Department for Transport

Aircraft Night Noise Effects

Peer Review

Prepared by

Dr Hannah Devine-Wright

Placewise Ltd

&

Stephen Turner, MA, MSc, HonFIOA

Stephen Turner Acoustics Limited

Report Number: PW/STA/08

Date: October 2024

PLACEWISE

STAcoustics

CONTENTS

1 Introduction

1.1 Placewise Limited (PW) and Stephen Turner Acoustics Limited (STA) have been commissioned by the Department for Transport (DfT) to carry out a peer review of the study concerning Aviation Night Noise Effects (ANNE).

1.2 The Peer Reviewers were initially provided with the following documents on 11th December 2023:

- The ANNE Study - Pilot Report;
- The ANNE Study -Technical Report;
- The ANNE Study - Subjective Study 071223 Version 0.4;
- Annex A Sample Design - Areas of low night noise exposure 051223;
- Annex B Sample Design - Stratification Scheme, distribution and minimum sample sizes 051223; and
- Annex C Sample Design -Target sample sizes and completion rates 051223.

1.3 The Inception meeting for the Peer Review was held at the offices of the Department for Transport on 11th December 2023.

1.4 This report describes the Peer Review process, sets out the views of the Peer Reviewers and provides overall conclusions.

2. The Peer Review Process

2.1 The ANNE study team comprised – the National Centre for Social Research (NatCen); St. George's University of London (SGUL), Noise Consultants Limited (NCL) and the University of Pennsylvania (Penn).

2.2 At the Inception Meeting, the reviewers met representatives of NatCen and SGUL along with officials from the Department for Transport. The study team gave a presentation describing the study and provided an example of some of the results.

2.3 It was explained that the ANNE project had two parts. The first was a subjective study comprising a social survey of respondents affected by aircraft noise at night. The second would be an objective study into the effects of night-time aircraft noise. Only the first part of the study had been completed and it is that study which is the subject of this review.

Aims of the Peer Review

2.4 The Social Scientist Peer Review had the following aims:

- Review the development and design of the questionnaire, including the phrasing of questions, response options and questionnaire length to meet the research aims.
- Review the approach to sampling airports and survey participants, and the representativeness of the sample.
- Assess the steps taken to ensure participation, for example, the ease of use of the questionnaire, accessibility and adjustments – paper questionnaire and online questionnaire participations options, translation services etc.
- Review the approach taken to maximise response rate/minimise attrition, including incentive payments and the use of reminder letters.
- Assess the appropriateness of data analytic techniques including the choice of statistical analysis.

- Quality assure key syntax used to produce the analysis.
- Consider whether the conclusions/recommendations derived in the report are appropriate given the results of the statistical analysis undertaken.
- Provide recommendations, arising from the review, for approaching future research in this area.

2.5 The Acousticians Peer Review had the following aims:

- The sampling approach (focusing on factors related to noise) and process for identification of airports for inclusion in the study. This should take in to account the information and data that was available.
- Whether relevant factors relating to aviation noise impacts were considered and taken account of in the methodology and the analysis e.g. ambient noise levels and non-acoustic factors.
- Whether the phrasing of questions in the questionnaire was consistent with current standards/best practice. If not, was the reason for this explained and justified?
- Whether the methodology for the noise modelling and the construct of any noise metrics was consistent with current standards/best practice. If not, was the reason for this explained and justified?
- An assessment of whether, given the research question, the analysis undertaken used appropriate techniques and if any potential opportunities for analysis have been overlooked.
- Whether sufficient information has been reported to allow the reviewer to gain confidence that the analysis has been carried out correctly and in accordance with the stated methodology.
- Whether sufficient reporting of results is included, such as confidence intervals, ranges, refusal rates etc. where appropriate.
- Whether the conclusions/recommendations derived in the report are appropriate given the results of the statistical/analytical analysis.

- Recommendations, arising from the review, for approaching future research in this area.

2.6 It can be seen there was, quite reasonably, an overlap in some of the aims for both the social scientist and acoustician review.

3. The Peer Review

3.1 Initially, the peer review was expected to consist of two phases:

a) An interim peer review of the ANNE Subjective Study draft report; and
b) A final peer review of the ANNE Subjective Study final report.

3.2 During the first phase, PW and STA both independently reviewed the material provided by DfT and suppliers and met online and in-person to discuss the independent reviews. As required, PW was the lead on the social science issues and STA lead on the acoustics element.

3.3 The interim peer review identified a number of issues. These were compiled into a presentation which was given to both DfT and the Suppliers at a meeting held at DfT's offices on 22nd January 2024. The issues included:

- The lack of detail regarding how the selection of the airports to include in the study was made;
- The need for more information on the sample weighting;
- The need for more information on exactly how the noise exposure for each respondent was determined; and
- The need for more information about the statistical analysis.

3.4 It was agreed that these various issues would be addressed by the Suppliers and that, in order to expedite the process there would be:

- direct engagement between the reviewers and some of the suppliers; and
- the production by the Suppliers, and subsequent peer review, of supplementary material.

3.5 The Supplementary material provided by Suppliers was in the form of various documents, including:

- A justification for selecting a logistic regression spline model with one knot;
- Information about modelling noise and annoyance data;

- A discussion of the issues when analysing HAv[1] weighted data; and
- A note about the 11-point numeric annoyance scale.

3.6 PW and STA provided further peer reviews covering:

- Comments on statistical modelling;
- Notes on the newly produced Noise Appendix; and
- Notes about the process for determining HAvw[2].

3.7 In early July 2024, PW and STA received the following documents to enable the peer review to be completed:

- Aviation Night Noise Effects – Subjective study report;
- Aviation Night Noise Effects – Technical report;
- Annex A – Sample Design: areas of low night noise exposure;
- Annex B – Sample design: stratification scheme, distribution and minimum sample sizes;
- Annex C – Sample design target sample sizes and completion rates; and
- Annex D – Noise Data for the ANNE subjective study.

3.8 These documents were reviewed and informal feedback was provided to DfT. The comments in section 4 below are based on the review of these study reports.

[1] The percentage of 'Highly Annoyed' (HA) is used to describe community response to noise. Using the 5-point verbal scale, 'Highly Annoyed' (HAv) is calculated as the 40% most annoyed. HAv includes those who report being 'very' annoyed (a score of 4) and those who report being 'extremely' annoyed (a score of 5).
[2] With the 11-point numeric scale, 'Highly Annoyed' (HA_N) is calculated as a score of 8, 9 or 10, equivalent to 28% being most annoyed. To enable comparison between the 11-point numeric scale and the 5-point verbal scale, a 28% 'Highly Annoyed' figure (HAvw) can be calculated using statistical methods.

4. Outcomes of the Peer Review

4.1 The aims of the Social Scientist review have been used as a basis for describing the outcomes of the peer review. The aims of the acoustician peer review have been included in the appropriate section. Each of the Social Scientist aims of the peer review are addressed in turn:

The development and design of the questionnaire, including the phrasing of questions, response options and questionnaire length to meet the research aims.

ISO/TS 15666:2021 compliance

4.2 To provide consistency with ISO/TS 15666:2021 and past surveys, the questionnaire included two measures of noise annoyance: a 5-point verbal response scale and an 11-point numerical rating scale. The format of the 11-point numeric response scale ranges from 0 to 10. Unfortunately, an error led to the response format of the 11-point scale in the paper version of the survey being 1 to 10 rather than 0 to 10.

4.3 As the format was incorrect, the 11-point numeric question about annoyance was non-compliant with ISO/TS 15666:2021. The reviewers agree that due to how the ANNE survey data are weighted, it was not appropriate to obtain a meaningful analyis of the sub-sample who completed the correct 11-point scale online. Therefore, data from the 11-point numeric scale was excluded from subsequent analysis.

4.4 Acknowledging the error, NatCen committed to investigating their internal processes to avoid similar errors in the future. Similarly, it is understood that DfT has reviewed their processes. It is recommended that, in the future, Suppliers provide DfT with an actual copy of the printed and online version of surveys prior to sign off.

Pilot study

4.5 The timescale for the pilot invitation and reminder letters was reasonable.

4.6 The address-level response rate (22%) for the pilot study was reasonable.

4.7 The 11-point numeric scale error was not picked up at the pilot study stage because the pilot version of the survey only contained the 5-point verbal scale due to concern about the length of the survey, and, therefore, time taken to complete the survey. It is recommended that future pilot surveys prioritise use of the 11-point numeric question as according to the Suppliers' the 11-point numeric scale '*offers the greatest options for statistical testing and cross-study comparisons*.'[3]

4.8 Concern was raised about the ethics of asking pilot participants to consent to participate in the objective sleep study when the research team knew that they would not be selected. Ethics procedures were reviewed by DfT and deemed satisfactory. It was reported that 63% of the Pilot sample agreed to be contacted.

Wellbeing measures

4.9 The Subjective survey included a single measure of wellbeing, the validated Short Warwick-Edinburgh-Mental-Wellbeing-Scale (SWEMWBS). Clark et al., (2021)[4] recommend that the ONS single life satisfaction item should be included in surveys as economists use it as a proxy for wellbeing. This item was not included in the survey due to concerns about the overall length of the survey. Excluding the ONS item limits the comparability of the findings of the Subjective Study to other studies, for example, SoNA14[5]. To enable comparison with other studies, consider including the ONS single life satisfaction item in future surveys.

[3] Clark et al., 2021, Revising ISO/TS 15666- the noise annoyance standard. ICBEN conference 14-17 June, Stockholm, 2021.
[4] Ibid.
[5] CAA – Survey of Noise Attitudes 2014: Aircraft Noise and Annoyance – Second Edition (CAP 1506 2021)

The Pittsburgh Sleep Quality Index (PSQI)

4.10 The University of Pennsylvania recommended inclusion of the PSQI as a way to measure sleep quality. The reviewers noted that in order to reduce the length of the survey it included two out of seventeen items from the PSQI and three out six items from the Brief-PSQI. The Suppliers considered that it was sufficient to include a single-item sleep quality question from the Brief-PSQI. The reviewers suggested that consideration is given to using Snyder et al (2018)[6] single item 10-point Sleep Quality Scale in future studies. The Suppliers stated that the full PSQI will be used in the follow-on field study. The reviewers queried the use of a 5-point response format rather than the typical PSQI 4-point response format. The Suppliers responded that using the 5-point version enables comparison with the USA FAA's National Sleep Study (NSS)[7], and the full PSQI with the 4-point response format will be used in the field study.

The order of survey questions

(a) Prioritisation of questions about annoyance and sleep disturbance

4.11 As questions about annoyance and sleep disturbance had a similar format, they were deliberately spaced apart within the questionnaire. This resulted in questions about sleep disturbance occurring relatively late in the survey. It is not clear what impact, if any, the ordering of the questions in this way had on the data collected. Consideration could be given to varying the order of key questions in future surveys. It is acknowledged that this may be easier to achieve with online, rather than paper surveys.

[6] Snyder et al., (2018). A new single-item Sleep Quality Scale: Results of psychometric evaluation in patients with chronic primary insomnia and depression. *Journal of Clinical Sleep Medicine, 14,* 1849–1857. http://dx.doi.org/10.5664/jcsm

[7] Basner et al., (2023) Effects of Aircraft Noise on Sleep: Federal Aviation Administration National Sleep Study Protocol. Int J Environ Res Public Health, 6;20(21):7024. doi: 10.3390/ijerph20217024.

(b) Routing of survey responses

4.12 The survey included four screening questions derived from the USA FAA Study. These questions enabled participants to be recruited for the second (objective) part of the ANNE study. In order to minimise potential confusion about the aims of the survey, and the time required to complete the survey, the reviewers suggest that screening questions are placed after asking for agreement to take part in follow-on studies.

Conclusion

4.13 Overall, the questionnaire was satisfactory and would be expected to achieve the survey aims.

Review the approach to sampling airports and survey participants, and the representativeness of the sample.

4.14 The methodology adopted for determining which airport to include in the study is described in the Technical Report. Quite sensibly, it was decided that the so-called designated airports of London Heathrow, London Gatwick and London Stansted would be purposely included. To identify the other airports, a range of data were obtained for over 20 airports and from that the remaining 5 airports were identified.

4.15 Exactly how that information was used has remained unclear, but in the end, the airports selected were appropriate and provided a good geographical distribution and a good range of night-time noise exposure.

4.16 The survey participants sampling was appropriate and met the requirements of the study. It was noted that although NatCen's experience suggested 3 or 4 variables could be used to stratify postcode address files, 5 variables were used in this study.

Assess the steps taken to ensure participation, for example, the ease of use of the questionnaire, accessibility and adjustments – paper questionnaire and online questionnaire participations options, translation services etc.

4.17 Lessons were learned from the pilot study and these included providing a freephone number and dedicated email address in the invitation letter along with an option to request a (large print) hard copy.

4.18 The timescale for the invitation letter and reminder letters was reasonable. It was also a reasonable assumption that 9% of addresses would be ineligible (non-residential). Across the pilot and the main study, there was a suitable address-level response rate and clear explanations were provided in the study report regarding where there was a lower response rate.

Review the approach taken to maximise response rate/minimise attrition, including incentive payments and the use of reminder letters.

4.19 The disruption caused by a postal strike was unfortunate but the study dealt with this inconvenience pragmatically in order to maximise the response rate.

4.20 The incentive in the form of a voucher and the value (£10) was considered appropriate for the study.

4.21 The format and frequency of reminder letters was appropriate and aided satisfactory response rates.

Assess the appropriateness of data analytic techniques including the choice of statistical analysis.

4.22 Additional information was requested and provided about the choice of statistical analysis, particularly the logistic regression analysis, the linear splines, knots and how the fit was established.

4.23 Clarification was also sought and provided about the use of weighted data, particularly how, when and why HAv and HAvw were calculated. Furthermore, crosschecking of all figures in the tables and the text was requested and addressed.

4.24 Consideration was given to ways in which data from the 11-point annoyance scale could be used in light of the typographical error associated with the paper copy of the survey. This resulted in adoption of the Monte Carlo approach using the weighted 5-point verbal scale (HAvw) as it is simple to understand, uses available data and avoids problems associated with how to assign a (0.4) weight across the sample.

4.25 Overall, appropriate detail has been provided about the type of statistical analysis undertaken.

4.26 Initially the methodology used for determining the noise exposure was barely explained. Following discussions between STA and NCL the approach taken became clear and as a result Appendix D was produced. The Suppliers are commended for the methodology used, basing the impact on the actual flightpath used by every flight of in the time periods being examined. This is a significant improvement over previously modelling methodologies which relied on assumptions about dispersion around departure tracks.

4.27 Pragmatically, the methodology used runway times to determine the time of the impact on a respondent. This does mean that for locations some distance from the airport, the time of impact would be a little before (for arrivals) or a little after (for departures) than that used in the analysis. Broadly, this would not be an issue, except for events close to the junction between two time periods. However, it is not felt that these potential differences had a bearing on the final results produced by the study.

Quality assure key syntax used to produce the analysis.

4.28 Queries were raised and addressed about the Stata syntax, for example, clarifying the meaning of codes, and the number and type of strata used to

weight by age and sex. The Stata syntax provided to the reviewers was considered appropriate for the study.

Consider whether the conclusions/recommendations derived in the report are appropriate given the results of the statistical analysis undertaken.

4.29 PW and STA are satisfied with the appropriateness of the analysis and the presentation of the results.

4.30 The report noted that, in order to determine a value for LOAEL[8], a decision had to be made regarding the level of effect to use. It was noted that the 2018 World Health Organisation Guidelines[9] used 10%HA[10] as a threshold and 11% for HSD[11]. The study suggests using a value of 10% for both. This seems reasonable.

4.31 On that basis, the study concludes that for annoyance at night, the LOAEL would be 43 dB, $L_{Aeq,8h}$. The corresponding value for L_{night} is 42 dB. These conclusions correctly interpret the data, including noting that account must be taken of the inherent uncertainties in calculating noise exposures at lower levels. DfT currently define LOAEL at night as 45 dB[12] and, as stated in the study report, the World Health Organization has set a value of 40 dB[5], L_{night}.

4.32 Asking the study to look at the impact during specific hours of the night was an appropriate next step in trying to understand the effects of night noise and how best to manage it. The approach adopted by the study was sensible although there needs to be some caution over the detailed conclusions drawn about individual hours of the night due to issues of sample size.

4.33 The results of the study demonstrate that consideration of the N60 indicator is helpful in showing how the number of events affect annoyance. Although it

[8] Lowest Observed Adverse Effect Level as described in Government noise management policy
[9] Environmental Noise Guidelines for Europe (WHO, 2018)
[10] Highly Annoyed
[11] Highly Sleep Disturbed
[12] Air Navigation Guidance (DfT 2017)

must be borne in mind that an aircraft can equally qualify to be counted in the N60 value with a maximum level of 61 dB or 81 dB.

4.34 Conversely, although the relationship between annoyance and overflights was examined, the conclusions drawn need to be treated with caution as for an aircraft to be counted, it only has to be flying in the cone of interest. Its impact could range from virtually silent to very loud and yet it would still count as one overflight.

4.35 For HSD, the results suggest that the LOAEL value is 44dB, $L_{Aeq,8h}$, 1dB higher than the LOAEL for Highly Annoyed.

4.36 Overall, the peer reviewers are satisfied with the conclusions drawn about the values of LOAEL from the results of this study.

Reflection on the study's overall contribution to the evidence base on aviation noise.

4.37 The inclusion/exclusion criteria for items in the survey both enabled and limited comparison with other UK and international studies. For example, including the Brief PSQI sleep quality item enabled comparison with the USA FAA NSS study. It is recommended that future surveys consider use of multiple measures of key indicators, and prioritise the use of validated, single-item measures, for example, sleep quality and wellbeing.

4.38 Consideration could also be given to potential measurement (ceiling) effects associated with using SWEMBWS (e.g., Melin et al., 2022)[13], and the sensitivity of SWEMWBS for assessing perceived wellbeing associated with events that occurred prior to the time at which the survey was completed.

4.39 Assessing the effects of night time aviation noise in hourly intervals and adopting the spline analytic approach has shown how the association between aviation noise and annoyance / sleep disturbance can vary across the night.

[13] Melin J, Lundin A., & Johansson M. (2022) An off-target scale limits the utility of Short Warwick-Edinburgh Mental Well-Being Scale (SWEMWBS) as a measure of well-being in public health surveys. Public Health, 202:43-48. doi: 10.1016/j.puhe.2021.10.009.

Provide recommendations, arising from the review, for approaching future research in this area.

4.40 In the Peer Review (CAP 2161a)[14] of the CAA Report – Survey of Noise Attitudes 2014: Aircraft Noise and Sleep Disturbance (CAP 2161 July 2021), it was noted that:

> *Caution needs to be exercised with both these conclusions, however. Firstly, as explained several times in the report, this analysis is, of necessity, exploratory as the SoNA2014 survey design was aimed at understanding the annoyance caused during the daytime and not the impact of aircraft noise on sleep disturbance.*
>
> *Some interesting results emerged, but as indicated above, they must be treated with caution and probably used only as issues that should be explored in any future night time disturbance study.*

4.41 Similar views were expressed in the Peer Review (CAP 2251a)[15] of the CAA Report Survey of Noise Attitudes 2014: Aircraft Noise and Sleep Disturbance – Further Analysis (CAP2251, December 2022).

4.42 Consequently, this study is welcomed as it focuses on the impact of aviation noise at night, including on sleep disturbance.

4.43 The methodology used to determine the noise exposure of each respondent was very sophisticated and has demonstrated how lower noise exposure levels can be obtained more robustly.

4.44 There was still the challenge of asking questions about how people felt they were affected by a noise exposure that occurred in the past and the reliability of their recollections. Furthermore, there is some uncertainty about how well

[14] Survey of Noise Attitudes (SoNA) 2014: Aircraft Noise and Sleep Disturbance – Peer Review (CAP 2161a), July 2021

[15] Survey of Noise Attitudes (SoNA) 2014: Sleep – Further Analysis – Peer Review (CAP 2251a), September 2022

respondents can differentiate their responses about the effect of the aviation noise impact at, say, 0200 hours and 0300 hours.

4.45 The follow-on objective study will examine these issues, with the results potentially increasing more robustly the understanding of the effects of aircraft noise exposure at night.

- - - -