden 00 (SE)	2000	341	other	postal	5-point verbal	Lden	34
Wind Turbines Sweden 05 (SE)	2005	754	other	postal	5-point verbal	Lden	33

Statistical analysis

For the present inventory, we assume that the data samples in each row of Table 1 represent a random or randomly stratified sample from the population of interest in each study or sub-study. To estimate explained variance of an exposure-annoyance model from survey data, one has to account for the random variation of the parameters of the regression model that describes the sample. We therefore used bootstrapping to estimate the probability distribution of R-squared as it emerges from regressing annoyance on exposure. The bootstrap method estimates the variability of a model parameter from the variability of that parameter between subsamples, rather than from parametric assumptions. Thus, while the bootstrap cannot estimate the "true" R-squared value in the population, it assesses the probability distribution of R-squared in the sample, based on calculating R-squared in many equally sized subsamples drawn from the original sample. As the resulting array of R-squared values obeys the central limit theorem, R-squared is normally distributed and can exhaustively be described by its mean and standard deviation. If the original study is an amply powered true random sample from the population, the mean R-squared from the bootstrap exercise is a straightforward estimate of the strengths of the exposure-annoyance relationship in the population whereas the standard deviation is an estimate of the confidence of the estimation. The larger the mean R-squared and the smaller the standard deviation that results from the bootstrap is, the better and the more confident the study was suited to describe the exposure-annoyance relationship in the sample. A study with a small sample size but relatively high mean R-squared has thus good potential to produce a more confident result if repeated with a larger sample from the same population.

Applied procedure in the linear modeling context. From each of 41 survey samples (cp. Table 1) with size N, we sampled 10'000 subsamples of size N, by sampling with replacement and modeled annoyance as a function of exposure using the sample() and lm() functions in R (R Development Core Team, 2012).

Applied procedure in the logistic modeling context. From each survey sample with size N, we sampled 10'000 subsamples of size N, by sampling with replacement, again using R's sample() function, and modeled the probability of high annoyance (HA) with logistic regression using the glm() function. HA was defined as a response on the upper 2 scale points in all verbal scales (except for the QUALIFEX study, which used a 4 point scale, where the uppermost answer alternative was used), and the upper 3 scale points (8, 9, 10) in 11-point numerical scales.
When analyzing data with logistic regression, a common procedure to e.g. model the percentage of "highly annoyed", the ordinary least squares approach to goodness-of-fit does not apply and hence, an equivalent statistic to R-squared does not exist. Therefore, to roughly asses the degree of explained variance in the model building process, the pseudo-R-squared statistic according to McKelvey and Zavoina (1975) was calculated.

R-squared as a function of study characteristics. As a second step, to investigate the role of study characteristics in the possible explanation of differences in R-squared, we modeled R-squared as a function of the source, exposure measure/metric, type of interview, and type of annoyance scale, using a straightforward GLM approach.

RESULTS

Inventory of R-squared values (and their confidence)

Figure 1 gives account of the variability of the R-squared distributions in different studies. The curves represent the normalized distributions of the bootstrap exercise's results of a few selected studies, sub-studies, or waves and demonstrate a considerable variability of R-squared values from below 0.05 up to 0.25, across studies. On average, R-squared in this selection amounts to about 0.15. It is important to note that all analyses are based on individual responses, and not on grouped data.

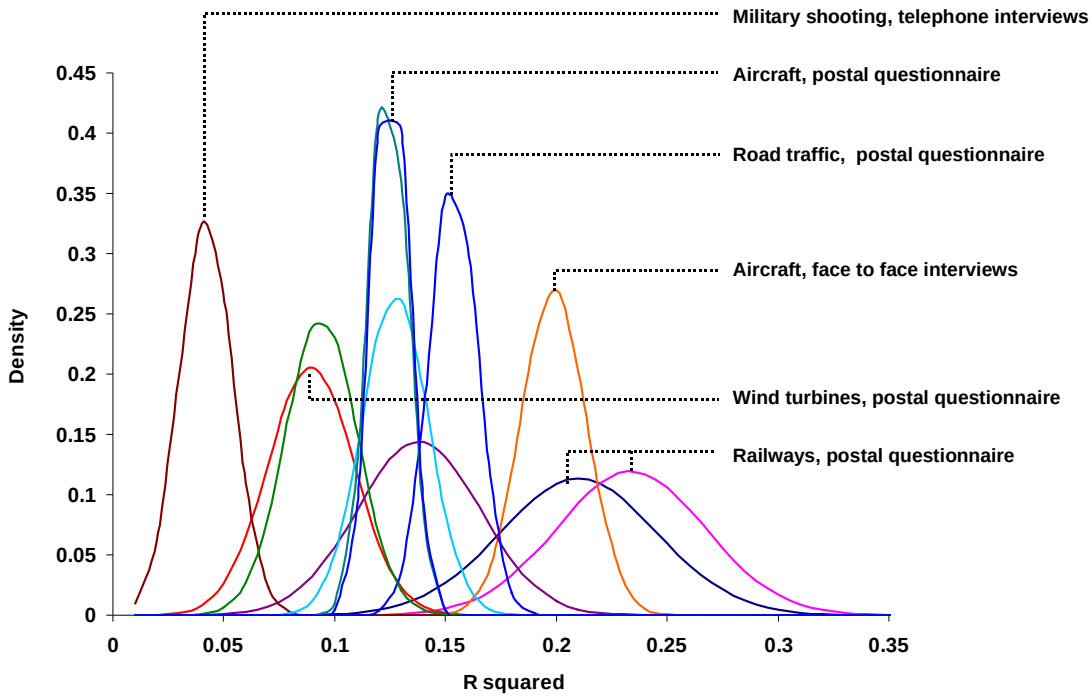

Figure 1: Estimated distributions of R-squared values of linear regression models in some selected study samples (11 out of 41) in the exercise

Effect of study characteristics on R-squared in the linear regression context

Due to the small number of studies/sub-studies and thus an essentially very small number of statistical cases (N=41, cp. Table 1) that were available for analysis, it was not possible to analyze differences of R-squared as explained by different study characteristics in a truly sound statistical multivariate hierarchical model. The following graphs show interaction plots with least squares (predicted) means, which were obtained by a simple GLM approach, essentially regressing mean R-squared (from the bootstrap) on noise source (Figure 2, left), exposure measure (Figure 2, right), type of interview (Figure 3, left), and annoyance scale type (Figure 3, right) in separate models. To control for the effect of exposure range, all four models included the interquartile range of the exposure as a covariate.

The analyses revealed slight differences of R-squared that may be explained by noise source, exposure measure, or type of interview. However, only the effect of source turned out to be significant. Contrary to expectation, the type of annoyance scale does not seem to affect R-squared.

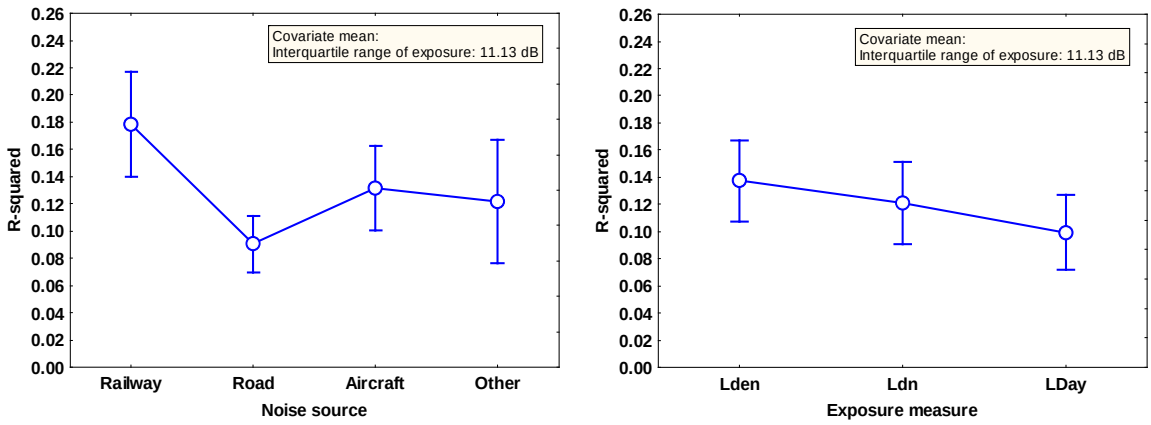

Figure 2: Least squares means of R-squared, explained by noise source (left) and exposure measure (right)

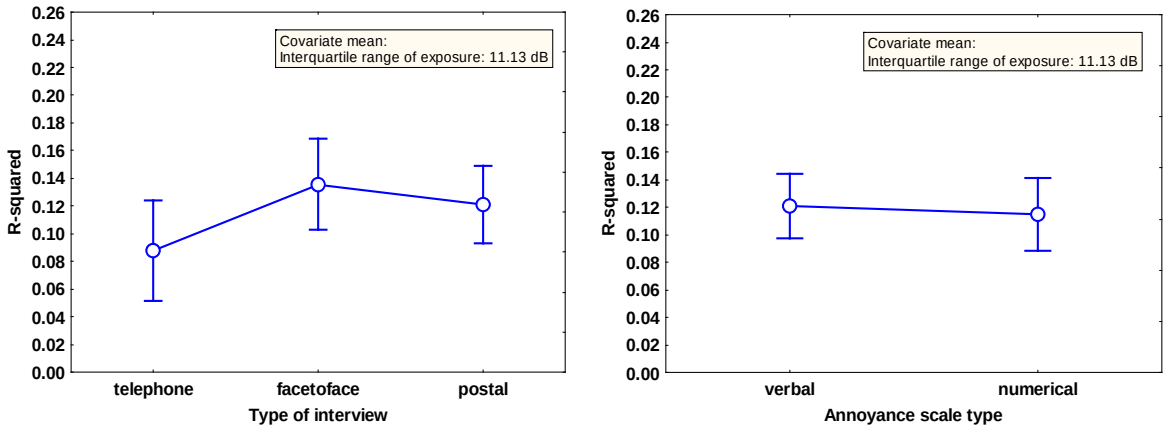

Figure 3: Least squares means of R-squared, explained by type of interview (left) and annoyance scale type (right)

Effect of study characteristics on R-squared in the logistic regression context

The above analyses were repeated with logistic regression and the calculation of the McKelvey and Zavoina pseudo-R-squared. By tendency, the logistic analysis yielded similar results, that just slightly differed from the linear results.

DISCUSSION & CONCLUSIONS

Referring to the aforementioned criticism concerning the required level of explained variance (as pointed out by Meyer, 2011), the exercise so far revealed that R-squared values for individual data in the region of 0.1 - 0.2 are by no means atypically low, but rather represent the normal range of this parameter that can be expected in noise annoyance surveys. Given today's available technology of refined exposure calculations and annoyance measurement, one would expect that R-squared values be larger than those from studies carried out decades ago (cp. Job, 1988), basically due to reduced exposure misclassifications. Strikingly, this does not seem to be the case.

The exercise provides some support for the hypothesis, that Lden better reflects the overall annoying properties of noise than Ldn, and Ldn better than LDay. Studies investigating railway noise effects in the present selection of data yielded higher R-squared values. The type of annoyance rating scale does not seem to affect levels of R-squared in a systematic way in the linear context, however, we found that models for the prediction of the percentage HA fitted slightly better with verbal scales.

Surely, the potential mechanisms to explain these differences remain to be investigated.

Limitations and outlook: The present exercise, that relied on a restricted number of non-systematically collected data sources, is of course burdened with several limitations, of which discussing even a few would go beyond the remaining available space here. For any exercise of this kind, it is essential to have enough statistical cases (i.e. studies in the present scope) in order to e.g. be able to control for cultural and language differences of response behavior as well as for potential researcher effects. The presented analyses document work in progress, further steps will hopefully encompass the collection of additional survey datasets to enhance overall power, the analysis of R-squared differences also in grouped (aggregated) data instead of only individual responses, and the necessary multilevel (hierarchical) approach to data analysis.

Acknowledgements – I would like to express my gratitude to the following people and institutions who were so kind to make available survey data from their studies to carry out this exercise: Gunn-Marit Aasvang, Peter Lercher, Martin Röösli, Tetsumi Sato, Dirk Schreckenberg, Danny Houthuijs, Eja Pedersen, Danielle Vienneau and Nicole Probst-Hensch, and the FORS Lausanne.

REFERENCES

Brink, M. (2013, 16. September 2013). *Annoyance assessment in postal surveys using the 5-point and 11-point ICBEN scales: Effects of scale and question arrangement.* Paper presented at the Internoise 2013, Innsbruck.

Fields, J. M. (1993). Effect of personal and situational variables on noise annoyance in residential areas. *Journal of the Acoustical Society of America, 93*(5), 2753-2763.

Fields, J. M., De Jong, R. G., Gjestland, T., Flindell, I. H., Job, R. F. S., Kurra, S., et al. (2001). Standardized general-purpose noise reaction questions for community noise surveys: Research and a recommendation. *Journal of Sound and Vibration, 242*(4), 641-679.

Job, R. F. S. (1988). Community response to noise: A review of factors influencing the relationship between noise exposure and reaction. *Journal of the Acoustical Society of America, 83*(3), 991-1001.

McKelvey, R. D., & Zavoina, W. (1975). Statistical-Model for Analysis of Ordinal Level Dependent Variables. *Journal of Mathematical Sociology, 4*(1), 103-120.

Meyer, A. (2011). Comment on "A field study of the exposure-annoyance relationship of military shooting noise" [J. Acoust. Soc. Am. 127, 230122311 (2010)] (L). *Journal of the Acoustical Society of America, 130*(2), 677–678.

R Development Core Team. (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.