U.S. Department of Transportation
Federal Aviation Administration

FINAL PROJECT REPORT

Form Approved:
O.M.B. No. 2120-0559
9/30/2013

PART I - PROJECT IDENTIFICATION INFORMATION

1. Institution and Address University of Pennsylvania	2. FAA Program	3. FAA Award Number 13-C-AJFE-UPENN
1019 Blockley Hall, 423 Guardian Drive, Philadelphia, PA 19104	4. Award Period From 10/1/2015 To 9/30/2018	5. Cumulative Award Amount $744,805.00

6. Project Title
Pilot study on aircraft noise and sleep disturbance.

PART II - SUMMARY OF COMPLETED PROJECT (For Public Use)

Aircraft noise can disturb sleep and impair recuperation. Research is needed to develop exposure-response relationships that are representative of noise-exposed communities and can be used to inform noise mitigation policy. For a national field study on physiologic response to aircraft noise during sleep to be feasible, an inexpensive yet sound study methodology is needed. We performed a pilot study around Atlanta Hartsfield-Jackson International Airport (ATL) using electrocardiography and actigraphy to monitor sleep. The primary objective was to evaluate the quality and quantity of data that could be obtained by recruiting participants by postal questionnaire, shipping them the physiologic and noise measurement equipment, and the unattended setup of the equipment and recording of data by the participants themselves. We mailed 4080 questionnaires containing items on sleep, health and noise disturbance to residences around ATL exposed to at least 35 dB Lnight aircraft noise. From 407 respondents, 34 participants completed five nights of unattended sleep measurements. Data of sufficient quality and quantity were obtained. Self-reported and physiologic awakenings increased as a function of maximum aircraft noise levels, although this effect was of only borderline statistical significance for physiologic awakenings likely due to the low sample size. The approach used in the pilot study was found to be feasible for the purpose of the larger-scale study among a representative population around multiple airports in the future.

PART III - TECHNICAL INFORMATION (For Program Management Uses)

1. ITEM (Check appropriate blocks)	NONE	ATTACHED	PREVIOUSLY FURNISHED	TO BE FURNISHED SEPARATELY TO PROGRAM	
				Check (X)	Approx. Date
a. Abstracts of Theses	✘				
b. Publication Citations		✘			
c. Data on Scientific Collaborators		✘			
d. Information on Inventions	X				
e. Technical Description of Project and Results		✘			
f. Other (specify)					

2. Principal Investigator/Project Director Name (Typed)	3. Principal Investigator / Project Director Signature	4. Date
Mathias Basner		3/24/20

FAA Form 9550-5 (03-03) Supersedes Previous Edition

PART III – TECHICAL INFORMATION

a) Abstracts of Theses

None

b) Publication Citations

Peer Reviewed Journal Publications

Basner, M., Witte, M., & McGuire, S. (2019). Aircraft noise effects on sleep – Results of a pilot study near Philadelphia International Airport. International Journal of Environmental Research and Public Health, 16(17): 3178.

Rocha, S., Smith, M., Witte, M., & Basner, M. (2019). Survey results of a pilot sleep study near Atlanta International Airport. International Journal of Environmental Research and Public Health, 16(22): 432.

Smith, M., Rocha, S., Witte, M., & Basner, M. (2020). On the feasibility of measuring physiologic and self-reported sleep disturbance by aircraft noise on a national scale: A pilot study around Atlanta airport. Science of the Total Environment, 718: 137368

Smith, M., Witte, M., Rocha, S., & Basner, M. (2019). Effectiveness of incentives and follow-up on increasing survey response rates and participation in field studies. BMC Medical Research Methodology, 19: 230.

Published Conference Proceedings

Basner, M., Smith, M., Rocha, S., & Witte, M. (2019). Pilot field study on the effects of aircraft noise on sleep around Atlanta International Airport. Presentation and conference paper 23rd International Congress on Acoustics, Aachen, Germany.

c) Data on Scientific Collaborators

Investigation Team

Mathias Basner, PI, University of Pennsylvania: Lead on all tasks

Michael Smith, Postdoctoral researcher, University of Pennsylvania: Data analysis on all tasks

Sarah Rocha, Research assistant, University of Pennsylvania: Data collection and technical and administrative support on all tasks

Maryam Witte, Research assistant, University of Pennsylvania: Data collection and technical and administrative support on all tasks

Katharine Casario, Research assistant, University of Pennsylvania: Technical and administrative support on all tasks

d) Information on Inventions

None

Project 017 Pilot study on aircraft noise and sleep disturbance. Final report.

University of Pennsylvania

Project Lead Investigator

Mathias Basner, MD, PhD, MSc
Associate Professor of Sleep and Chronobiology
Department of Psychiatry
University of Pennsylvania
1019 Blockley Hall, 423 Guardian Dr.
Philadelphia, PA 19104-6021
215-573-5866
basner@pennmedicine.upenn.edu

University Participants

University of Pennsylvania

- PI(s): Mathias Basner, associate professor
- FAA Award Number: 13-C-AJE-UPENN-011
- Period of Performance: October 01, 2015 to September 30, 2018
- Task(s):
    - Pilot study on aircraft noise and sleep disturbance around Atlanta (ATL) airport

Investigation Team

Mathias Basner, Associate Professor, University of Pennsylvania: P.I. and lead on all tasks

Michael Smith, Postdoctoral researcher, University of Pennsylvania: Data analysis on all tasks

Sarah Rocha, Research assistant, University of Pennsylvania: Data collection and technical and administrative support on all tasks

Maryam Witte, Research assistant, University of Pennsylvania: Data collection and technical and administrative support on all tasks

Katharine Casario, Research assistant, University of Pennsylvania: Technical and administrative support on all tasks

Project Overview

The long-term goal of this line of research is to derive exposure–response relationships for aircraft noise-induced sleep disturbance that are representative of the exposed U.S. population. Studies will have to investigate samples around multiple airports; therefore, it will not be possible to use polysomnography [i.e., simultaneous recording of the electroencephalogram (EEG), electromyogram, and electrooculogram] to monitor sleep because this would require trained personnel at the measurement site in the evening and morning, which would be too costly. An alternative method of using a single-channel electrocardiogram (ECG) and actigraphy to monitor sleep has been examined. This would allow investigation of a greater number of subject samples at lower cost because individuals can be taught how to apply the electrodes themselves. Also, in contrast to polysomnography, awakenings can be identified automatically. Awakenings are defined as brain activations (so-called EEG arousals) that last 15 s or longer. As part of previous research, we refined an algorithm for identifying EEG arousals (Basner et al., 2007) based on increases in heart rate to identify only those arousals ≥15 s in duration, which is the most agreed upon indicator of noise-induced sleep disturbance. High agreement was obtained between arousals scored visually from the EEG and those identified using the refined ECG-based algorithm. The method of using ECG and actigraphy to monitor sleep has been implemented in two pilot field studies to evaluate the quality of data that can be obtained for unattended physiological and noise measurements. Based on lessons learned, the study protocol is being refined to inform the design and cost of a potential multi-airport study on the effects of noise on sleep.

Task 1 Pilot study on aircraft noise and sleep disturbance around Atlanta airport

Objective(s)

Aircraft noise can disturb sleep and impair recuperation. Research is needed to develop exposure-response relationships that are representative of noise-exposed communities around multiple airports and that can be used to inform noise mitigation policy in the United States. To achieve this goal, we will conduct a field study around airports throughout the U.S. in which we will measure both aircraft noise exposure in the bedroom and physiologic response to this noise during sleep. In order for this National Sleep Study (NSS) to be feasible, which is anticipated to involve scores of airports and several hundred participants, an inexpensive yet sound study methodology is needed. In an earlier pilot study around Philadelphia International Airport (PHL) we demonstrated that electrocardiograph (ECG) electrodes and actigraphs measuring body movements could easily and non-invasively be applied to the torso by study participants themselves. This greatly reduces the methodological study cost compared to fully attended studies. In a second pilot study, which forms the basis of this report, the methodology of using ECG and actigraphy to monitor sleep was implemented around Atlanta Hartsfield-Jackson International Airport (ATL). The primary objective of this study was to continue improving study methodology, in particular evaluating the quality and quantity of data that could be obtained when recruiting participants by postal questionnaire, shipping them the physiological and noise measurement equipment, and the unattended setup of the equipment and recording of data by the participants themselves, in preparation for the larger-scale NSS. A secondary objective of the study was to compare objective and subjective measures of sleep and health between groups exposed to different levels of nocturnal aircraft noise.

Research Approach

I. Summary

We mailed 4080 questionnaires containing items on sleep, health and noise disturbance to residences around ATL that were exposed to at least 35 dB L_{Night} aircraft noise. A number of different mailing strategies were adopted to maximize response rates. Prepaid cash incentives and sending follow-up reminder and survey waves were an effective method of improving response rates.

Completed questionnaires were received from 407 respondents, who were broadly representative of their geographical region. Among these respondents, calculated outdoor nighttime air traffic noise was significantly associated with self-reports of worse overall sleep quality, trouble falling asleep within 30 minutes, annoyance, and sleep disturbance. Residents in areas exposed to higher levels of aircraft noise coped by closing the windows at night.

From among the questionnaire respondents, 37 participants were initially recruited into the field study, with 34 participants completing five nights of unattended sleep measurements and 3 recruits dropping out before the study began. Data of sufficient quality and quantity to investigate the effects of aircraft noise on sleep were obtained, despite some data loss in the field study due to technical issues with the equipment and non-compliance among the participants. The technical issues were the main cause of data loss however, and non-compliance was low, with both physiologic and acoustic data collected by the participants in 87.6% of all study nights.

Concerning the primary objective of the study, evaluation of the feasibility of the study methodology, we demonstrated both the feasibility of recruiting field study participants by postal questionnaire in a larger, more nationally representative sample for future studies around multiple airports, and the feasibility of mailing equipment to participants to obtain unattended physiologic and acoustic measurement data.

Regarding the secondary objective of the study, investigating noise-induced effects on physiologic and self-reported sleep, a number of statistically significant outcomes were found, including associations between aircraft noise and physiologic and recalled awakenings. However, these findings are from a sample population of limited size, living close to a single airport. The findings of physiologic and self-reported effects of aircraft noise on sleep may not be representative of response among a demographically diverse national study population exposed to different patterns of nocturnal aircraft noise. A larger-scale study among such a population should be performed in the future, and the approach used in the present pilot study has been demonstrated to be feasible for this purpose.

II. Glossary of terms

ATL	Hartsfield-Jackson Atlanta International Airport
A-Weighting	Frequency weighting filter applied to a sound measurement to mimic the frequency-dependence of human hearing
dB	Decibel, relative to the threshold of human hearing (2×10^{-5} Pa)
dB(A)	A-weighted decibel
CI	Confidence interval
DLR	German Aerospace Center (Deutsches Zentrum für Luft- und Raumfahrt)
ECG	Electrocardiogram
EEG	Electroencephalogram
EMG	Electromyogram
EOG	Electrooculogram
FAA	Federal Aviation Administration
FRA	Frankfurt Airport
H5	Zoom H5 Handy Recorder
ICBEN	International Commission on the Biological Effects of Noise
INM	Integrated Noise Model
$L_{AEq,sleep}$	A-weighted equivalent continuous sound pressure level during an individual's sleep period time from sleep onset to sleep cessation
$L_{AEq,t}$	A-weighted equivalent continuous sound pressure level over specified time period t
$L_{AF,max}$	Maximum A-weighted sound pressure with fast (0.125 s) time constant
$L_{AS,max}$	Maximum A-weighted sound pressure with slow (1 s) time constant
L_{Night}	Nighttime (23:00-07:00) A-weighted outdoor equivalent sound pressure level from aircraft
$L_{Night,cat}$	Nighttime (23:00-07:00) A-weighted outdoor equivalent sound pressure level from aircraft, categorized into 5 dB bins
NSS	National Sleep Study
PHL	Philadelphia International Airport
PSQI	Pittsburgh Sleep Quality Index
PSG	Polysomnography
SPL	Sound pressure level
SSS	Stanford Sleepiness Scale
UPenn	The University of Pennsylvania
XL2	NTi Audio XL2 Class 1 sound level meter

III. Background and introduction

Humans spend approximately one third of their lives asleep, yet the core function or functions of sleep remains elusive. Some of the proposed functions of sleep include clearance of neural waste products that build up in the central nervous system during wakefulness, reducing cellular stress, synthesis of cellular components in preparation for the next period of wakefulness, consolidation of memories and restoration of cognitive performance [1-5]. Whatever the core function of sleep, it is critical for good physical and mental health, and chronic short sleep duration is associated with increased risk for obesity in both adults and children, diabetes, hypertension, cardiovascular disease and all-cause mortality [6-10]. Nocturnal traffic noise can impair physiologic and subjective sleep, by causing cortical awakenings and self-reported sleep disturbance [11]. With the most recent US sleep study dating back to 1996 [12], US research on the effects of aircraft noise on sleep, particularly compared to the efforts of some European countries, has lagged over the past 20 years. During the intervening time, US air traffic has changed significantly, with substantial increases in traffic volume over the past 30 years on one hand, and significant reductions in noise levels of single aircraft on the other. Due to inter-cultural differences and different operational procedures, results from studies performed outside the US may not translate directly to US domestic airports. Therefore, it is important that field studies be conducted in the US to acquire current data on sleep disturbance relative to varying degrees of noise exposure.

The long-term goal is to perform a National Sleep Study (NSS) throughout the U.S. to derive exposure-response relationships for aircraft noise-induced sleep disturbance that are representative for the exposed US population. Since airports differ in nocturnal traffic volume and pattern, it will be necessary to investigate several airports across the US that are representative for all US airports with relevant nocturnal air traffic to achieve this goal. The pilot study presented in the current report represents a preparatory step towards implementing the NSS. Prior to this point, we made significant progress during our work within the FAA Centers of Excellence PARTNER and ASCENT to achieve this long-term goal (Table 1).

Table 1 Overview of previous accomplishments made as part of the PARTNER COE.

Funding Period	Result
2010-11	Proposed an initial study design for a US field study on the effects of aircraft noise on sleep.
2011-12	Refined the ECG-based algorithm for the automatic detection of cortical arousals to better reflect EEG awakenings. This refinement was based on the 2011 NORAH[1] data.
2012-13	Validated the refined ECG-based algorithm with the 2012 NORAH data. Wrote a MatLAB™ software interface that facilitates the automatic identification of EEG awakenings based on a single channel ECG and body movements.
2013-14	Completed preparation for a field study examining the effects of aircraft noise on sleep around Philadelphia International Airport (PHL). GIS modeling of socio-demographic characteristics were completed to select the control area. Developed study materials including recruitment flyers and questionnaires. New hardware was purchased and coupled with software.
2014-15	Completed a pilot field study on the effects of aircraft noise on sleep around PHL and in a control area not exposed to aircraft noise.

In 2010/2011, we proposed an initial study design for the NSS [13]. Models relating noise characteristics of single aircraft events (e.g. maximum A-weighted sound pressure level, $L_{AS,max}$) and physiological reactions (e.g. awakenings) will be the primary outcome of the NSS, which will have to investigate samples representative of exposed populations, and therefore sample more subjects than similar studies that have been conducted in the past. The gold standard for measuring sleep is polysomnography (PSG), which is the simultaneous measurement of the electroencephalogram (EEG), electrooculogram

[1] NORAH was a multi-disciplinary study on the effects of aircraft noise performed around FRA Frankfurt Airport (Frankfurt, Germany).

(EOG), and electromyogram (EMG). This method has been implemented in a few field studies on the effects of road, rail, or aircraft noise on sleep [14-17]. However, PSG is methodologically expensive to implement. Trained staff are needed at the measurement site in the evening and the morning to respectively apply and remove the electrodes. Trained sleep technologists are needed to visually score sleep stages, which has both high intra- and inter-rater variability [18, 19]. Finally, the methodology is somewhat invasive and may itself influence sleep, especially during the first night(s) [20]. For these reasons, it is not viable to implement PSG in studies of the planned scale; as of July 2019 the NSS is anticipated to involved 400 field study participants living around 77 airports within the U.S. Based on the 2010/2011 results of PARTNER Project 25B, it was proposed to use a combination of actigraphy (skeletal muscle movement) and electrocardiography (heart rate) instead of PSG, which will allow a cost-effective and methodologically sound investigation of large subject cohorts.

Awakenings are typically associated with arousals of the autonomic nervous system, which include increases in heart rate and blood pressure. In prior publications, we were able to show the potential of an automatic ECG-based algorithm to predict cortical arousals [21, 22]. During an earlier project , this algorithm was refined in order to only identify cortical arousals that are 15 seconds or longer in duration [23], which is the indicator of noise-induced sleep disturbance most commonly used [24].

In 2011/2012, the University of Pennsylvania (UPenn) and the German Aerospace Center (DLR) collaborated to develop common methodological approaches to be used both in the NSS and in a DLR field study (called NORAH) [25-27]. The first two waves of the NORAH study (summers of 2011 and 2012) used standard PSG to investigate 120 subjects living around Frankfurt Airport (FRA) for 3 consecutive nights. In the third wave, 187 volunteers (including 39 who participated in all 3 waves) were investigated with the less methodologically expensive ECG-based method for the detection of awakenings [28]. The advantage of replacing PSG with the less costly actigraphy and ECG-algorithm is that much larger and representative subject populations can be investigated at an acceptable cost. However, the validity of the ECG-based algorithm is crucial for the success of the NSS that will rely only on actigraphy and the ECG.

The ECG algorithm was originally programmed to detect cortical arousals (defined as activations lasting 3 s or longer) rather than EEG awakenings (defined as cortical activations lasting 15 seconds or longer). In terms of noise effects prediction and noise policy, EEG awakenings may be superior indicators of noise induced sleep disturbance than cortical arousals [29]. Noise policy and noise indices based on awakening probability are already in use at the airports in Leipzig/Halle, Zurich, and Frankfurt [30, 31]. A 2012 assessment of the effects of aircraft noise on sleep at Montreal airport was also based on awakening probability [32].

In the 2011-2012 period, the ECG algorithm was thus refined to better reflect EEG awakenings (i.e., it was the goal to detect cortical arousals 15 seconds or longer). However, with kappa=0.733, the agreement fell short of an a priori set goal of kappa=0.80 which marks the beginning of "almost perfect" agreement [33].

In 2012-2013, the ECG algorithm was thus further refined. It now combines arousals that are scored based on the ECG and actigraphically-determined body movements, and it is able to estimate sleep onset and offset based on heart rate and movement activity alone. A comparison of kappa values based on the refined algorithm is shown in Figure 1.

The pre-defined threshold of kappa=0.80 was surpassed (0.86). As UPenn's algorithm outperformed DLR's algorithm, we moved forward with Penn's algorithm only. We developed a MatLAB™ software interface that allows an easy analysis of ECG and actigraphy data, and automatically outputs start and end times of automatically detected arousals.

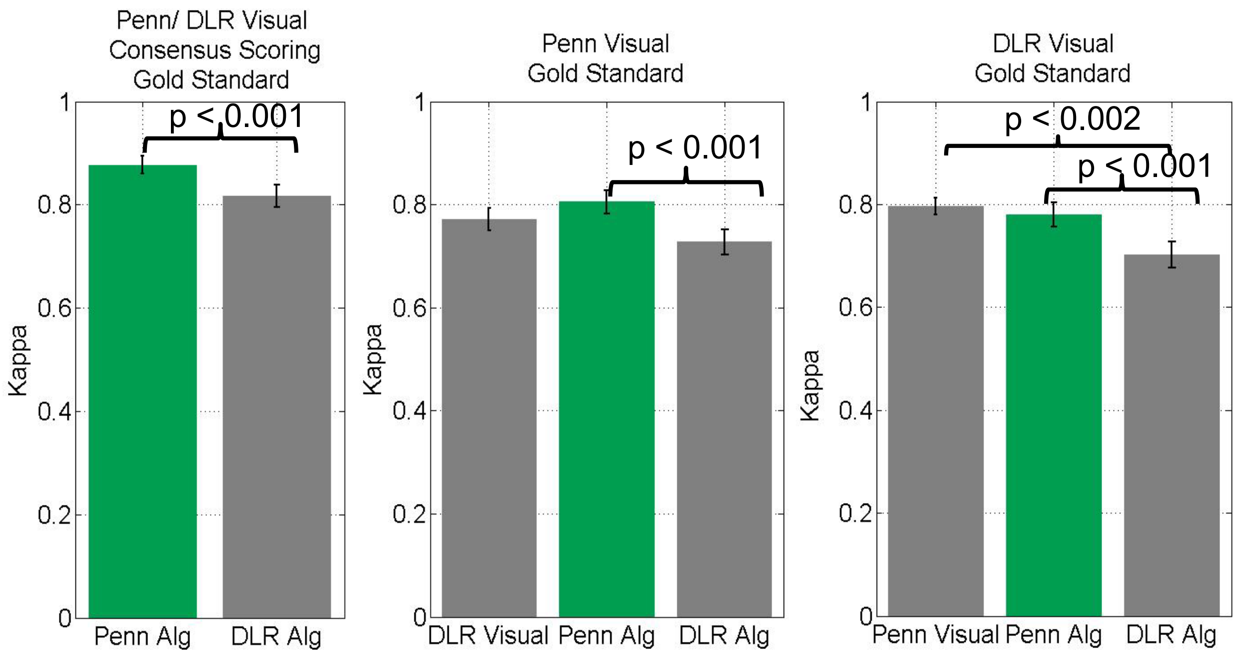


Figure 1. Chance corrected agreement (kappa) between visual (DLR, Penn) and automatic (DLR Alg, Penn Alg) arousal scorings is shown for a consensus arousal scoring (left graph), for Penn visual scoring being the gold standard (middle graph), and for DLR visual scoring being the gold standard (right graph). Kappa values indicated almost perfect (kappa>0.80) agreement between both algorithms and the consensus scoring. Penn's algorithm significantly outperformed DLR's algorithm in all three comparisons. Importantly, the agreement with the gold standard did not differ significantly between Penn's algorithm and both of the two visual scorings (p>0.05). Arousals had to last 15 s or longer to better reflect traditionally defined EEG awakenings [34].

In 2013-2015, we performed a pilot field study around Philadelphia International Airport (PHL) using the developed ECG and actigraphy methodology, with measurements performed unattended in order to assess the feasibility of such an approach in the NSS. In order to determine the airport for the study we examined flight operations for 4 months: from June 2012 to September 2012, for PHL. Cumulative nighttime metrics (LNight) and single event metrics (LAS,max) were predicted using the FAA's Integrated Noise Model (INM). Although the number of people exposed to high noise levels (≥55 dB LNight) was found to be low around the airport, due to the airports close proximity to UPenn and the number of night events (on average 130 events between 11:00 PM and 7:00 AM), the decision was made to conduct this pilot study at PHL. To select a control region where dwellings were not exposed to aircraft noise, GIS modeling of data from the American Community Survey was performed on the census tract level. Eighty participants were recruited, 40 from a region with aircraft noise exposure near the airport and 40 from a control region in Philadelphia County. Control region participants were comparable to the exposed group of subjects in terms of sociodemographic characteristics and non-aircraft traffic exposure, but without relevant amounts of nighttime air traffic. Each participant completed three consecutive nights of ECG and actigraphy measurements with concomitant noise level measurements and sound recordings each night in their bedroom. Additionally, participants completed brief questionnaires subjectively assessing their sleep each morning. All objective and subjective measurements were performed unattended, with staff going to the participant's home only on the first and last day of the study to setup and collect the equipment, respectively. Overall, it was found that participants were able to follow the study protocol well. For 93.4% of the nights, there were no missing periods of ECG data due to participants not wearing the device or due to improper use of the device, electrodes, or cables. For 5.7% of the nights, partial ECG recordings were obtained and for only 0.9% of nights no valid ECG data was recorded. For 89.4% of the nights, full sound recordings were obtained. Data loss was due to either equipment problems or participants failing to turn on the sound recorder at night. All questionnaires for the study were completed. The surveys were web-based which allowed staff members to verify completion of the surveys in real time and contact participants if the study protocol was not being followed.

Single event awakening analysis based on random effect logistic regression was conducted to examine whether the indoor noise level of single aircraft events ($L_{AS,max}$) was related to awakenings determined with the ECG and actigraphy. The coefficient for $L_{AS,max}$ was positive and statistically significant (i.e., higher noise levels were associated with increased awakening probability). One limitation of the derived exposure-response relationship was the wide confidence interval due to the small sample size and the comparatively low number of events per subject in this pilot study. The results of the PHL study indicated that the protocol needed further refinement for a potential future multi-site US field study on the effects of aircraft noise on sleep. While the target enrollment was met, the response rate was low, which restricts generalizability of the findings.

In 2015-2017, we performed a follow-up pilot study, around Atlanta Hartsfield-Jackson International Airport (ATL), and completed data analysis in 2019. The method and results of this study are presented in this report. The primary objective of this current study was to evaluate the feasibility of the study methodology that could be implemented in the future NSS, in particular the quantity and quality of data that could be obtained when recruiting participants by postal questionnaire, shipping them the physiological and noise measurement equipment, and the setup of the equipment and recording of data by the participants themselves, completely unattended. A secondary objective of the study was to compare objective and subjective measures of sleep and health between groups exposed to different levels of nocturnal aircraft noise.

IV. Study Methodology

A. Overview

The study was designed to assess the feasibility of obtaining in-home aircraft noise measurements and physiologic measurements of awakening from sleep, without the need for trained staff on-site. Atlanta Hartsfield-Jackson International Airport (ATL) was one of several US airports with relevant nocturnal air traffic, and chosen by the FAA as the study site for this pilot study. Modelled nighttime noise exposure around ATL and census tract demographic data were used as the basis for selecting the field study target population (section IV.C). Participants for the study were recruited by postal questionnaires (section IV.D), with a number of different mailing strategies used in order to determine how to maximize response rates (sections IV.E and V.A). Prospective study participants received one of three recruitment surveys of different length (section IV.D and Appendix 2). Field study eligibility (see section IV.E.2 for eligibility criteria) could be determined with the long and medium versions of the survey. Participants had to be re-contacted to determine eligibility for the short survey. Participants were then shipped equipment to measure aircraft noise and physiologic data during sleep (sections IV.B, IV.F), which they set up themselves in their own bedrooms. After recording five nights of data (Monday night/Tuesday morning through to Friday night/Saturday morning) and completing questionnaires each morning on subjective sleep (sections IV.F.3 and V.D), participants mailed the equipment back. Data were then downloaded and analyzed using a suite of software developed for the project in collaboration with investigators at DLR (section IV.G). Noise and ECG recordings were used to determine noise-induced event-related awakening probabilities (section V.E), with particular attention given to the efficacy of the methodology on providing usable data in the future NSS (section V.G).

The protocol of the pilot study was approved by the Institutional Review Board of the University of Pennsylvania (IRB #823726). Participants in the field study provided informed, written consent prior to taking part in the study. All private contact information for study participants was stored in a Redcap database, a secure web application designed to support data capture for research studies. Web-based community surveys were implemented through Redcap's secure system. Participant responses to paper copies of the community survey were entered separately by two staff members into Redcap's online survey database. Any discrepancy between the two data entries were resolved in consensus. For participants interested in participating in the in-home sleep study, eligibility was determined (see section IV.E.2 for eligibility criteria). Information on those participating in the in-home sleep study was stored in Redcap as well. Data were recorded on when participants were scheduled to complete measurements, which equipment was shipped to their home, when it was returned, and if there were equipment failures or damage to equipment.

B. Equipment identification and testing

For the study to be feasible on a national scale, it was important to obtain high quality acoustic and physiologic data while keeping equipment costs low . A breakdown of the equipment used in the field study is given in section IV.B.1. Equipment was tested before buying multiple units to ensure it met the required data acquisition specifications (section IV.B.2).

1. Equipment selection and cost breakdown

Study equipment (see Appendix 1) was shipped directly to participants, who unpacked and set-up equipment unattended (i.e., without research staff on site). It was therefore necessary that the noise and sleep measurement equipment we used could be set up and operated easily, with the participants able to follow simple instructions to do so, even if they did not have technical knowledge. Just as importantly, in order for the study to be feasible on a large scale, it was necessary to select recording equipment that was both low-cost and accurate in its measurement.

The H5 Handy Recorder (Zoom Corp, Tokyo, Japan) with an Earthworks M23 measurement microphone (Earthworks Inc., Milford, NH) was selected for recording acoustic data in participants' bedrooms (see section IV.B.2 for equipment testing results). Prior to shipment, the H5 recorder and microphone were fastened to a tripod and a remote control was provided to subjects for their convenience.

The Faros 90 (Bittium Corp, formerly eMotion, Oulu, Finland) was chosen to measure heart rate and actigraphy data. We have previously demonstrated the ability of the Faros 90 devices to reliably capture ECG and actigraphy data for the scoring of noise-induced awakenings among field study participants at PHL and FRA airports [35].

A total of twenty sets of equipment were prepared for use in the field study. A breakdown of equipment cost for a set of study equipment is given in Appendix 1. A single set of equipment cost $1261. In total, purchasing of study materials and testing of potential equipment designs cost $28,381. These costs do not include those for personnel, storage, or expenses for shipping the study equipment to and from study participants.

2. Equipment testing

Noise recorder testing

Prior to purchasing all twenty Zoom H5 Handy Recorders and Earthworks M23 measurement microphones, two units were purchased and tested to ensure they met the manufacturer stated specifications, and that they were suitable for accurate measurement of aircraft noise levels.

To measure the noise floor of the H5 we used the following approach. A recording was initialized, the recorder was isolated from noise by placing a cap over the microphone, sealing the recorder in a box filled with foam, and then placing the sealed box in a cupboard in the quietest room available at our laboratory. The resulting noise floor of the equipment was 22 dB(A).

Measurements were made with the H5 and compared against measurements of the same sound signal made with two Class 1 sound level meters (XL2, NTi Audio). All systems were first calibrated using a 1 kHz calibration signal at 94 dB (Larson Davis CAL200). This calibration signal was stored for the H5 recorders. As in the actual field study, the sounds recorded with the H5 were stored as MP3 files (320 bit). The stored calibration signal was used to convert these MP3 files into A-weighted sound pressure levels (see section IV.G.1 for a description of the software that was developed for this conversion). One XL2 unit was owned by us, and is hereafter termed XL2-UPenn. The second XL2 was loaned to us by the manufacturer NTi, and is hereafter termed XL2-NTI. An audio file of airplane flyovers and train pass-bys was used as the acoustic test signal, since the H5 recorders were to be used for traffic noise measurements. Sound pressure level measurements made with the H5, XL2-UPenn and XL2-NTI are presented in Figure 2. The region around the highest measured level (173-177 s) is presented in higher sound level resolution in Figure 3 for clarity. The difference in level measured with H5 and XL2-NTI relative to the level measured with XL2-UPenn during traffic noise playback is given in Figure 4. As expected there was almost no difference between both XL2 units. The noise floor of the XL2 units was around 3 dB lower than the H5. During noise measurement, there was close agreement between the H5 and XL2-UPenn, agreeing to within approximately 1.5 dB. There were very short intervals with slightly higher deviation between 160-175s (Figure 4), but at these points there were also deviations between both XL2 units. These deviations could be due to slight spatial variation in the microphone positions during measurement.

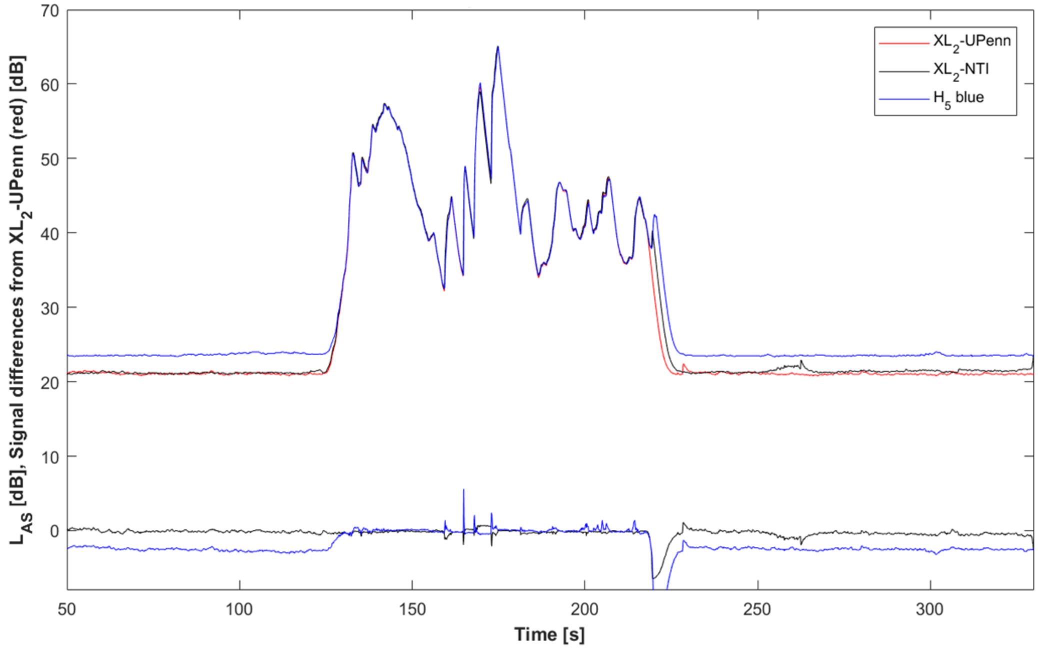


Figure 2 Measurement of traffic noise made with ZoomH5 (blue) and two XL2 sound level meters (black and blue). The upper lines represent the sound pressure level measurement (A-Weighted, slow time filter) made with each device. The lower lines represent the difference between the sound pressure level measured with H5 and XL2-NTI compared to measurements made with the XL2-UPenn. Note that the disparity between devices around 220-230 s is due to slight differences in noise cessation timing.

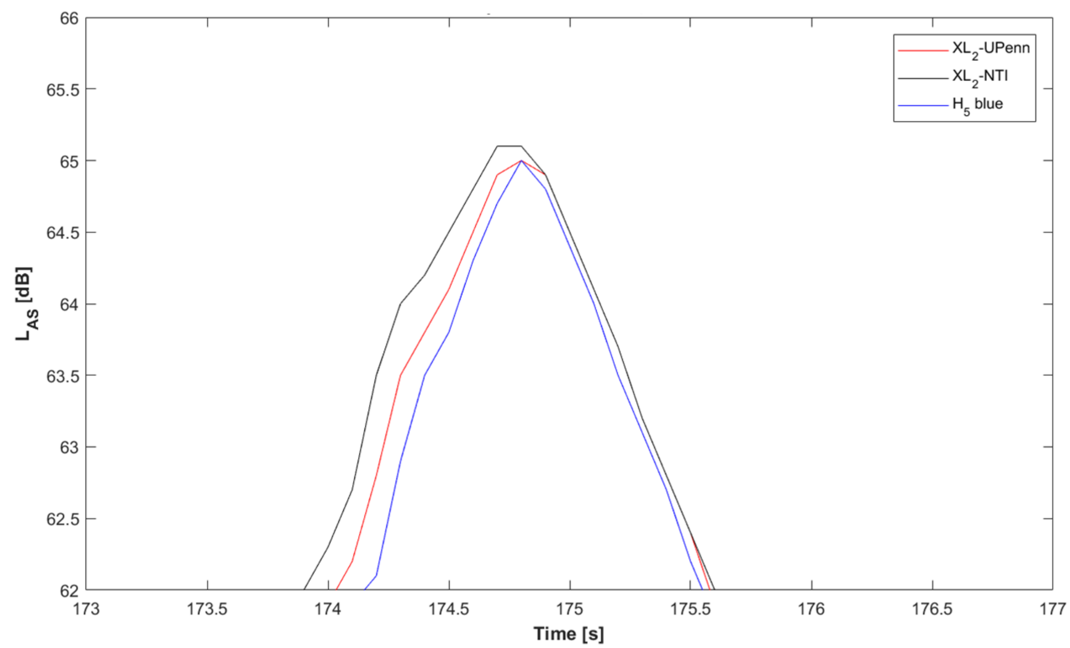


Figure 3 Measurement of traffic noise made with ZoomH5 (blue) and two XL2 sound level meters (black and blue) around the noise maximum.

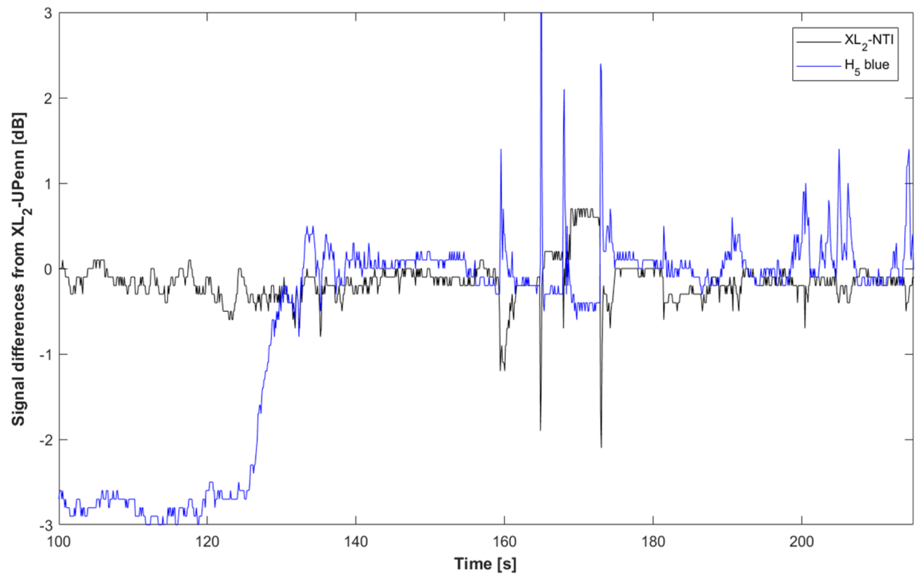


Figure 4 Difference in sound pressure level made with XL2-NTI (black) and H5 (blue) relative to measurement of the same noise signal made with XL2-UPenn.

We also compared the H5 unit used in the above measurements against a second H5 unit to examine inter-unit variability. The same procedure as above was used, and recordings were compared against those made with XL2-UPenn. The difference in level during the noise signal is given in Figure 5. Both H5 units generally agreed to within ±1 dB, which is within the tolerance limits for Class 1 sound level meters [36].

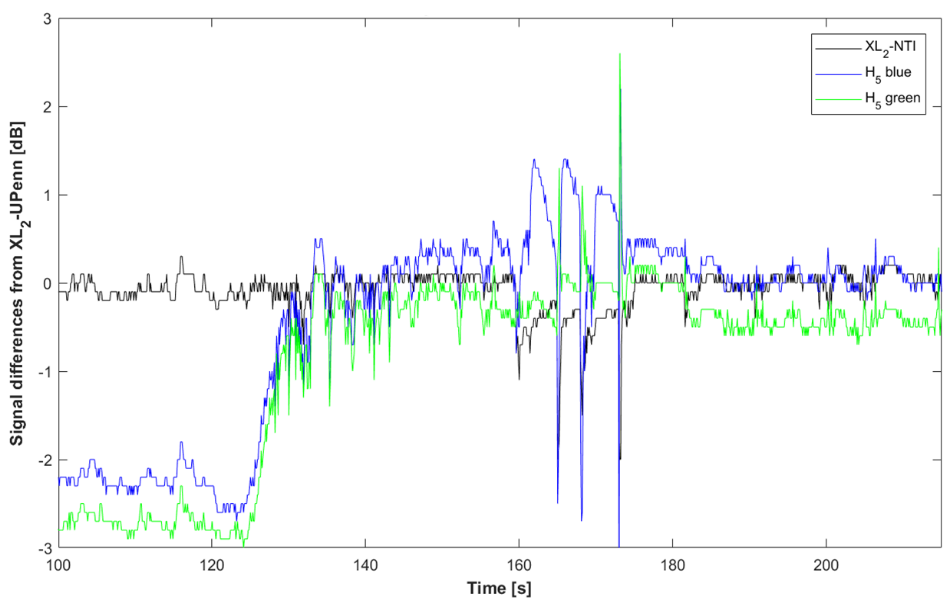


Figure 5 Difference in sound pressure level made with XL2-NTI (black), H5 (blue) and a second H5 (green) relative to measurement of the same noise signal made with XL2-UPenn. Measurements with H5 units generally agreed to within ±1 dB.

In summary, Zoom H5 recorders using with Earthworks M23 microphones represent a cost-effective approach of performing accurate measurements of aircraft noise in a field study. All microphones were calibrated by the manufacturer.

Faros and H5 time drift testing

For the event-related analysis, it is very important that acoustic and physiological events are recorded on a synchronized same timeline, so that an awakening in the physiologic data can be attributed to a concurrent aircraft event in the noise data. As we used two separate devices to record sounds (H5) and physiological data (Faros 90), we needed to ensure that there was minimal time drift between the devices, or alternatively develop a method allowing us to synchronize both data streams post-hoc. Prior to shipment to study participants, the internal clocks on the Faros 90 and H5 sound recorder were synchronized with the network time; however, study equipment was in the field for approximately 20 days, during which time was the potential for time drift in either or both devices. To investigate \time drift between the devices, we performed a study in which movement detected in the physiological data and noise events detected in the acoustic data were matched. We tested all 20 Faros 90 devices (Figure 6-Figure 8) and four H5 recorders (Figure 9). We also tested an updated version of the Faros device, the Faros 180, for comparison (this device was not used in the ATL study, but may be used in future studies). The Faros devices and H5 recorders were initialized with the network time and then powered off. They were kept in a cool location for 1 week, simulating the time devices are in transit to participants. After 1 week, the Faros 90 and 180 were placed on a rotating table that rotated the devices at fixed intervals. The start of the rotation was indicated by a clicking sound which was recorded by the H5 sound recorders. We recorded differences in the event times, relative to the network master clock, between the acoustic and physiologic data throughout the 5 study days. These recordings were completed under a variety of test conditions to simulate common scenarios expected in the field. Recordings on the Faros devices were made in either a room-temperature environment (23 ˚C, Figure 6) or in a warm room (35 ˚C, Figure 7). In both the warm and cold room scenarios, the Faros devices recorded for 8 hours per day, simulating an anticipated 8 hour recording of sleep during the field study, and were turned off for the remaining 16 hours. Additionally, we also examined the time drift when the Faros devices were left running for the duration of the simulation (Figure 8), i.e. not turned off for the 16 hours each day, as subjects may forget to turn off the devices in the morning.

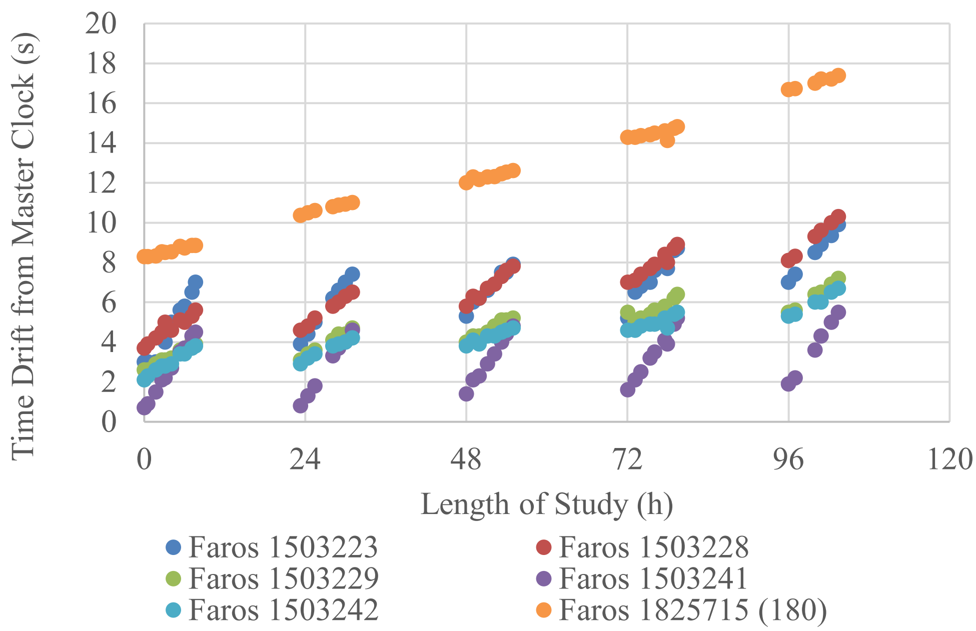


Figure 6 Time drift between master clock and five Faros 90 and one Faros 180 recorder internal clocks, recorded at room temperature (23 ˚C). Different colored points indicate different Faros units.

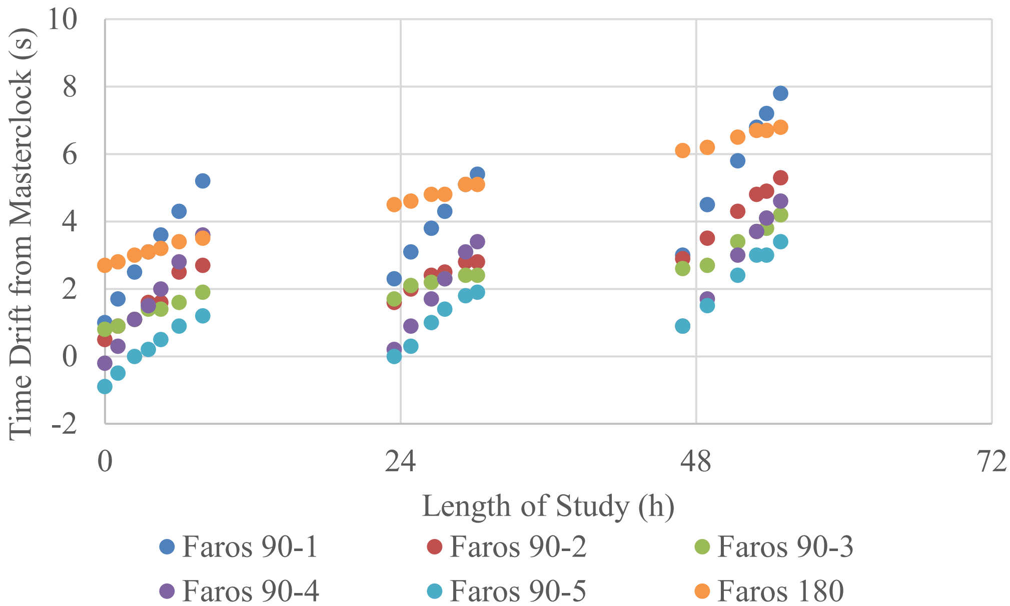


Figure 7 Time drift between master clock and five Faros 90 and one Faros 180 recorder internal clocks, recorded in a warm room (35 °C). Different colored points indicate different Faros units.

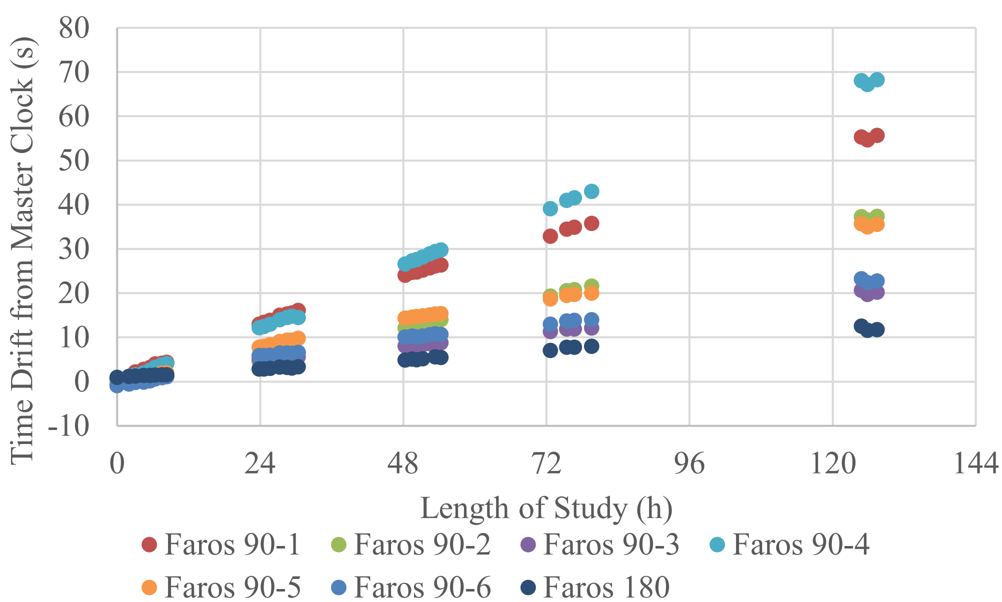


Figure 8 Time drift between the Faros devices and the master clock for the simulation in which the Faros are left running for the duration of the simulation, mimicking the scenario in which a participant forgets to turn off the Faros prior to charging. At the end of each day, the Faros were plugged into a charging port but continued running. It was found that the Faros 180 automatically turns off when plugged into a power source, and so this device did not run continuously during the simulation. At 96 hours of the simulation, technical staff were unavailable to run the simulation, and so the Faros were not rotated on the rotating table until hour 120 of the simulation.

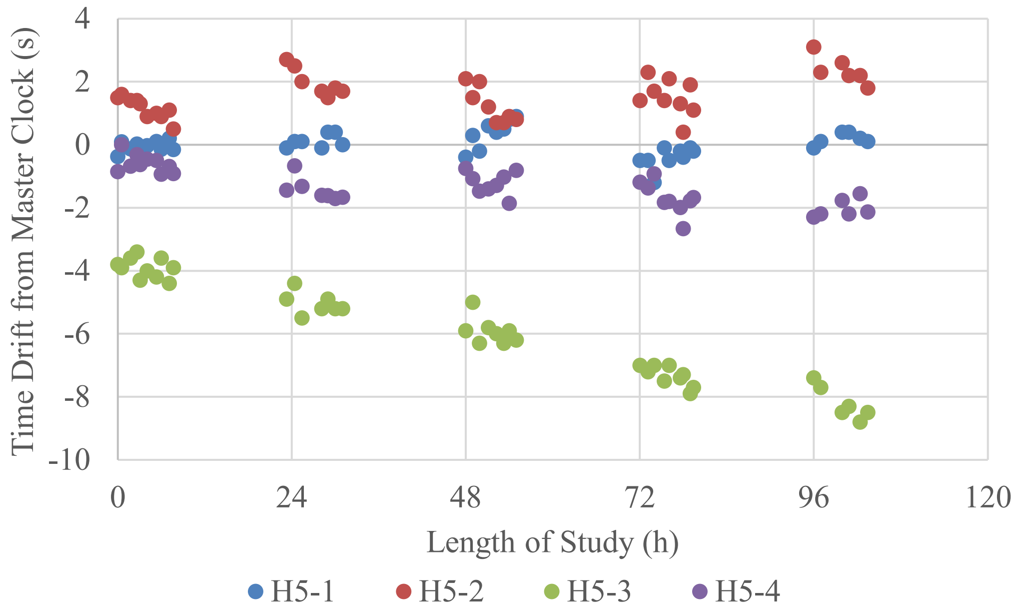


Figure 9 Time drift between master clock and H5 recorder internal clocks. Different colored points indicate different H5 units.

It was found that the Faros 90 clocks drifted approximately linearly within recording nights, both at room temperature and in a warm room, but did not appear to drift between nights when the devices were turned off. The Faros 180 clock also drifted linearly within recording nights, but continued drifting between nights when turned off. Out of the four H5 sound recorders that were tested, three drifted approximately ±2 seconds from the master clock in a seemingly random pattern. A fourth H5 recorder drifted approximately 4 seconds from the master clock during the simulated transit week, and drifted a further 4 seconds during the five recording nights in a linear fashion. When switched on (and therefore recording) for extended periods of time, the Faros devices were found to continue to drift linearly for the duration of the simulation. Based on this evidence of time drift between the acoustic and physiologic data streams from our simulations, our DLR collaborator, Dr. Uwe Müller generated a time-synchronization software that matches body movements scored in the acoustic data with the body movements recorded in the physiological data (see section IV.G.3). Based on the simulation results above, a linear time drift across the measurement night was assumed for correction purposes.

C. Selection of field study target sample region

The purpose of the field study was to investigate effects of aircraft noise on sleep. It was therefore necessary to stratify the sample population by nighttime aircraft noise exposure levels, so that recruitment from appropriate regions could be performed.

1. Generating and validating noise contours around ATL airport

Noise exposure around ATL was modelled using the FAA's Integrated Noise Model (INM) [37], implemented using the ArcGIS software (Esri, Redlands, CA).

Radar track data and flight plan data from the Performance Data Analysis and Reporting System (PDARS) [38] around ATL were provided by the FAA for the period of September 1st 2014 to August 31st 2015. Along with runway location and orientation, the PDARS data were used to model individual nighttime (23:00-07:00) aircraft noise events over 84 nights. These noise data were used to calculate outdoor nighttime A-weighted noise level ($L_{Night,outdoor}$) contours around ATL. These modelled contours are presented as filled contours in Figure 10.

To validate the modelled contours, they were visually compared with yearly average L_{Night} contours from 2012 for 45, 50, 55 and 60 dB, which were also provided by the FAA. These are presented as lines in Figure 10. There was a good agreement between the FAA contours and our own modelled contours.

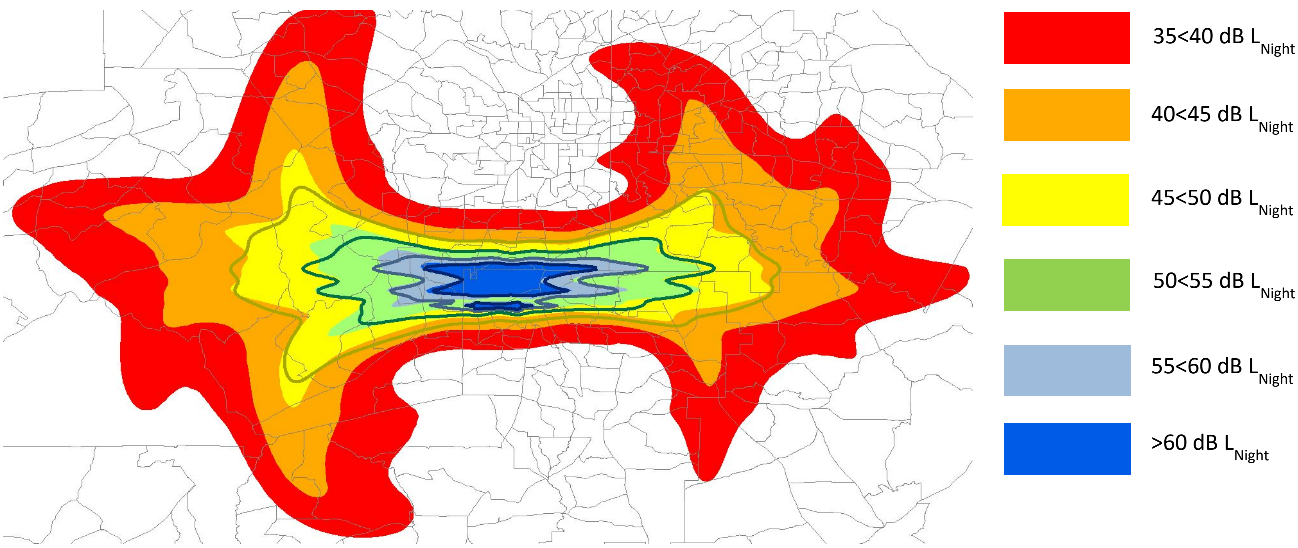


Figure 10 L_{Night} noise contours around ATL. Filled contours represent those calculated by UPenn. Line-only contours represent the 2012 average, provided by the FAA, and used only to validate the UPenn contours. Contours are overlaid on Atlanta census tract geographical boundaries.

Since L_{Night} was the primary exposure variable of interest, it was necessary to sample the study population from addresses with different noise exposure. We therefore stratified into five sampling regions: 35-39.9 dB, 40-44.9 dB, 45-49.9 dB, 50-54.9 dB and ≥55 dB. This stratification was performed based on the UPenn contours since the FAA contours had a lower limit of 45 dB L_{Night}, as compared to the UPenn contour lower limit of 35 dB L_{Night}.

2. Population sampling procedure

Geographical shape information for the census tracts in and around Atlanta were extracted from TIGER/Line® Shapefiles (https://www.census.gov/geo/maps-data/data/tiger-line.html). These shapefiles are an extract of selected geographic and cartographic information from the U.S. Census Bureau's Master Address File / Topologically Integrated Geographic Encoding and Referencing (MAF/TIGER) Database (MTDB). Demographic data for these census tracts were extracted from

American FactFinder (http://factfinder2.census.gov/). For each census tract in each noise exposure category, the population weighted centroid was calculated using the extracted geographical and demographic information. The noise levels at each centroid were then calculated, before assigning the census tract into the 35, 40, 45, 50 or 55 dB L_{Night} category. The resulting assignment of each census tract is shown in Figure 11.

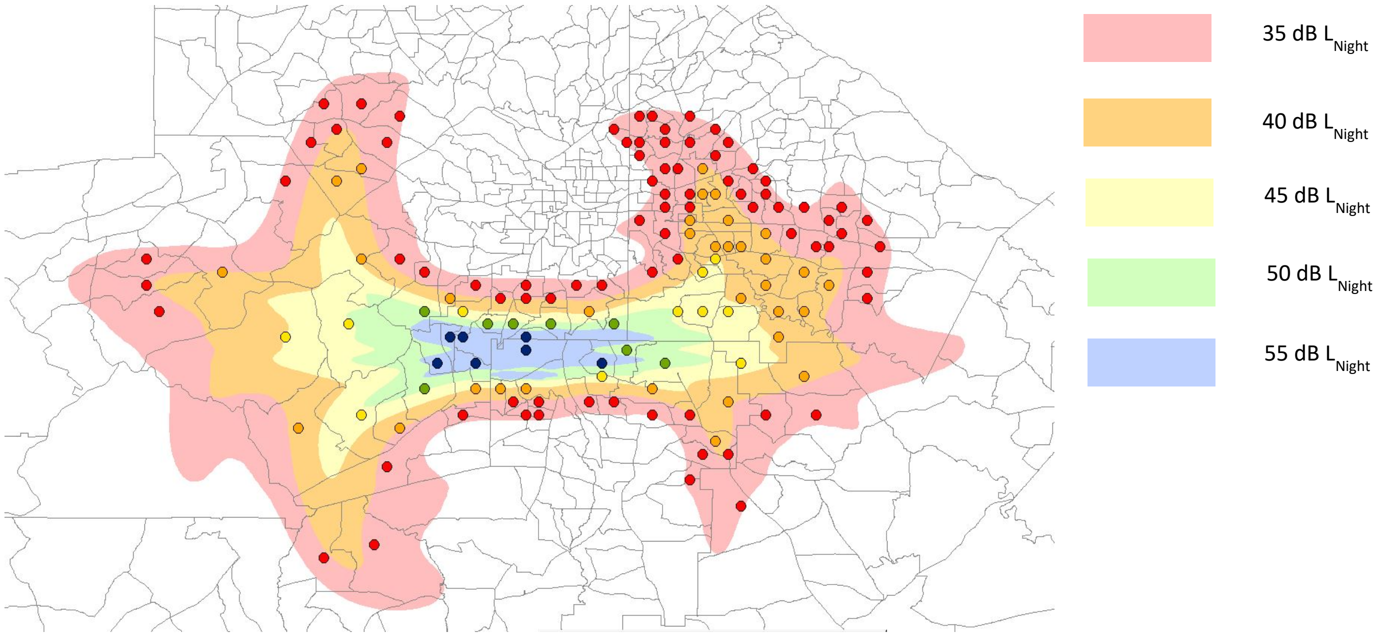


Figure 11 Population weighted centroid of each census tract, colored according to noise exposure category (noise contour) in which it is located.

In addition to classifying census tract by noise exposure, they were further sub-divided into their orientation relative to ATL airport, either west or east. The location of the population weighted centroid of each census tract relative to the airport coordinate (33.640444˚ N, 84.4269444˚ W) was used to assign whether the census tract was east or west of the airport. The number of census tracts in each noise exposure category is given in Table 2. Demographic data from the 2010-2014 American Community Survey 5-Year Estimates for the census tracts in each noise exposure category are given in Table 3.

Table 2 Number of census tracts in each noise category

Noise category	n	West	East
≥55 dB	5	4	1
50-54.9 dB	8	4	4
45-49.9 dB	11	4	7
40-44.9 dB	34	10	24
35-39.9 dB	79	22	57
Total	137	44	93

Table 3 Demographic characteristics of census tracts within each noise category.

Noise category	Direction re: ATL	Houses (n)	% No College Education	% Black or African American	Mean income ($)	Mean age (years)	Mean house price ($)
≥55 dB	East	1949	59.5	55.1	33,624	29.4	60,300
≥55 dB	West	7305	50.8	90.7	26,737	29.4	105,975
50-54.9 dB	East	9464	59.7	59.6	31,126	30.7	78,950
50-54.9 dB	West	11,123	34.8	77.3	40,938	35.5	161,200
45-49.9 dB	East	14,489	46.3	83.6	46,964	35.4	102,971
45-49.9 dB	West	20,457	32.2	32.2	59,955	35.1	138,625
40-44.9 dB	East	53,391	41.9	77.4	50,249	38.4	126,300
40-44.9 dB	West	30,674	45.1	81.2	39,677	30	101,260
35-39.9 dB	East	118,182	35.7	52.7	50,684	35	182,782
35-39.9 dB	West	55,842	41.1	58.5	54,040	36.6	139,109

The 35-39.9 dB category was the control region for the study. The cost for obtaining addresses was $50 for each census tract. To minimize cost we selected 16 census tracts from the 35-39.9 dB category (8 west and 8 east). These 16 control region census tracts were chosen so as to have a similar mean and variance of household income as in all 79 census tracts in the <40 dB category (Table 4).

Table 4 Demographic characteristics of 35-39.9 dB census tracts. Demographics of the census tracts selected as the control region are highlighted.

Direction re: ATL	Houses (n)	% No College Education	% Black or African American	Mean (M), range (R) and standard deviation (SD) income ($)	Mean age (years)	Mean house price ($)
East (n=57)	118,182	35.7	52.7	M: 50,684 R: 14,879-136,813 SD: 25,689	35.0	182,782
East, selected (n=8)	12,300	47.1	36.8	M: 50,376 R: 14,879-92,000 SD: 25,710	35.1	156,157
West (n=22)	55,842	41.1	58.5	M: 54,040 R: 24,129-103,333 SD: 19,177	36.6	139,109
West, selected (n=8)	22,302	38.4	60.2	M: 54,302 R: 37,446-83,969 SD: 19,191	35.7	148,450

Once the 74 census tracts from which we would sample was finalized, 10,000 residential addresses and inhabitant names within these tracts were purchased from MSG Marketing Group at a cost of $1,325 ($425 initial setup cost, $50 for each of the 9 additional survey tracts, and $450 for the 10,000 address-based sampling records). Each address was provided with its associated latitude and longitude. L_{Night} was then calculated for each individual address. Addresses were reclassified into the appropriate noise categories based on these L_{Night} noise levels and not based on the census tract population weighted centroid noise levels.

D. Postal surveys

Postal questionnaires are an inexpensive and unobtrusive method of data sampling among large study populations, and so are widely used in epidemiological research. One of the challenges faced by public health research is the current trend for decreasing response rates to all survey modes [39], which leads to reduced effective sample sizes, and furthermore may bias the acquired data and subsequent conclusions [40]. To minimize threats to internal and external validity, the highest attainable response rate is therefore desirable. Researchers have adopted a number of methods to improve response rates, which include monetary and non-monetary incentives, changes in the length and appearance of questionnaires, different methods of returning completed questionnaires, pre-notification and different approaches to follow-up contact [41]. Reduced survey length, the use of incentives and follow-up contact for postal surveys can improve response rates, but these findings are not found universally across different studies [41, 42]. There is also a risk that incentives may introduce bias, by being more appealing to those with lower socioeconomic status [43]. Survey follow-up and incentivization also increases methodological expense, although this may be offset by the reduced need for further sampling from a study population to obtain an equivalent sample size.

Postal questionnaires can, in addition to furnishing researchers with valuable epidemiological data, serve as useful pre-screening instruments. Pre-screening questionnaires can determine a person's eligibility for, as well as their interest in, recruitment into later studies, although when relying on self-report there can be some risk for respondents to misrepresent themselves so that they can participate in the study [44]. Low response rates for questionnaires used for pre-screening may lead to non-representative sample populations in any subsequent studies, so it remains important to obtain the greatest achievable number of responses. For this pilot study, we therefore adopted a number of different survey strategies in order to determine how to maximize survey response and field study recruitment while minimizing cost.

1. Survey instruments

The primary purpose of the postal surveys was to recruit participants for the field study on the effects of nocturnal aircraft noise on sleep. Of primary importance therefore were questions regarding suitability as pertains to the study inclusion criteria (see section IV.E.2). The survey included a checkbox for respondents to indicate whether they were interested in participating in the field study, along with their contact details.

The secondary purpose of the surveys was comparison of eventual field study participants with non-participants. This allows for determining whether those who are eligible for the field study are representative of those who respond. This comparison can potentially inform weights to adjust for non-response bias.

Of tertiary importance in the questionnaires were items regarding the effects of noise on annoyance, sleep disturbance and health outcomes, to allow a cross-sectional analysis of community response to aircraft noise. The addition of these items increased the questionnaire length, which as a result could risk lowering response rates, while at the same time providing useful data on the effects of aircraft noise. We therefore used questionnaires of different lengths to investigate if longer questionnaires had a significant adverse effect on response rate.

Survey instructions indicated that only a single household member should fill out the survey (the person who most recently celebrated a birthday). Complete versions of the questionnaires are given in Appendix 2, and are only summarized here. Questionnaires differed in length and were characterized as short (11 questions), medium (26 questions) or long (57 questions). The long form of the survey asked respondents to provide basic demographic information, such as age, sex, race, income, marital status, education level, and employment status. Respondents were asked to rate their overall sleep quality on a 4-point Likert-type scale over the past month, which is an item taken directly from the Pittsburgh Sleep Quality Index (PSQI) [45]. They also indicated how often (on a 4-point scale from "not during the past month" to "three or more times per week") they experienced trouble falling asleep, waking up in the night or early morning, took medication for sleep, or had difficulty staying awake during the day, all of which are items from the PSQI. The survey asked about coping behaviors to environmental noise. Survey respondents were asked to estimate over the past month how often (on a 5-point Likert-type scale from "never" to "always") they "wear earplugs," "use alcohol," "use medication," "turn on the TV," "turn on music," "close windows," "use a sound machine," or "turn on a fan" because of noise when trying to sleep. Sensitivity to noise in the community was another variable examined, and respondents were asked to estimate on a 6-point ordinal scale their agreement with statements: "I am easily awakened by noise," "I get used to most noises without difficulty," "I find it hard to relax in noisy places," "I am good at concentrating no matter what is going on around me," "I get mad at people who make noise," and "I am sensitive to noise." All of the noise sensitivity questions and response scales were taken from the Weinstein Noise Sensitivity Scale [46]. Also, participants were asked to describe how much they were annoyed over the last 12 months (on a 5-point Likert-type scale with endpoints "not at all" and "extremely", per recommendations by the International Commission on the Biological Effects of Noise (ICBEN) [47]) to “road traffic,” “trains,” “aircraft,” “industry/factory,” “construction,” “neighbors,” and “air conditioner” noise. They also indicated on the same ICBEN scale how often their sleep was disturbed by those noise sources over the past 12 months. Respondents estimated their general health on a 5-point Likert-type scale (poor to excellent) and indicated if they had ever been diagnosed with any of the following sleep disorders: sleep apnea, periodic limb movement syndrome, narcolepsy, insomnia, or restless leg syndrome. Participants also reported any diagnosis of hypertension, migraines, arrhythmia, heart disease, stomach ulcer, or diabetes, and indicated whether they had received treatment in the past month.

The short and medium questionnaires did not include the items on habitual sleep and wake times, frequency of sleep difficulties, expanded noise sensitivity, annoyance by traffic, industry and community noise, diagnosis and treatment for a number of the medical conditions, marital status, income, education level, employment status or residence sound proofing treatment. Furthermore, the short form questionnaire did not include items on sleep medication, sleep disorders, sleep-promoting coping strategies, hearing acuity, diagnosed hypertension and/or arrhythmia, shift work, residence duration, household children, height or weight.

The medium and long versions were sufficiently comprehensive to determine whether a respondent met the field study inclusion criteria, but the short version required us to contact the respondents via telephone for additional information. This telephone contact was only done if the respondent indicated that they were interested in participating in the study and as such gave permission to be contacted.

E. Field study participant selection process and recruitment

1. Survey protocol

Between September 2016 and July 2017, we sent paper surveys along with a letter of introduction to 4080 randomly selected households around ATL. The introduction letter, provided in Appendix 3, briefly described the purpose of the survey, informed the recipient that participation was voluntary, assured the confidentiality of their responses, and provided contact information for the research group. Also provided was the survey eligibility criteria: 21 or more years of age and only one respondent per household, preferably the adult whose birthday was most recent. Respondents returned surveys by mail using an included pre-paid addressed envelope, or completed them online by following a URL or scanning a QR code.

The surveys indicated the financial compensation that would be awarded for participating in the field study (which varied between $100, $150 or $200; see below), and included items on whether respondents would be interested in taking part in such a study.

Surveys were sent in batches of 240 in seventeen mailing rounds. An equal number of surveys were sent to each noise exposure category within each round (24 surveys to each of the 10 noise exposure categories). Mailing rounds differed in the incentive for completing the survey, the length of the survey, the number of follow-up (reminder) waves issued after the initial mailing, and the monetary incentive for participating in the field study if eligible (Table 5). The incentive for completing the survey was either $2 cash included in the initial survey mailing wave, or an Amazon gift card of $2, $5 or $10 value provided upon completion of the survey. The United States Postal Service could not always deliver the surveys to the listed address. We classed a survey as “non-deliverable” if at least one survey, from any wave within a round, was returned to sender. Such reasons for returning to sender included vacant address, unable to be forwarded, incorrect address or reasons unknown. The percentage of surveys that were deliverable within each mailing round are given in Table 5. On average, (87.6%) of the surveys were deliverable. If a completed survey was received for a recipient that had been classed as non-deliverable (n=9), we reclassified the survey as deliverable. A number of surveys were returned to the sender because the recipient was deceased (n=1), refused delivery of the survey (n=23) or returned a blank survey indicating they were not interested (n=5): these instances were classed as deliverable but as non-response.

Prior to the initial survey wave, a pre-survey notification postcard was sent out only in round 5. Following the initial survey wave within each round, there were 0, 2 or 3 follow-up waves sent if a completed survey had not yet been received from a specific household. The first follow-up, sent 7 days after the initial survey, consisted of a postcard encouraging the recipient to return and complete the original survey if they had not yet already done so. The second follow-up, sent 21 days after the initial survey, consisted of a reminder letter, a new paper copy of the survey and a new pre-paid envelope for returning the survey. The third follow up, sent 42 days after the initial survey consisted of a reminder letter, a further new paper copy of the survey and a further new pre-paid envelope for returning the survey.

Mailing rounds 1-2 were addressed to “Current Resident” and rounds 3-17 were personalized and addressed to a named individual or current resident, for example “A. N. Onymous or Current Resident”. Rounds 1-2 were mailed in envelopes measuring 24×10.5 cm, and rounds 3-17 were sent in 23×15.5 cm envelopes. In addition to a University of Pennsylvania logo on the envelope of all mailing rounds, rounds 1-2 indicated that “Perelman School of Medicine, University of Pennsylvania, Department of Psychiatry, Division of Sleep and Chronobiology” sent the mail, and rounds 3-17 indicated only “University of Pennsylvania” as the sender.

Table 5 Overview of each survey round.

Round	Incentive for completing the survey	Survey length	Number of follow-up waves	Incentive for participating in field study	Addressee	% deliverable
1	Gift card	Long	0	$100	“Current Resident”	91.3
2	Gift card	Long	0	$100	“Current Resident”	92.9
3	Gift card	Long	0	$100	Personalized	91.7
4	Gift card	Long	0	$100	Personalized	88.8
5	Gift card	Long	0†	$100	Personalized	91.3
6	$2 cash	Long	3	$150	Personalized	88.3
7	$2 cash	Long	3	$150	Personalized	89.6
8	$2 cash	Medium	3	$150	Personalized	87.5
9	$2 cash	Short	3	$150	Personalized	86.3
10	$2 cash	Long	3	$200	Personalized	84.6
11	$2 cash	Long	0	$200	Personalized	91.3
12	$2 cash	Long	3	$200	Personalized	85.0
13	$2 cash	Long	3	$200	Personalized	86.3
14	$2 cash	Long	2	$200	Personalized	85.4
15	$2 cash	Long	2	$200	Personalized	84.2
16	$2 cash	Long	2	$200	Personalized	83.8
17	$2 cash	Long	2	$200	Personalized	82.1

† Included pre-survey notification postcard sent before the initial survey mailing

2. Recruitment into field study

Upon receiving completed surveys where respondents indicated they were interested in participating in the field study, responses were checked to see whether an individual was eligible for the field study. In the case of short survey respondents, follow-up contact via telephone was required to determine eligibility.

Exclusion criteria were as follows:

- Use of medication (either prescribed or "over-the counter") to help with sleep three times or more per week, over the past month.
- Diagnosed by a heath professional with any sleep disorder, including but not limited to the following: sleep apnea, narcolepsy, restless leg syndrome, period limb movement syndrome, insomnia.
- Diagnosed by a heath professional with arrhythmia.
- Self-reported problems or difficulties with hearing.
- Overnight shift work, defined as working for at least 4 hours between 00:00 to 06:00.
- Under 21 years of age.
- Any children in the household under 5 years of age.
- Body mass index (BMI) of >35 or <17 kgm^{-2}, corresponding to classification as Obesity Class II ("severely obese") and moderately underweight respectively [48].

Out of 407 completed surveys, 237 respondents (58.2%) were interested in participating in the field study. Among respondents interested in the field study, 79 respondents (19.4% of all completed surveys, 33.3% of those interested) met the eligibility criteria. Of those interested and eligible, 37 respondents (9.1% of completed surveys, 15.6% of those interested) were enrolled into the field study. Three participants dropped out before the study commencement. Demographic data of the 34 remaining participants who completed the study are given in Table 6. Further analysis on the effectiveness of the different survey protocols for eliciting questionnaire response, interest for participating in the field study, and eventual participation in the field study, are given in section V.A.

Table 6 Demographics of participants completing the field study.

Variable		Mean (±S.D.)	Range
Age, years (n=34)		50.2 (±14.7)	21-81
BMI, kgm^{-2} (n=34)		27.0 (±3.25)	21.8-33.5
Categorical variable	**Level**	**Count (n)**	**% of responses**
Sex (n=34)	Women	22	64.7
	Men	12	35.3
General health (n=34)	Poor	1	2.9
	Fair	2	5.9
	Good	8	23.5
	Very good	18	52.9
	Excellent	5	14.7
Race (n=34)	White	11*	32.4
	Black	19	55.9
	Other	3*	8.8
	Prefer not to answer	2	5.9
Marital status (n=23)	Single	11	47.8
	Married	6	26.1
	Widowed	1	4.3
	Separated	1	4.3
	Divorced	3	13.0
	Dom. Partner	1	4.3
Education (n=23)	< High school	0	0
	High school	9	40.9
	College or more	13	59.1
Job status (n=23)	Employed	15	65.2
	Unemployed	2	8.7
	Retired	6	26.1
Household income (n=23)	<$25k	5	21.7
	$25-50k	6	26.1
	$50-75k	4	17.4
	$75-100k	2	8.7
	$100-150k	2	8.7
	>$150k	2	8.7
	Prefer not to answer	2	8.7

* One participant listed race as both White and Other and is counted for both categories.

F. Field study procedure

1. Telephone recruitment

Survey respondents who indicated that they would like to be contacted about participating in the in-home sleep study were contacted by telephone. These prospective participants were read a script detailing the study length, procedures and compensation. They were informed that the study was a 5 consecutive night, in-home, unattended sleep study, and that sounds inside the bedroom would be recorded at night using a sound recorder. Participants would wear a small device attached to two electrodes that would measure heart rate and body movement. In the morning, study participants complete a brief questionnaire concerning their sleep. The eligibility of prospective participants was verified. Those determined ineligible according to exclusion criteria were informed that they did not meet eligibility criteria for the in-home sleep study, and thanked for their time. Eligible participants were mailed an informed consent form for their review together with a pre-paid return envelope. Prospective participants who completed and signed a consent form were called and scheduled for participation in the in-home study.

2. Field study procedures

Unpacking Study Equipment

Study equipment was shipped directly to participants by staff (Figure 12B). Participants received an instruction manual detailing step-by-step instructions for setting up the equipment and completing measurements. Included in the manual was a link for online-instructional videos on how to unpack and setup the equipment. Participants were called on the first and last day of the study to review procedures and answer questions. Participants were encouraged to call the 24 hour hotline to contact staff for questions regarding study procedures. Also included in the equipment package were five copies of morning surveys (Appendix 4), a photocopy of their signed consent form, return shipping instructions, and forms for payment. Participants were instructed to setup the sound recorder on the first evening (Monday), at any time prior to bedtime. For five consecutive nights (Monday to Friday), immediately before going to bed, they would put on and start the heart rate device, and begin the recording on the sound recorder. On each of the following mornings (Tuesday to Saturday), they would stop the sound recorder, stop and remove the heart rate device, and complete the morning survey. During the day after the final study morning (Saturday), the participants would then pack up and return the measurement equipment.

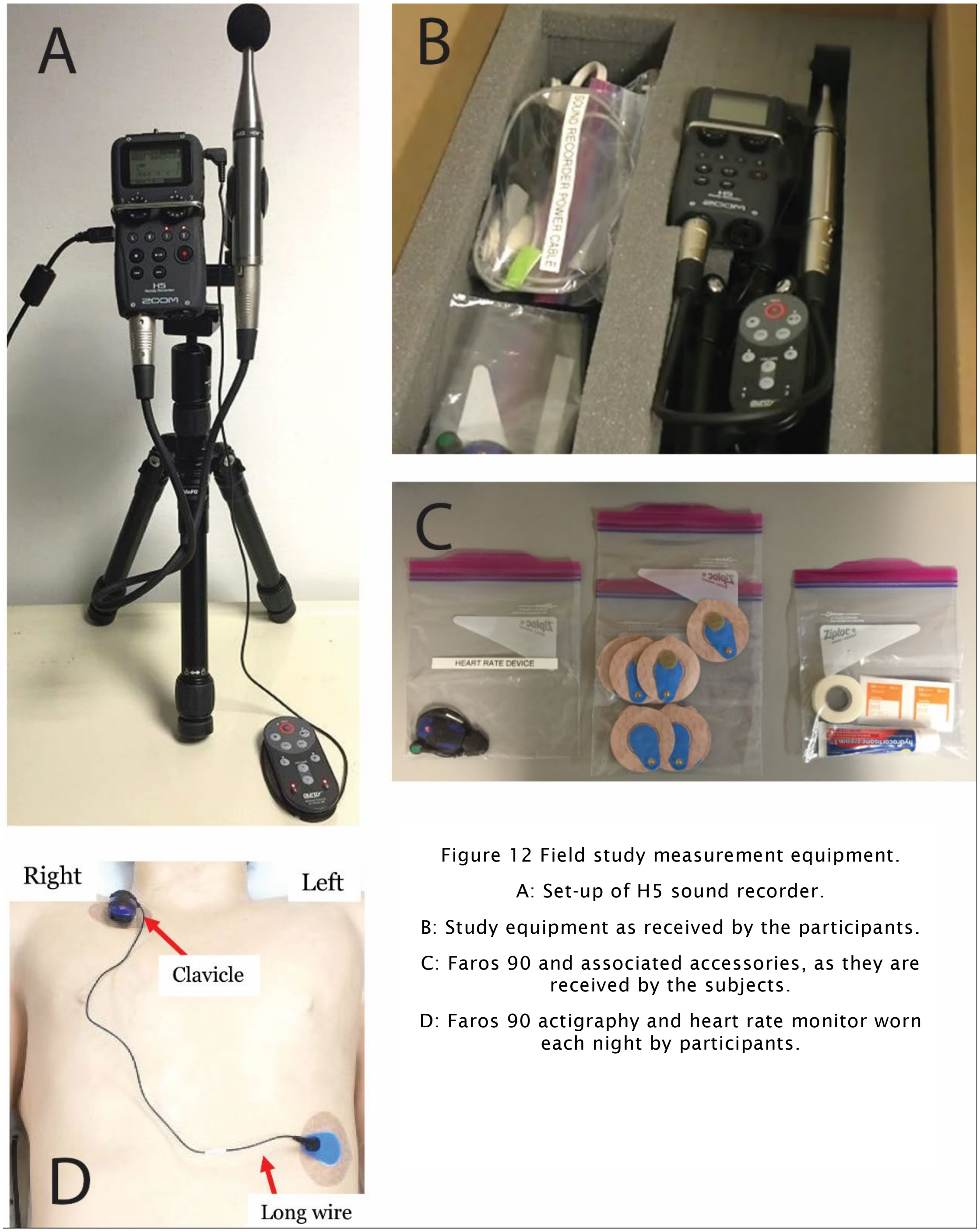


Figure 12 Field study measurement equipment.

A: Set-up of H5 sound recorder.

B: Study equipment as received by the participants.

C: Faros 90 and associated accessories, as they are received by the subjects.

D: Faros 90 actigraphy and heart rate monitor worn each night by participants.

Setting up the Sound Recorder:

Participants were allowed to sleep at their normal times and wake up at their normal times each day. They were asked to turn off any noise producing items such as the TV, radio, or music during the night. However, in order to preserve a typical sleeping environment, participants were allowed to turn on fans, air conditioners and heaters for their comfort. Also, participants were allowed to sleep with their pets (such as dogs and cats) as they would have normally in their bedrooms. It was desired to have participants maintain as close to their normal sleep routine as possible. Participants were instructed to place the sound recorder near where they slept at night, preferably on a night stand near their head, and to keep the recorder plugged in during measurements (Figure 12A). An extension cord was provided in case it was required. A remote control was supplied for convenience in turning the recorder on/off. The recorder was to be turned on before getting into bed and turned off once awake in the morning.

Setting Up the Heart Rate Device:

During the night, participant's sleep was monitored using one device (eMotion Faros 90) which measured both heart rate and body movements. The device was battery powered and attached with two electrodes to the chest of the subjects. The ECG was sampled at 1 kHz and the peak of each R-wave was detected and recorded. Movement was also measured using a 3-axis accelerometer at a sample rate of 10 Hz, 14 bit resolution, range set to 2 g. As movement was recorded with a high resolution, breathing patterns could be inferred from movements of the chest and it could be determined whether participants had chest movements that would be suggestive of sleep apnea during the night.

Along with the Faros 90 device, participants received a charger, electrodes, tape, alcohol wipes, and cortisone cream in case of skin irritation from the electrodes (Figure 12C). Participants were instructed to place one electrode just below the right clavicle, and another below the left breast (Figure 12D). The heart rate device snaps onto the electrode below the clavicle and the cable snaps onto the bottom electrode. The device is secured with Velcro and medical tape is supplied for extra security if needed. Participants were instructed to turn the device on when they get in bed, and turn it off when waking up. They were instructed to charge the device every morning after awakening.

3. Morning survey

On each morning after measurements took place, participants were instructed to fill out a short questionnaire on the previous night's sleep. Surveys could be completed either online or on the provided hard copies (Appendix 4). The morning survey asked participants at what time they went to sleep, how long it took them to fall asleep, and how many times (if any) they woke up during the night. They were also asked about their quality of sleep, how refreshed or tired they felt in the morning, and whether they felt disturbed by environmental noise during the night. Online morning surveys were checked daily and participants contacted if survey comments mentioned difficulties or concerns with equipment.

To ensure accuracy of the data when coding the paper versions of the morning questionnaires, we adopted an approach to minimize human error. The responses indicated on the questionnaires were manually entered into RedCap by two or three different investigators using the same coding scheme. An automated algorithm was then implemented to check for any discrepancies between the entered data. If a discrepancy was identified, i.e. at least one of the investigators had entered a value that did not match exactly with the entries of the other investigators, the data point was cross-checked against the original questionnaire and the correct value entered.

4. Returning study equipment

After completing five nights of measurements, participants were instructed to pack all equipment back into the shipping box. Photos of how the box should appear when properly packed were included for their assistance. Participants filled out their personal information on payment forms in order to receive compensation for participating in the study. Return shipping instructions indicated the FedEx phone number and shipping order number to schedule an at-home pick-up of study equipment. Subjects could also drop off the equipment at any location that accepts FedEx shipments.

G. Data Analysis

1. SPL converter

The H5 recorder used in the in-home sleep study records noise in mp3 format. Acoustic data from the field study thus had to be converted from mp3 to sound pressure level (SPL) prior to analysis. A sound pressure level converter program was developed to calculate the correct A-weighted sound pressure levels with fast (0.125 s) and slow (1 s) time constants (L_{AF} and L_{AS} respectively), for a given mp3 file using an existing calibration file for each measurement. Calibration files (1 kHz at 94 dB) were recorded prior to shipment into the field study, and again upon return.

First, the L_{AS} and L_{AF} of the initial and final calibration files were calculated (Figure 13). If the deviation between the two calibration files was less than ±2 dB, then the SPL for the measurement was calculated. In total, of the data of 9 subjects were excluded from the analysis due to large deviations in the pre- and post-calibration files. This deviation was due to shifting in the dials of the sound recorders, and was remedied for future subjects by securing the dials in a fixed position with adhesive prior to shipment.

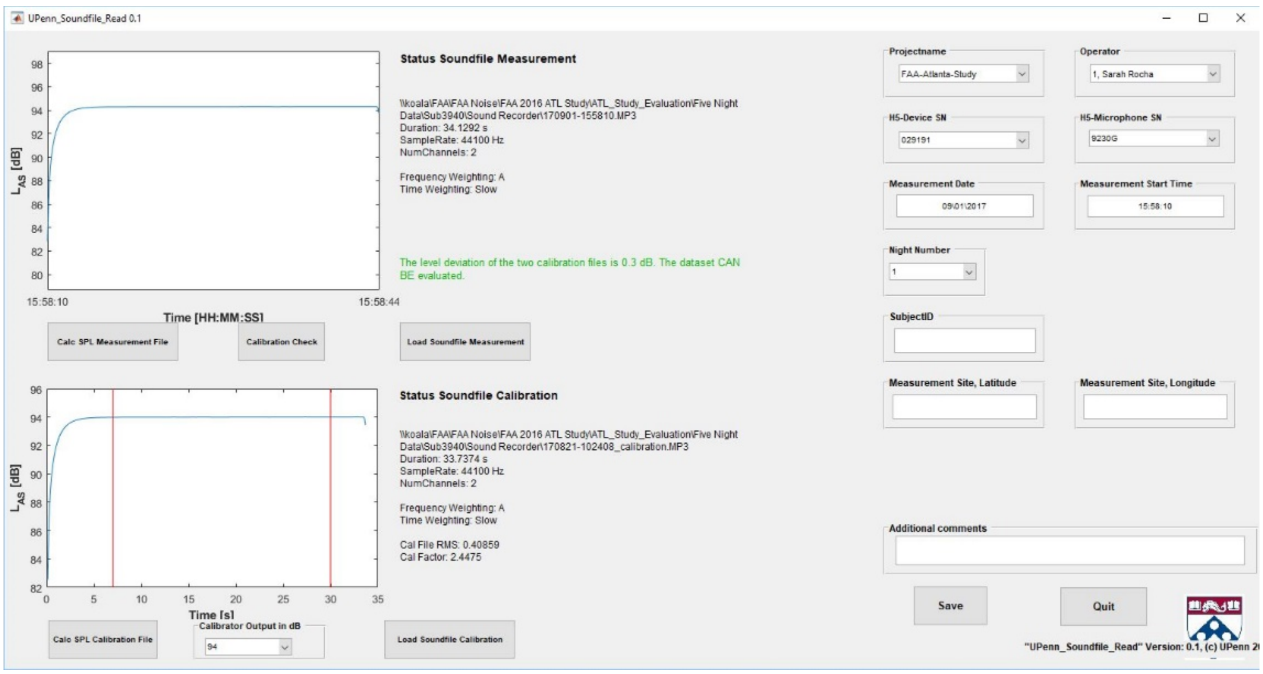


Figure 13 Sound Pressure Level Converter compares the initial and final calibration files for a given subject

Next the program calculated the L_{AS} and L_{AF} of the measurement file using the calibration file and the calibrator output value. The converted sound pressure level could then be scored for aircraft noise in the acoustic scoring program, Akustikview (see section IV.G.2).

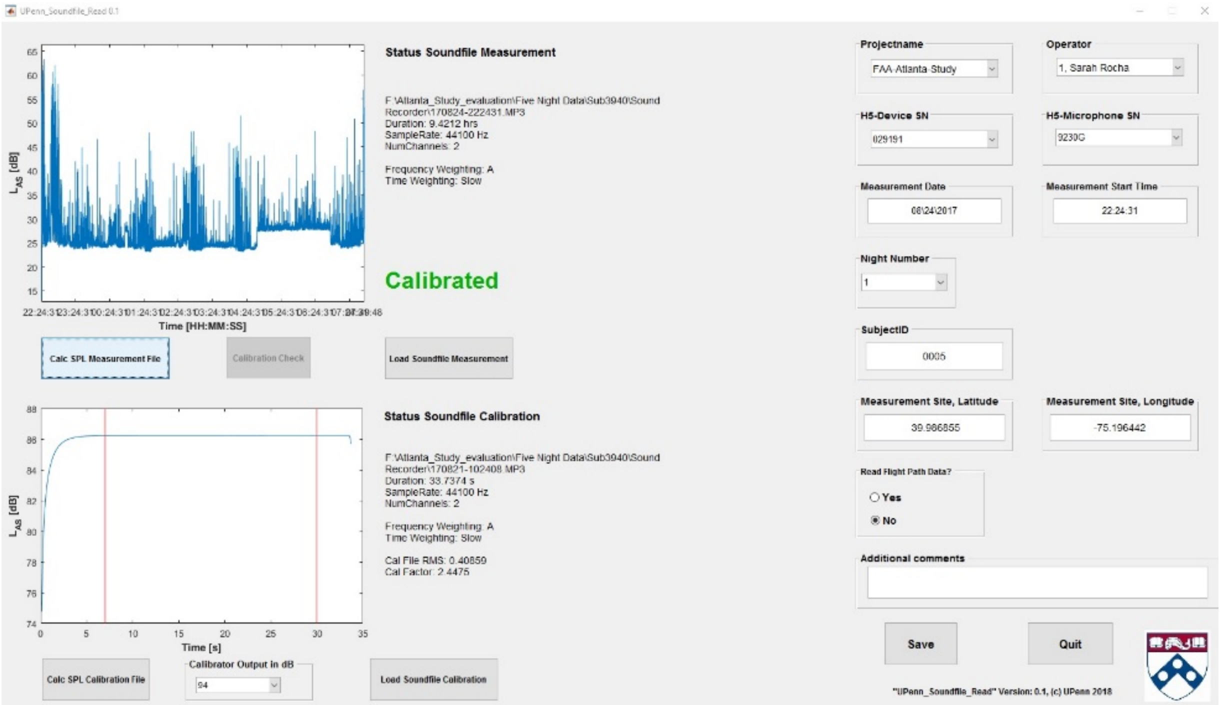


Figure 14 L_{AS} of a measurement file plotted using the Sound Pressure Level Converter program

2. Akustikview

Research assistants listened to and scored acoustic data using the acoustic scoring software, Akustikview, which is a non-commercial software developed in-house by our collaborators at DLR. Staff members marked when they heard subjects get into and out of bed, aircraft noise, background noise, traffic sounds, and any other relevant noise events in the bedroom. These notations were used to determine periods of time where the subject was not sleeping or was affected by non-aircraft noise events. In case of other noise events during an aircraft noise event, the maximum SPL of the aircraft noise event had to be the highest noise level for the aircraft noise event to be scored as the primary noise event. Akustikview recorded the L_{ASmax} for aircraft noise events as well as a number of other acoustical whole night and event related acoustical indicators (e.g., the average sound level in the minute preceding the start of the noise events). Once staff had scored a full night of acoustic data, Akustikview generated a text file with information on nightly aircraft noise events, for later use in the statistical analysis. The background noise level is automatically selected and scored by Akustikview, but the selection can be manually overwritten. Akustikview also synchronized the timeline of the acoustic data with the physiological data timeline using input from a time adaptation software (see section IV.G.3). Scoring all acoustic events in a given night is cumbersome, can take 2 hours or more, and is likely not feasible for a larger National Sleep Study. For this study, we plan to integrate flight rack radar data into the Akustikview software, that, based on the minimal distance to the receiver site on the ground, suggests times of expected aircraft noise events that can then be listened to and scored, probably including the minute before the start of the aircraft noise event.

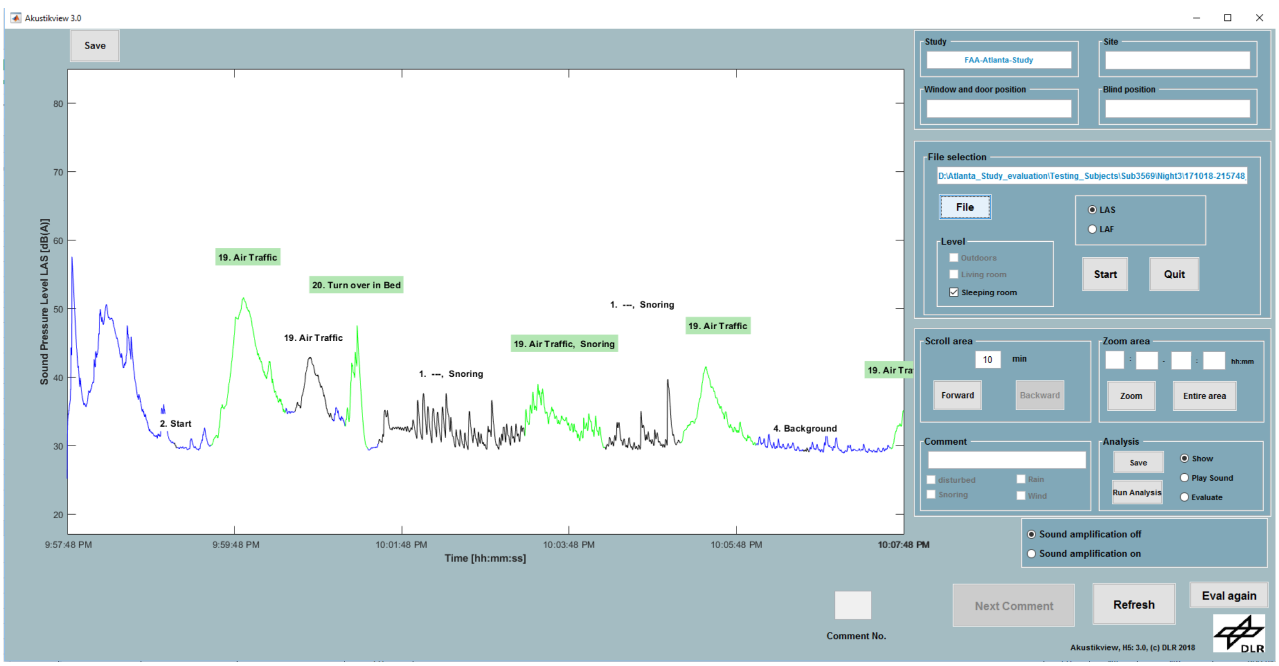


Figure 15 A ten-minute window of an acoustic file scored for aircraft noise and other sounds heard in the bedroom. Noise events are scored by staff and are displayed alternately in green and black. A caption appears above a noise event describing the type of noise heard by staff (e.g. air traffic). Unscored periods of the acoustic file appear in blue.

3. Time adapt

Physiological signals recorded with the Faros 90 heart rate devices and acoustical signals recorded with the H5 sound recorders were recorded on the individual devices. Although the devices were synchronized before they were shipped, it often took more than a week before data collection began. Therefore, over the course of the study, their internal clocks could potentially drift apart in time. A software called “Time Adapt” was developed by our collaborator at DLR to synchronize the timeline of the acoustic and physiological data (Figure 16). In this time adaptation software, body movements scored in the acoustic data were paired against movements detected in the physiological data. Time Adapt recorded the difference between movement events in the acoustic and physiological data. The differences were then plotted across the measurement night, as the time drift may increase throughout the night. Time Adapt fitted a linear regression to the time drift data and outputted this information in a text file to be read by the acoustic scoring software, Akustikview (section IV.G.2).

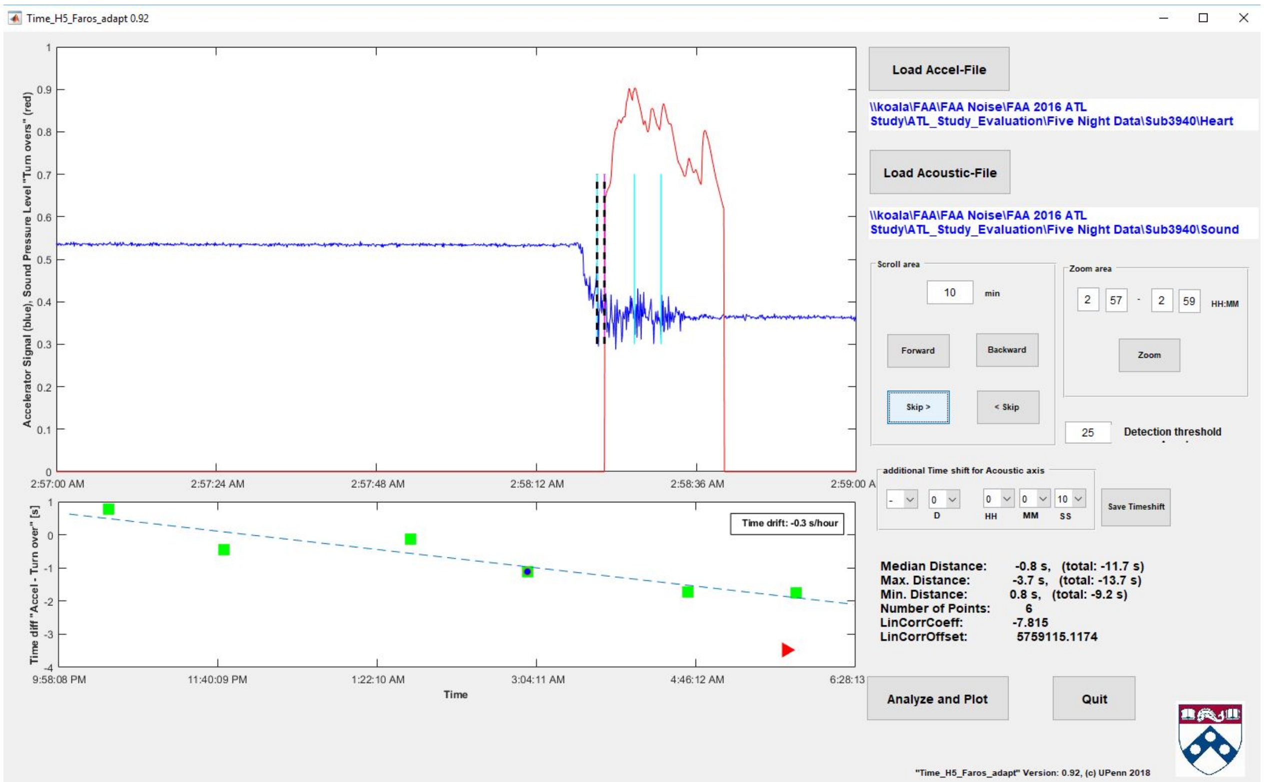


Figure 16 The Time Adapt program matches the start of major body movements with movements scored in the acoustic data. In the upper window, the accelerometer signal is plotted in blue for a given time window and marks the start of major body movements with pale blue lines. A body movement scored in the acoustic data is depicted in red. When the program pairs a body movement scored in both the physiologic and acoustic data, it adds a dashed line to the start of each event. In cases where there are multiple body movements in succession, staff can manually adjust which movement in the physiologic data is paired with the movement in the acoustic data by adding or subtracting time in the program. In the lower window, within-night time drift is shown (see section IV.B.2).

4. Heart rate file splitter

If subjects forget to turn off the heart rate device when they woke up and took off the device, it was possible that movement and heart rate data from multiple days were stored in one large file. Before physiological data could be read by the arousal detection software, the large file had to be split in two or more separate files. A software was developed for this purpose, which detects body movement recorded in the Faros 90 above a minimum threshold. The program then marks these periods of movement, and a human scorer manually adjusted the boundaries to encompass the actual time spent in bed (see Figure 17). Once adjustments had been made, the program then generated separate new data files.

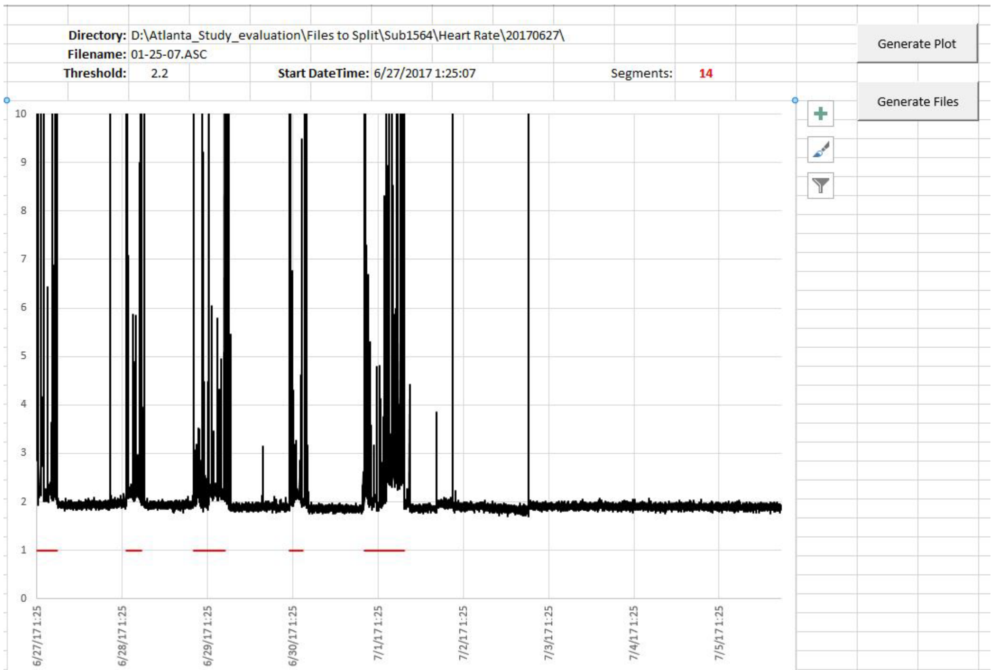


Figure 17 The Heart Rate Splitter program detects periods of body movement, indicated by red horizontal dashes, which can be exported into separate files.

5. Automatic identification of awakenings based on heart rate and actigraphy data

Awakenings during the night were identified automatically based on the heart rate and actigraphy data. The software (Figure 18) was based on the algorithm of Basner et al. [21] which identified EEG arousals (≥3 seconds) based on heart rate alone. This algorithm was refined to identify EEG awakenings (>15 seconds) using heart rate and actigraphy data, which is a more specific indicator of noise-induced sleep disturbance due to the lower frequency of occurrence on nights without noise exposure [49]. Awakenings are identified in the algorithm by using matrices of likelihood ratios which indicate whether the difference in the beat to beat heart rate to a 3 minute median heart rate or the amount of movement is associated with an awakening [50]. Awakenings were calculated for every subject night. After the calculations were completed, artefacts in the heart rate signals or missing data were visually identified, and these periods were removed from data analysis.

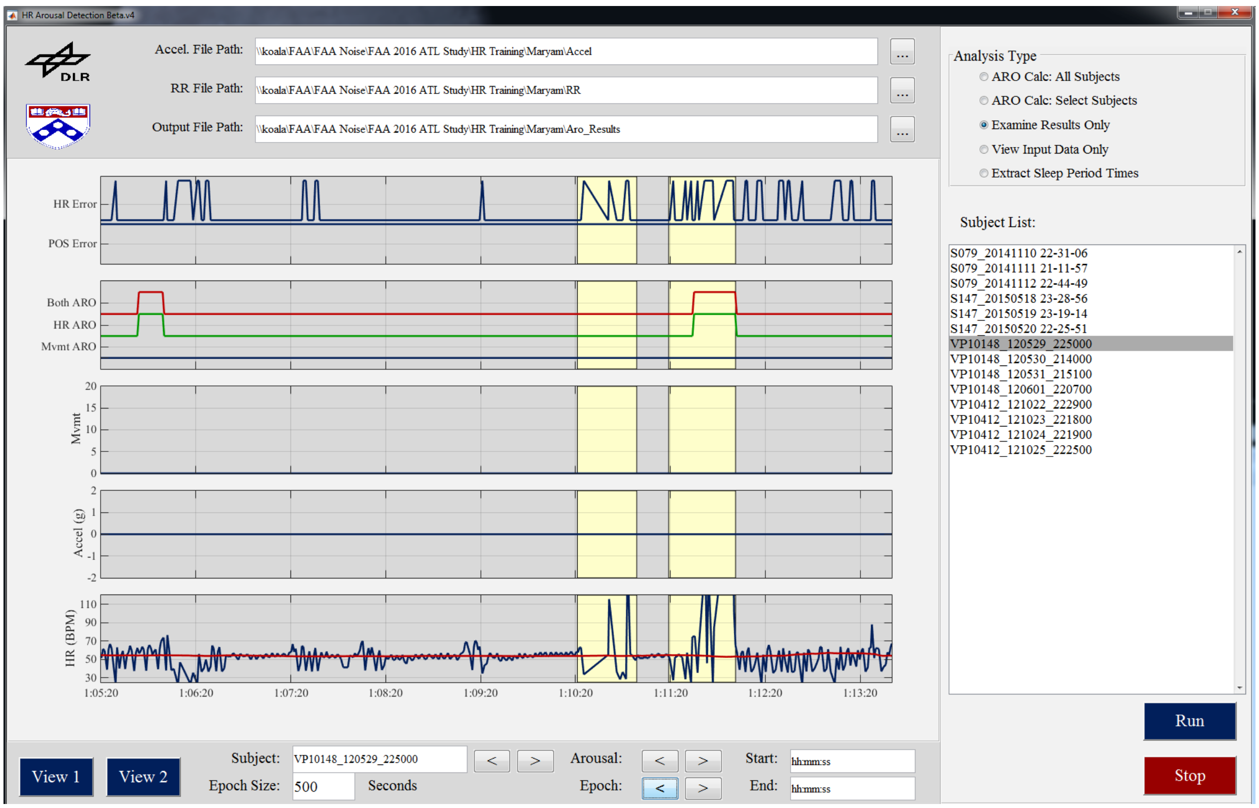


Figure 18 Physiological arousals were detected using the software's algorithm. Artefacts in the data were highlighted by staff (yellow sections) and removed from the dataset.

6. Respiratory signal viewer

We tried to recruit only subjects without intrinsic sleep disorders (like sleep apnea, restless leg movements syndrome, or periodic limb movements in sleep) into the study. However, subjects are often not aware of these sleep disorders, and therefore some intrinsic sleep disorders may not be captured by the questions of the recruitment survey addressing these disorders. Obstructive sleep apnea is characterized by partial or complete obstructions of the upper airways during sleep, that lead to decreases in blood oxygenation levels that ultimately cause an arousal which re-opens the airway. In a sleep laboratory, several physiological signals are used to identify obstructive respiratory events (measurements of movements of movements of the rib cage and abdomen, airflow measurements at the mouth and nose, and blood oxygenation measurements with pulse oximetry). Most of these signals were not available in our study, but the FAROS device, which was attached to the rib cage, is very sensitive, and we thus developed a software that displayed movements of the rib cage along all three orthogonal axes (Figure 19). We inspected rib cage movement for all subject nights for signs of possible obstructive or central sleep apnea, which would be indicated by repeated periods of no activity during times of restricted respiration, followed by an abrupt increase in activity as respiration was resumed. In this case, participants would be notified with a recommendation to seek out their primary care physician for further diagnostic procedures, and the collected data would be excluded from data analysis. In this study, none of the participants demonstrated potential signs of sleep apnea, and thus no data were subsequently excluded from our analyses.

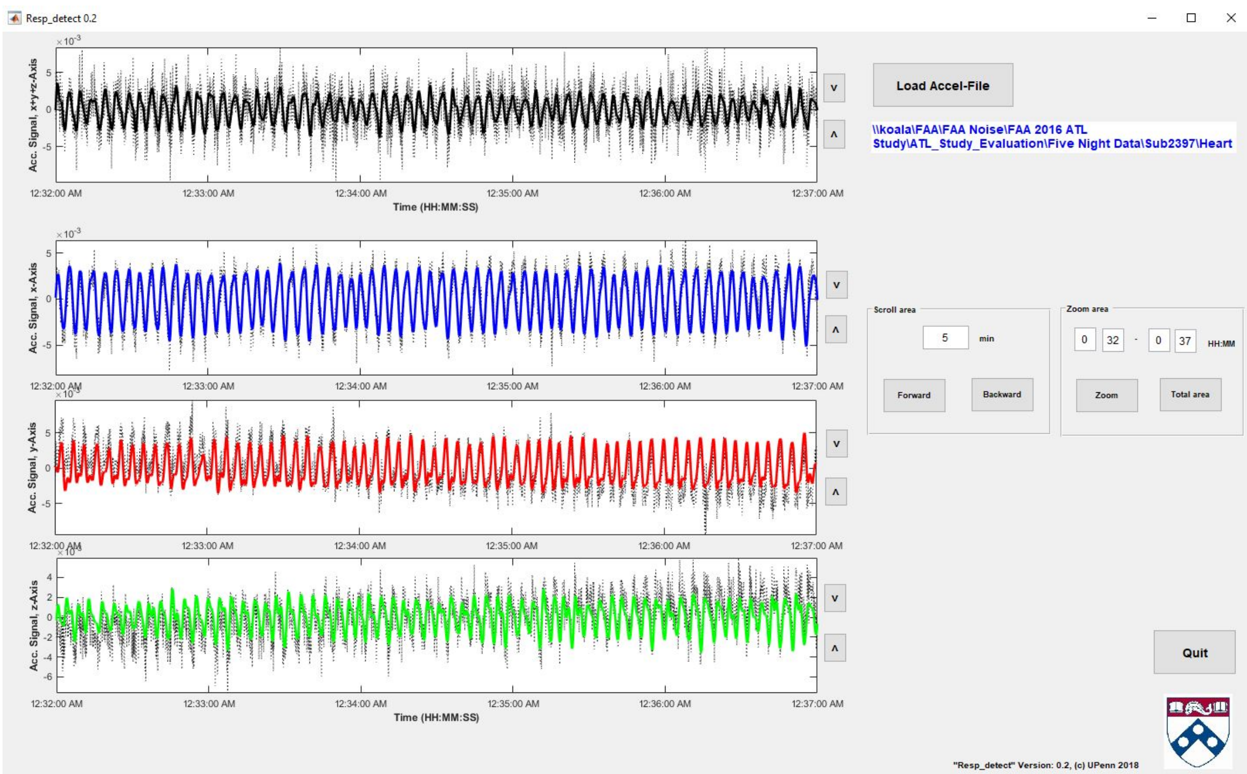


Figure 19 Respiratory signal of a healthy subject shown along the three axes and in a combined axis view.

7. Physiological analysis

The main outcome of interest of the event-related analysis is an exposure-response function between the maximum sound pressure level $L_{AS,max}$ of an aircraft noise event and the probability of the exposed subject to wake up.

Acoustic analysis – Aircraft event scoring

As described in detail earlier, sounds were continuously recorded in the bedroom of study participants with calibrated sound recorders. Sound levels were calculated based on these recordings. Trained research personnel listened to the sound recordings of each night and marked the beginning and the end of each aircraft noise event using Akustikview (see section IV.G.2). An aircraft noise event was only scored as such if it was the dominant noise source. For example, if a car drove by the house at the same time and generated a higher $L_{AS,max}$ than the aircraft, the event was classified as road traffic noise (primary) and aircraft (secondary). Only aircraft noise events characterized as the dominant (primary) noise source contributed to data analysis. In addition to the maximum SPL of aircraft noise events, the average noise level L_{AEq} in the minute prior to the start of the aircraft noise event was calculated as a proxy for the background noise level prior to the start of the aircraft noise event.

Automatic identification of awakenings based on heard rate and actigraphy data

Awakenings during the night were identified automatically based on the heart rate and actigraphy data, using the procedure and software described in IV.G.5.

Time drift correction

Time measured both by the sound recorders and by the Faros devices drifted in an approximately linear fashion relative to actual time determined by Network Time Protocol Internet servers. We wrote special software (see section IV.G.3 above) to correct for the time drift between acoustical and physiological data. We also added 5 seconds prior to the start of an aircraft noise event to the screening window to allow for minor inaccuracies in the time drift correction (see below).

Single event awakening analysis

All aircraft events were included in the single event analysis regardless of whether another noise source occurred at the same time, such as an aircraft event occurring at the same time as a car pass-by, as long as the aircraft noise was the dominant noise source. In analyses performed for WHO based on data from DLR's STRAIN study, it was found that for aircraft noise, exposure-response relationships did not vary relevantly when including all events or only events that did not co-occur with noise events from other sources [50]. A 50-second time window extending from -5 seconds until +45 seconds relative to the start of each aircraft noise event was screened for an awakening. A noise event was excluded from analysis if an awakening started before the start of this screening window and extended into or even beyond it. Five seconds before the start of the aircraft noise event were added to the screening window to account for any inaccuracies in synchronizing acoustical and physiological measurement equipment (see 2.5.3). The 50-second duration of the screening window was derived empirically from data collected at four different airports (PHL, ATL, FRA, and CGN), which maximized slope estimates for the maximum sound pressure level.

8. Statistical analyses

Survey protocol

We performed statistical analysis in IBM SPSS Statistics (version 25). We excluded surveys that were non-deliverable from all analyses with the exception of analysis of survey delivery rates. Binomial logistical regression models were constructed with completed survey (yes/no), interest in taking part in the field study (yes/no), or participation in the field study (yes/no) as the dependent variables. A number of regression models were constructed, including a combination of survey incentive (gift card/$2 cash), survey length (short/medium/long), number of follow-up waves (0/2/3), noise exposure category (<40/40-45/45-50/50-55/>55 dB) and orientation to the runway (West/East) as nominal predictor variables. Furthermore, sex (woman/man) and age category (18-29/30-39/40-49/50-59/60-69/70+) data from completed surveys were used as predictor variables in a regression model for both interest and participation in the field study. For each model, we performed an overall omnibus test (χ^2 tests) relative to the intercept-only model, and χ^2 tests within each model to examine whether there were significant fixed effects for any of the independent variables. Respondents with missing data were excluded from analyses involving the missing variables. Age data were missing for 43 respondents (10.6%), sex data were missing for 21 respondents (5.2%), and interest in the field study data were missing for 5 respondents (1.2%). The level of statistical significance was set at α=0.05. Results are reported as odds ratios (OR) and 95% confidence intervals (CI).

We calculated the cost effectiveness of the different survey strategies based on the cost of envelopes (both for mailing the surveys to the study population and the enclosed pre-paid envelopes for returning the completed surveys), paper, color printing, survey incentive and postage. Color printing cost $0.075 per page, with 3 pages for the short survey and 4 pages for the medium and long surveys. Mailing envelopes cost $0.086 each, which also required printing in color. Pre-printed return envelopes cost $0.093 each. We used the current cost of first class postage ($0.50) rather than the cost when we mailed the surveys.

Postal questionnaire results

Statistical analysis of the postal questionnaire data are described in detail in Rocha et al. 2019 [51] and are only summarized here. Only the long questionnaire versions were included in the analysis, corresponding to 3600 surveys across 15 mailing rounds. A logistic regression was performed using SAS (version 9.4, SAS Institute, Cary, NC). Each survey response variable was re-coded on a binomial scale. Responses in the top two categories (i.e. "very" & "extremely") were coded as "1" and all responses below as "0". $L_{Night,}$ was analyzed as a continuous variable using the outdoor L_{Night} estimate for each household.

We first analyzed each outcome separately in a crude, unadjusted model, with L_{night} only as an independent variable. We then analyzed each outcome in an adjusted multilevel regression model. We used directed acyclic graphs in DAGitty v2.3 to determine the minimal adjustment required to estimate the total effect of L_{night} on outcomes of interest [52]. Adjustment for age and income were minimally necessary, so we did not include occupational status or education in analysis models. In addition to L_{night} and income, we furthermore included sex, BMI, noise sensitivity and hearing problems as independent covariates in the adjusted model since we were interested in their influence on our outcomes. Fifteen missing values for age and 14 missing values for BMI were replaced with the mean age (53 years) and mean BMI (29 kg/m^2). Where categorical covariate data (sex, income, hearing problems and/or noise sensitivity) were missing, we excluded the respondent from analysis.

Wald Chi-Squared tests were performed to determine the significance of the predictor variables, and statistical significance was set to α=0.05. We did not correct for multiple testing in this exploratory analysis of pilot study data. Odds ratios with 95% confidence intervals and p-values are reported for both the crude and fully-adjusted models.

Field study morning questionnaires

Data were analyzed in repeated measures multiple logistic regression assuming an independent working correlation matrix (SPSS Generalized Estimating Equation). For each outcome variable, four models were performed. Two crude models used either the equivalent indoor aircraft noise over the individualized sleep period from physiologically-determined sleep onset to sleep cessation ($L_{AEq,sleep}$) or the maximum aircraft noise level during the sleep period ($L_{AS,max}$) as the primary independent predictor variable. Two adjusted models used the same noise exposures as the primary independent variables of interest but were further adjusted to account for the number of measured aircraft noise events during sleep (covariate), sex (dichotomous), age (covariate) and if the window was open or closed. There was only one single study night where the participant slept with fully open windows, therefore window closing was coded as a dichotomous variable as "fully closed" or "partially or completely open".

Numerical outcome variables (sleep latency, number of awakenings, tiredness, difficulty sleeping, sleep restlessness and sleep quality) were analyzed as continuous outcomes. Categorical outcome variables (Stanford Sleepiness Scale [SSS, question 7] [53] and sleep disturbance by aircraft, road, rail and general noise) were analyzed as dichotomous outcomes, where a score of ≥4 on the 7-point Likert scale for SSS was classified as "sleepy", and scores of ≥4 on the 5-point Likert scales for sleep disturbance were classified as "disturbed".

Event-related physiological data

Statistical analysis was performed using SAS (version 9.4, SAS Institute, Carey, NC). For the calculation of single event exposure-response relationships for the probability of an awakening, logistic mixed models with random subject intercept were calculated using Proc NLMIXED. The random intercept term accounts for the correlation of the repeated observations within each subject. In this case, the repeated observations are multiple reactions to aircraft noise events observed per subject. A p-value of 0.05 or less was considered statistically significant. We ran an unadjusted model with $L_{AS,max}$ as the only predictor, as well as models adjusting for age (continuous), BMI (continuous), time from sleep onset (continuous), and sex (nominal; value of 1=male, 0=female).

V. Results and discussion

A. Survey protocol

1. Delivery rates

Across all 17 mailing rounds, 3576 out of 4080 surveys (87.6%) were deliverable. A breakdown of the delivery rate, by survey round, is given in Table 5. When the survey was addressed only to "Current Resident", the mean deliverable rate was 92.1% (95% CI: 89.3-94.2%). When the survey address was personalized, the mean deliverable rate was 87.1% (95% CI: 85.9-88.1%). Regression analysis showed that there were lower odds (OR=0.578, 95% CI: 0.409-0.817) of delivery to personalized individuals than "Current Resident" only ($\chi^2(1, n=4080)=9.668$, $p=0.002$).

The delivery rate was lower for surveys sent to named individuals, perhaps due to the mail carrier not delivering if the name on the envelope did not match a name at the address despite the appended "or Current Resident", but this was more than offset by higher response rates among those named addressees. This increased response rate when personalizing the surveys is generally in agreement with previous research. A meta-analysis of 14 trials including over 12,000 participants found that the inclusion of names on health survey letters increased the odds of response by one fifth [54]. A later study however found that addressing surveys to named individuals significantly increased the response rate to reminder letters, but the increased response rate to the initial survey waves was not significant, although in this study of 1000 participants the absence of significance could be due to insufficient power [55]. As well as personalization, the higher response rate could be in part due to the removal of "School of Medicine" and "Department of Psychiatry" from the envelope, since psychiatry as a medical profession continues to suffer from public stigma [56]. We would not anticipate the change in envelope size to influence response [57].

2. Response rate

Out of 3576 delivered surveys, 407 were completed, a response rate of 11.4%. The majority (n=309; 75.9%) were returned by mail, with a minority (n=98; 24.1%) completed online. There was a statistically significant effect of respondent age category on the response mode (χ^2=54.9, p<0.0001), with younger respondents generally preferring to respond online and older respondents generally preferring to respond by mail (Figure 20).

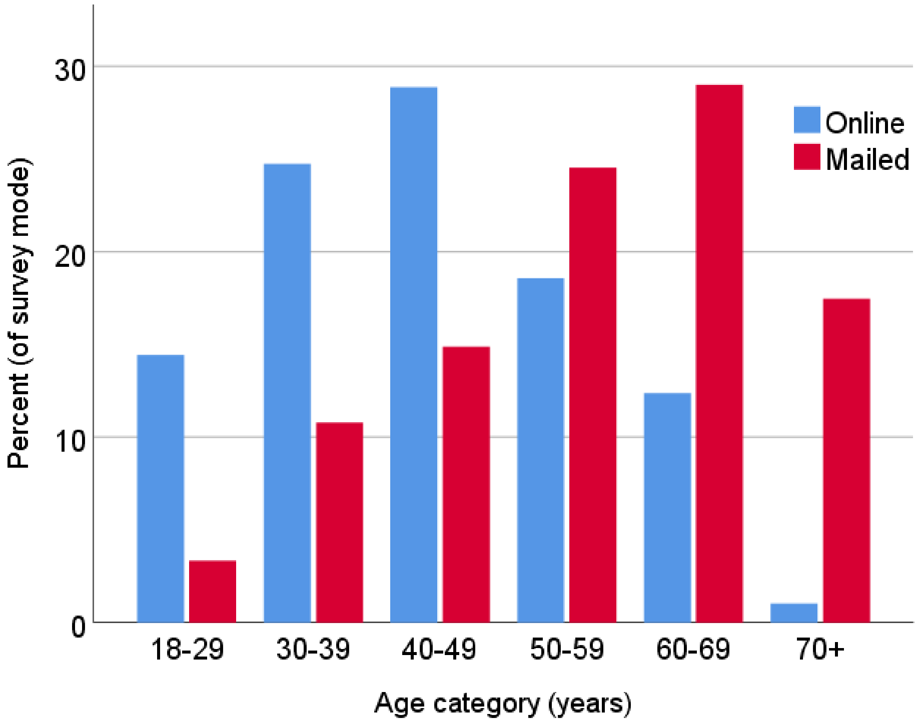


Figure 20 Effect of age of respondent on preferred response mode.

Among deliverable surveys within rounds 1-5, there was a 4.3% response rate when addressing the survey to a named individual in larger envelopes that indicated only “University of Pennsylvania” as the sender. The response rate was 1.4% when addressing the survey to only “Current resident” in smaller envelopes that indicated “Perelman School of Medicine” and “Department of Psychiatry, Division of Sleep and Chronobiology” as the sender. The higher response rate among personalized, larger envelope, “University of Pennsylvania” sender surveys was statistically significant (Wald χ^2(1, n=1094)=6.772, p=0.009, OR=3.261, 95% CI: 1.339-7.942).

A total response rate of 11.4% is lower than rates of 30-76% for postal surveys on aircraft noise annoyance in Europe and East Asia that were reported in a recent systematic review [58]. Our response rate is however in line with some more general attitudinal surveys [55, 59]. Possible reasons for non-response in our sample might include concerns about privacy and confidentiality despite assurances given in the introduction letter [60], illiteracy or language issues [61] or lack of interest in the survey topic or low community engagement [62]. In the United States, 37.6 million people speak Spanish at home [63], and including Spanish language surveys along with the English versions could improve response rates among this population without lowering response rates from non-Spanish speakers [64].

We received the majority of responses by mail, at a ratio of around 3:1 compared to online response. There is inconsistency among earlier studies regarding the influence of response mode, with some reporting higher response rates for paper surveys compared to online surveys e.g. [59, 65], and others finding an increased preference for completing questionnaires electronically e.g. [66]. We do not know whether those who completed our survey online would have returned it by post if the online option was not available, or vice versa for respondents who completed the survey by mail, and therefore cannot draw any conclusions regarding the optimal choice if only one survey mode were to be used in future studies. Offering web and mail response modes concurrently, rather than sequentially, may have reduced the overall response rate [67], although evidence is mixed [68]. Hypothesized reasons for this effect include, firstly, increased complexity in the decision to respond by introducing the choice of response mode; secondly, respondents choosing to respond online but never actually doing so since it involves a break in the response process; and thirdly sample members attempting to respond by web but not completing the survey due to computer or internet connectivity issues [69]. Initial mail contact offering a web-based response, and withholding paper surveys until later mailing rounds, may increase response rates compared to a paper-only method, but without significantly improving respondent representativeness [70]. A higher response rate, while not necessarily indicating greater respondent representativeness or data quality [71-73], may at least reduce the risk of nonresponse bias [67]. The pilot study presented in the current paper is a preceding step towards a national study of the potential effects of aircraft noise on sleep, and this future study offers the opportunity to more rigorously address nonresponse bias. One approach that has been widely used is comparing respondent characteristics to known characteristics of the whole population of interest [74, 75], in this case residents exposed to a certain minimum level of aircraft noise, using demographic data at the census tract level from the decennial U. S. Census [76] and the American Community Survey [77].

The survey rounds were not issued concurrently, but the earlier rounds were sent in autumn, the middle rounds were sent in winter or spring and the final rounds were sent in early summer. We cannot totally exclude there are subsequent effects on response rate, perhaps because residents were not home at certain times of year, or that there are seasonal effects influencing the predisposition of an individual to complete the questionnaire [78].

3. Effect of protocol on survey completion

We performed a regression analysis including the only round with pre-notification (round 5) and the two rounds that were otherwise identical except for pre-notification (rounds 3 and 4). There were higher odds for survey response when issuing a pre-notification postcard (OR=1.759, 95% CI: 0.821-3.765), but the effect was not statistically significant (Wald $\chi^2(1, n=652)=2.113$, $p=0.146$).

Results of the regression models for completing the surveys are presented in Table 7, and are graphically illustrated in Figure 21 in green. Regression model 1 (survey incentive, survey length, follow-up waves and field study incentive) indicated that a survey was more likely to be completed if including a $2 cash incentive compared to a gift card of any value (OR=2.792), and if 3 follow-up waves were issued compared to no follow-ups (OR=2.121). Survey length and field study incentive had no significant effect on survey completion rate. The inclusion of noise exposure category as a predictor (model 2) revealed results similar to that of model 1, with higher response rates for the $2 cash incentive (OR=2.798) and 3 follow-up waves (OR=2.120), but there was no effect of noise exposure or direction on survey completion rate.

Table 7 Results of the regression models for recipients completing the survey (including only deliverable surveys). All analyses excluded surveys that could not be delivered for any reason. OR=Odds Ratio. CI=Confidence Interval. Ref=Reference category. df=Degrees of Freedom. Statistically significant ($p<0.05$) results are indicated with bold typeface.

Model and test relative to intercept-only model	Variable	Fixed effects			Variable level	Completing survey		
		df	Wald χ^2	p		p-value	OR	95% CI
Model 1 $\chi^2(6, n=3576)=158.793$, $p<0.0001$	**Survey incentive**	**1**	**11.599**	**<0.001**	Gift card	Ref		
					$2	**<0.001**	**2.792**	**1.546-5.041**
	Survey length	2	2.569	0.277	Short	Ref		
					Medium	0.752	0.927	0.579-1.484
					Long	0.139	0.730	0.482-1.107
	Follow-up waves	**2**	**9.627**	**0.008**	0	Ref		
					2	0.114	1.530	0.903-2.591
					3	**0.005**	**2.121**	**1.250-3.597**
	Field study incentive	1	0.150	0.699	150	Ref		
					200	0.699	0.936	0.671-1.306
Model 2 $\chi^2(11, n=3576)=162.574$, $p<0.0001$	**Survey incentive**	**1**	**11.643**	**<0.001**	Gift card	Ref		
					$2	**<0.001**	**2.798**	**1.550-5.054**
	Survey length	2	2.505	0.286	Short	Ref		
					Medium	0.759	0.929	0.580-1.488
					Long	0.144	0.733	0.483-1.112
	Follow-up waves	**2**	**9.592**	**0.008**	0	Ref		
					2	0.114	1.530	0.903-2.592
					3	**0.005**	**2.120**	**1.249-3.596**
	Field study incentive	1	0.170	0.680	150	Ref		
					200	0.680	0.932	0.668-1.301
	Noise exposure category	4	3.397	0.494	<40	Ref		
					40-45	0.562	0.907	0.651-1.263
					45-50	0.306	0.839	0.599-1.175
					50-55	0.671	1.073	0.776-1.484
					>55	0.594	1.093	0.787-1.519
	Direction	1	1.073	0.300	West	Ref		
					East	0.538	0.936	0.758-1.156

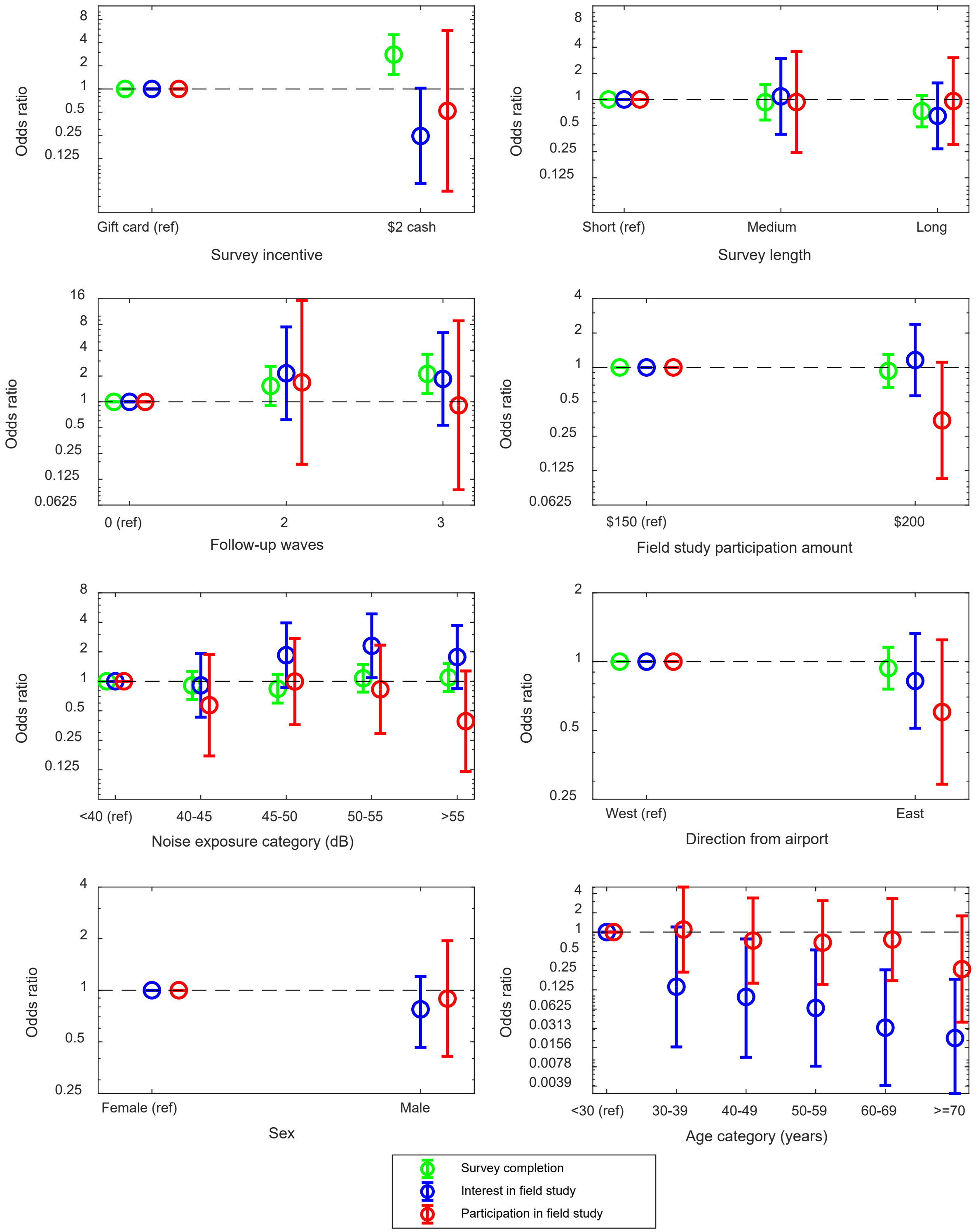


Figure 21 Odds ratios and 95% confidence intervals for the effect of different survey approaches and situational factors on receiving completed surveys (green), eliciting interest in the study (blue) and recruiting a participant into the study (red). The horizontal dashed line indicates the reference value OR=1.0.

Our findings on the effectiveness of different surveying strategies are in good agreement with the existing literature. For instance, a previous meta-analysis found that response to health research postal questionnaires could be improved by implementing repeat mailing strategies and, to a lesser degree, using shorter questionnaires [42]. In particular, the effectiveness of follow-ups on increasing response is rather well established in the existing literature [41, 79]. Similarly, we attained the highest response rate when using the most intensive follow-up strategy, but observed no significant increases in response when shortening the questionnaire length.

Only the mailing rounds with gift card incentives offered $100 for field study participation, and only the rounds with cash incentives offered $150 or $200 for field study participation, which is a limitation of the study design. The almost three times higher odds in survey response when we used a cash incentive is most plausibly due to the $2 cash outperforming the gift card as an incentive, rather than the difference in field study participation incentives. This is supported by the lack of observed differences in response rates between $150 and $200 field study incentives, the fact that monetary incentives have previously been found to outperform non-monetary incentives and that prepaid incentives outperform promised incentives [41, 80-83]. Furthermore, completion of the survey did not obligate field study participation, so we did not anticipate that field study compensation would influence survey response rates.

4. Effect of protocol on interest in field study

Out of 407 completed surveys, 237 respondents (58.2%) were interested in participating in the field study. Regression models for interest, calculated only using data from completed surveys, are given in Table 8, and are graphically illustrated in blue in Figure 21. The crude model (model 1) was not significantly different from the intercept-only model. In the fully adjusted regression model 3, residents exposed to 50-55 dB L_{Night} were more interested in taking part than those exposed to <40 dB (OR=2.304). There was a significant effect of age, with a monotonic decrease in the odds of interest in the field study with increasing age. There was also a statistically borderline effect (p=0.054) of survey incentive, whereby recipients of the $2 cash incentive were less likely to be interested in the field study (OR=0.245). No effects of survey incentive, survey length, number of follow-up waves or the field study participation incentive were found.

Older people are, for multiple reasons, frequently more difficult to recruit into experimental studies [84]. Accordingly, younger people in our survey sample were more interested in taking part in the field study. When endeavoring to recruit evenly distributed age groups in studies, oversampling from the target population might be needed.

Table 8 Results of the regression models for respondent interest in participating in the field study (including only completed surveys). All analyses excluded surveys that could not be delivered for any reason. df=Degrees of Freedom. OR=Odds Ratio. CI=Confidence Interval. Ref=Reference category. Statistically significant (p<0.05) results are indicated with bold typeface. Results of borderline statistical significance (p=0.05-0.1) are indicated with italic typeface.

Model and test relative to intercept-only model	**Variable**	**Fixed effects**			**Variable level**	**Interest in field study**		
		df	**Wald χ^2**	**p**		**p-value**	**OR**	**95% CI**
Model 1	Survey incentive	1	2.106	0.147	Gift card	Ref		
$\chi^2(6, n=402)=6.885$,					$2	0.147	0.417	0.128-1.359
p=0.332	Survey length	2	2.628	0.269	Short	Ref		
					Medium	0.819	1.111	0.452-2.733
					Long	0.233	0.621	0.284-1.358
	Follow-up waves	2	1.735	0.420	0	Ref		
					2	0.366	1.595	0.581-4.384
					3	0.811	1.130	0.414-3.090
	Field study incentive	1	0.001	0.971	150	Ref		
					200	0.971	1.011	0.550-1.861
Model 2	Survey incentive	1	2.095	0.148	Gift card	Ref		
$\chi^2(11, n=402)=20.832$,					$2	0.148	0.408	0.121-1.373
p=0.035	Survey length	2	2.854	0.240	Short	Ref		
					Medium	0.753	1.158	0.463-2.899
					Long	0.234	0.615	0.277-1.369
	Follow-up waves	2	1.564	0.457	0	Ref		
					2	0.422	1.529	0.543-4.310
					3	0.876	1.086	0.388-3.038
	Field study incentive	1	0.010	0.921	150	Ref		
					200	0.921	0.969	0.519-1.808
	Noise exposure category	**4**	**10.830**	**0.029**	<40	Ref		
					40-45	0.311	0.721	0.383-1.358
					45-50	0.150	1.619	0.841-3.118
					50-55	*0.072*	*1.775*	*0.949-3.318*
					>55	0.171	1.558	0.826-2.940
	Direction	1	2.049	0.152	West			
					East	0.152	0.738	0.487-1.119
Model 3	*Survey incentive*	*1*	*3.719*	*0.054*	Gift card	Ref		
$\chi^2(17, n=359)=63.308$,					*$2*	*0.054*	*0.245*	*0.059-1.023*
p<0.0001	Survey length	2	1.659	0.436	Short	Ref		
					Medium	0.873	1.086	0.396-2.973
					Long	0.330	0.647	0.270-1.553
	Follow-up waves	2	1.461	0.482	0	Ref		
					2	0.228	2.153	0.619-7.489
					3	0.332	1.851	0.534-6.421
	Field study incentive	1	0.164	0.685	150	Ref		
					200	0.685	1.160	0.565-2.381
	Noise exposure category	*4*	*8.904*	*0.064*	<40	Ref		
					40-45	0.803	0.909	0.430-1.924
					45-50	0.114	1.846	0.863-3.949
					50-55	**0.029**	**2.304**	**1.088-4.875**
					>55	0.132	1.768	0.842-3.713
	Direction	1	0.642	0.423	West	Ref		
					East	0. 423	0.823	0.511-1.326
	Sex	1	0.961	0.327	Female	Ref		
					Male	0. 327	0.774	0.464-1.202
	Age category	**5**	**33.150**	**<0.0001**	<30	Ref		
					30-39	*0.073*	*0.140*	*0.016-1.202*
					40-49	**0.029**	**0.094**	**0.011-0.781**
					50-59	**0.010**	**0.065**	**0.008-0.525**
					60-69	**0.001**	**0.032**	**0.004-0.257**
					≥70	**<0.001**	**0.022**	**0.003-0.183**

5. Effect of protocol on participation in field study

Among respondents interested in the field study, 79 respondents (19.4% of all completed surveys, 33.3% of those interested) met the eligibility criteria. Of those interested and eligible, 37 respondents (9.1% of completed surveys, 15.6% of those interested) were enrolled into the field study (see section V.G.2 for discussion of attrition at the different stages of recruitment). Regression models for participating in the field study, calculated only using data from completed surveys, are given in Table 9 and illustrated in red in Figure 21. In no models were any statistically significant effects of survey incentive, survey length, follow-up waves, field study incentive, age or sex found for the likelihood that respondents would participate in the field study.

The lack of significant difference in the odds of participation for different field study compensation amounts could suggest that the participants had more self-determined motivational traits [85], and/or that general interest in the research was a primary reason for taking part rather than financial interests alone. The hypothesis for personal interest is supported by the doubled odds of interest in the study for respondents exposed to 50-55 dB noise relative to the lowest noise category. Populations exposed to higher noise levels could be expected, through personal experience, to be more acutely aware of the issue of nocturnal aircraft noise, and therefore more willing to contribute to research on its effects. The odds in the highest exposure category (>55 dB) were not significantly higher than in the lowest category, which on one hand would not substantiate the idea for greater interest among those most affected, but could alternatively be explained by the most adversely affected people self-selecting themselves out of the area by moving to a quieter neighborhood.

Although rounds 1-5 offered $100 for field study participation, these mailing rounds also exclusively included gift cards as survey incentives, and so we cannot draw conclusions regarding differences in participation rates between $100 and $150/$200 amounts. Furthermore, the absence of significant findings could result from insufficient statistical power, since only 34 subjects eventually participated in the field study.

The highest probability of field study participation, achieved with the short survey - although not statistically significant - may reflect a modest advantage of using a reduced survey length. On the other hand, the short survey required additional telephone contact, which may be the cause of a potential higher participation likelihood, rather than the short survey *per se*.

The study design was not perfectly balanced, so we cannot conclude whether increasing the field study compensation from $100 to $150 or $200 would have affected recruitment. To avoid possible confounding, an alternative study design, but with additional expense, could involve a 2×2×3×3 factorial design with the factors of pre-/post-completion incentive, $2/gift card incentive, short/medium/long survey length and 0/2/3 follow-up waves.

Table 9 Results of the regression models for recipients participating in the field study (including only completed surveys). All analyses excluded surveys that could not be delivered for any reason. df=degrees of freedom. OR=Odds Ratio. CI=Confidence Interval. Ref=Reference category.

Model and test relative to intercept-only model	Variable	Fixed effects			Variable level	Field study participation		
		df	Wald χ^2	p		p-value	OR	95% CI
Model 1	Survey incentive	1	0.174	0.677	Gift card	Ref		
χ^2(6, n=407)=4.707,					$2	0.9677	0.608	0.059-6.305
p=0.582	Survey length	2	0.058	0.809	Short	Ref		
					Medium	0.809	0.855	0.241-3.040
					Long	0.896	0.929	0.307-2.811
	Follow-up waves	2	0.805	0.669	0	Ref		
					2	0.698	1.528	0.179-13.022
					3	0.936	0.914	0.100-8.300
	Field study incentive	1	2.828	0.093	150	Ref		
					200	0.093	2.657	0.851-6.588
Model 2	Survey incentive	1	0.294	0.588	Gift card	Ref		
χ^2(9, n=407)=10.502,					$2	0. 588	0.521	0.049-5.505
p=0.486	Survey length	2	0.065	0.968	Short	Ref		
					Medium	0.810	0.854	0.236-3.095
					Long	0.843	0.892	0.290-2.748
	Follow-up waves	2	1.012	0.603	0	Ref		
					2	0.628	1.703	0.197-14.691
					3	0.971	0.960	0.104-8.834
	Field study incentive	1	3.254	0.071	150	Ref		
					200	0. 071	2.890	0.912-9.153
	Noise exposure category	4	3.662	0.454	<40	Ref		
					40-45	0.258	0.519	0.166-1.619
					45-50	0.906	1.061	0.399-2.818
					50-55	0.605	0.770	0.285-2.079
					>55	0.142	0.427	0.137-1.330
	Direction	1	1.917	0.166	West	Ref		
					East	0.166	0.607	0.299-1.231
Model 3	Survey incentive	1	0.286	0.593	Gift card	Ref		
χ^2 (17,					$2	0. 593	0.520	0.047-5.730
n=364)=13.496,	Survey length	2	0.011	0.995	Short	Ref		
p=0.702					Medium	0.919	0.933	0.244-3.569
					Long	0.944	0.959	0.303-3.036
	Follow-up waves	2	1.092	0.579	0	Ref		
					2	0.642	1.687	0.187-15.238
					3	0.935	0.910	0.094-8.817
	Field study incentive	1	3.190	0.074	150	Ref		
					200	0.074	2.904	0.901-9.354
	Noise exposure category	4	3.432	0.488	<40	Ref		
					40-45	0.354	0.570	0.173-1.873
					45-50	0.992	0.995	0.360-2.746
					50-55	0.722	0.828	0.293-2.340
					>55	0.119	0.391	0.120-1.274
	Direction	1	1.877	0.171	West	Ref		
					East	0.171	0.602	0.291-1.245
	Sex	1	0.081	0.776	Female	Ref		
					Male	0.776	0.894	0.411-1.942
	Age category	5	3.223	0.666	<30	Ref		
					30-39	0.906	1.096	0.237-5.064
					40-49	0.696	0.737	0.159-3.410
					50-59	0.624	0.686	0.152-3.093
					60-69	0.722	0.764	0.173-3.368
					≥70	0.173	0.263	0.039-1.793

6. Questionnaire completion and field study participation probabilities

Probabilities of completing the survey and participating in the field study were calculated using regression model 1. The probability of surveys being completed for each observed combination of survey incentive, survey length and follow-up waves are given in Table 10. The more follow-up waves were sent and the shorter the survey length, the more likely it was to receive a completed survey, with a response rate of 21.7% for survey rounds with 3 follow-up waves, a short survey and a $2 cash incentive.

Table 10 Predicted probability and 95% confidence intervals (CI) of receiving a completed survey, stratified by number of follow-up waves, survey length and survey incentive. Data calculated excluding non-deliverable surveys.

Sample size (n)	Probability of completing survey and 95% CIs (%)	Follow-up waves	Survey length	Survey incentive
207	21.7 (16.6-27.9)	3	Short	$2
210	20.5 (15.6-26.5)	3	Medium	$2
1041	16.3 (14.2-18.7)	3	Long	$2
805	12.0 (10.0-14.5)	2	Long	$2
219	8.2 (5.2-12.7)	0	Long	$2
1094	3.1 (2.2-4.3)	0	Long	Gift card
Total=3576				

Since the $2 cash incentive was superior to gift cards for receiving completed surveys, and therefore likely a more representative sample, we restricted analysis of field study participation to rounds where only the cash incentive was used (rounds 6-17). The probability of respondents participating in the field study for each combination of survey length, follow-up waves and field study incentive, are given in Table 11. We calculated probabilities based on both the total number of surveys mailed and from among completed surveys only. Since the field study incentive of $100 was offered only in rounds 1-5, probabilities are presented for incentive amounts of $150 and $200 only. The shorter the survey length, the more likely it was for a respondent to participate in the field study. Generally, participation was more likely with more follow-up waves and with the lower field study incentive, although there may be some confounding among these variables due to the unbalanced design.

Table 11 Predicted probability and 95% confidence intervals (CI) of a recipient participating in the field study, stratified by number of follow-up waves, survey length, and field study participation amount. Data calculated excluding non-deliverable surveys and gift card incentive rounds.

Sample size (n)*	Probability of participating in field study (% with 95% CIs)*	Probability of participating among survey respondents (% with 95% CIs)†	Follow-up waves	Survey length	Field study participation amount
207	2.9 (1.3-6.3)	13.3 (6.1-26.7)	3	Short	$150
210	2.4 (1.0-5.6)	11.6 (4.9-25.1)	3	Medium	$150
427	2.1 (1.1-4.0)	12.5 (6.6-22.3)	3	Long	$150
805	1.0 (0.5-2.0)	8.2 (4.2-15.6)	2	Long	$200
614	0.8 (0.3-1.9)	5.1 (2.1-11.7)	3	Long	$200
219	0.5 (0.1-3.2)	5.6 (0.8-30.7)	0	Long	$200
Total=2482					

*Based on total number of surveys mailed (n=2482)
†Based only on completed surveys (n=407)

B. Postal questionnaires

1. Delivery rates

Out of 3600 surveys mailed, 3159 surveys could be delivered. Of deliverable surveys, 319 were completed and returned, resulting in a response rate of 10.1%. Twenty one surveys with missing information for sex (n=21), income (n=9), noise sensitivity (n=7) and/or hearing problems (n=33) were excluded from analysis, resulting in an effective sample size of n=268 (8.5 %) of the surveyed population.

2. Respondent demographics

summarizes the demographics for respondents to the noise and sleep survey for whom there were no missing sex, income, noise sensitivity or hearing problem data. There were 57 participants in the 35<40 dB noise exposure category, 46 in the 40<45 dB category, 51 in the 45<50 dB category, 64 in the 50<55 dB category and 50 in the ≥55 dB category. Respondents ranged in age from 21 to 97 years, with a mean of 52.4 years (SD±15.2) and had a mean BMI of 29.3 kg/m2 (SD±6.5). Sixty one percent of respondents were black, which is a similar proportion to the 62.5% mean proportion for the sampled sampling region. For highest level of completed education, 46.4% of respondents had no college education, which is slightly higher than the 40.0% of the population without college education in the sampling region. Among respondents who disclosed their income, 50.9% had a household income below $50,000, with 31.3% of respondents and the median value lying in the $25-50k category. This is in agreement with the mean household income for the sampled census tracts of $49,100.

Table 12 summarizes the demographics for respondents to the noise and sleep survey for whom there were no missing sex, income, noise sensitivity or hearing problem data. There were 57 participants in the 35<40 dB noise exposure category, 46 in the 40<45 dB category, 51 in the 45<50 dB category, 64 in the 50<55 dB category and 50 in the ≥55 dB category. Respondents ranged in age from 21 to 97 years, with a mean of 52.4 years (SD±15.2) and had a mean BMI of 29.3 kg/m2 (SD±6.5). Sixty one percent of respondents were black, which is a similar proportion to the 62.5% mean proportion for the sampled sampling region. For highest level of completed education, 46.4% of respondents had no college education, which is slightly higher than the 40.0% of the population without college education in the sampling region. Among respondents who disclosed their income, 50.9% had a household income below $50,000, with 31.3% of respondents and the median value lying in the $25-50k category. This is in agreement with the mean household income for the sampled census tracts of $49,100.

Table 12 Demographic characteristics of survey respondents (N=268) for whom complete data were available for regression analysis. Respondents could provide multiple answers for Race.

Variable	Level	Percent
Sex (n=268)	Women	64.9
	Men	35.1
Race (n=268)	Black	61.2
	White	24.6
	Other	8.2
	Prefer not to answer	10.4
Marital Status (n=267)	Single	36.9
	Married or domestic partners	38.6
	Widowed	7.9
	Separated/divorced	16.5
Income (n=268)	<$50,000	41.8
	$50,000-$100,000	27.2
	>$100,000	13.1
	Prefer not to answer	17.9
Education (n=265)	<High School	4.2
	High School	42.3
	College or greater	53.6
Employment (n=265)	Working	53.6
	Unemployed	9.1
	Student	1.9
	Retired	30.9
	Homemaker	4.5
Hearing (n=268)	No problems	85.8
	Problems	14.2
Noise sensitivity (n=268)	Not sensitive	69.0
	Sensitive	31.0

3. Survey responses

Sleep disturbance by noise, annoyance by noise and sleep quality

Results of the unadjusted logistic regression models for annoyance, sleep disturbance and sleep quality are presented in Table 13. With increasing nocturnal aircraft noise exposure L_{night} there were significant increases in the following outcomes: sleep disturbance by aircraft noise; annoyance by aircraft noise; likelihood of rating overall sleep quality as “bad” or “fairly bad”; trouble falling asleep within 30 minutes at least once a week; trouble sleeping at night due to nocturnal awakenings or waking too early in the morning at least once a week; and trouble staying awake during the daytime at least once a week. Only use of sleep medications was not significantly associated with L_{night}. Nighttime aircraft noise was therefore associated with higher sleep disturbance and decreased subjective sleep quality.

Table 13 Odds ratios and 95% confidence intervals from unadjusted logistic regression models for sleep quality variables.

Covariate	Outcome measure: Sleep disturbance	Annoyance	Overall sleep quality	Trouble falling asleep	Trouble sleeping at night	Sleep medication	Trouble staying awake
L_{night}	1.15 [1.10-1.20]****	1.17 [1.11-1.22]****	1.05 [1.01-1.08]*	1.05 [1.02-1.09]**	1.04 [1.01-1.08]*	0.99 [0.95-1.04]	1.06 [1.01-1.11]*

P-values for odds ratios that are statistically significant are denoted with asterisks (*<0.05; **<0.01; ****<0.0001).

The odds ratios for the associations between L_{night} and sleep in the adjusted regression models (Table 14) closely match results from the unadjusted models, although trouble staying awake during the daytime is no longer significant. Furthermore, there were significant effects of noise sensitivity for all of the sleep outcomes, with noise sensitive individuals reporting higher disturbance, annoyance and trouble sleeping than non-sensitive individuals. Respondents with hearing problems were more likely to report trouble falling asleep and staying asleep. There were also effects of income bracket, with respondents in the highest annual income bracket (>$100k) less annoyed and sleep disturbed by aircraft noise than respondents in the lowest income bracket (<$50k).

Table 14 Odds ratios from logistic regression models for sleep quality variables, adjusted for age, BMI, sex, hearing problems, noise sensitivity and income.

Covariate	Level	Outcome measure						
		Sleep disturbance	Annoyance	Overall sleep quality	Trouble falling asleep	Trouble sleeping at night	Sleep medication	Trouble staying awake
L_{night} [95% CI]	Continuous	1.15 [1.10-1.21]****	1.17 [1.11-1.23]****	1.04 [1.00-1.08]*	1.06 [1.02-1.10]**	1.04 [1.00-1.08]*	0.98 [0.94-1.03]	1.05 [1.00-1.11]
BMI	Continuous	0.95*	0.95*	1.00	0.99	1.00	1.04	1.00
Sex[a]	Male	1.05	1.13	0.99	1.23	0.59	0.66	0.58
Age	Continuous	1.00	1.00	0.99	0.99	0.99	0.99	0.99
Hearing problems[b]	Hearing problems	0.72	0.66	1.44	2.46*	2.51*	1.57	1.98
Noise sensitivity[c]	Noise sensitive	3.05***	3.10***	2.09**	2.74***	4.01****	2.10*	2.03*
Income[d]	$50-100k	0.49	0.69	1.08	0.94	0.89	1.13	1.53
	>$100k	0.21*	0.17*	0.72	0.63	0.72	1.70	0.83
	Prefer not to answer	0.67	0.85	0.82	0.61	0.56	1.88	0.62

Reference categories as follows: [a]Female; [b]No hearing problems; [c]Not noise sensitive; [d]<$50k; P-values for odds ratios that are statistically significant are denoted with asterisks (*<0.05; **<0.01; ***<0.001; ****<0.0001). CI: Confidence interval.

Use of sleep aids in response to noise

Results of the unadjusted logistic regression models for often or always using different sleep aids are presented in Table 15. With increasing nocturnal aircraft noise exposure L_{night}, respondents were significantly more likely to report using alcohol, television, music closing their windows in response to noise. Nighttime aircraft noise was therefore positively associated with increased prevalence of a number of coping behaviors.

Table 15 Odds ratios and 95% confidence intervals from unadjusted logistic regression models for always or often using sleep aids because of noise.

Covariate	Outcome							
	Earplugs	Alcohol	Medication	TV	Music	Close windows	Sound machine	Fan
L_{night}	1.04 [0.98-1.12]	1.11 [1.01-1.21]*	1.01 [0.97-1.06]	1.06 [1.02-1.10]**	1.08 [1.02-1.13]**	1.05 [1.01-1.08]**	0.97 [0.90-1.05]	1.02 [0.99-1.06]

P-values for odds ratios that are statistically significant are denoted with asterisks (*<0.05; **<0.01; ***<0.001).

The odds ratios and tests of significance for the associations between L_{night} and sleep aid use in the adjusted regression models (Table 16) closely match results from the unadjusted models. Furthermore, there were some significant effects of age, hearing problems and noise sensitivity. Older individuals were increasingly less likely to use music or fans as a sleep aid. Noise sensitive respondents and respondents with hearing problems were more than twice as likely to use either medication or television as a sleep aid. Noise sensitive individuals were also more likely to close their windows and use fans than non-sensitive individuals. Individuals with hearing problems were over 5 times as likely to use music as a sleep aid against noise compared to individuals without hearing problems.

Table 16 Odds ratios from logistic regression models for always or often using sleep aids because of noise, adjusted for age, BMI, sex, hearing problems, noise sensitivity and income.

Covariate	Level	Outcome							
		Earplugs	Alcohol	Medication	TV	Music	Close windows	Sound machine	Fan
L_{night} [95% CI]	Continuous	1.04 [0.96-1.12]	1.10 [1.00-1.21]*	1.01 [0.96-1.06]	1.05 [1.01-1.10]*	1.07 [1.01-1.13]*	1.05 [1.01-1.09]**	0.99 [0.91-1.07]	1.01 [0.97-1.06]
BMI	Continuous	1.08*	0.99	1.03	1.02	1.04	0.96	0.95	1.03
Sex[a]	Male	0.90	1.12	0.85	0.82	0.84	1.16	0.97	0.77
Age	Continuous	1.00	0.97	1.00	0.98	0.94****	0.99	1.00	0.97**
Hearing problems[b]	Hearing problems	3.00	0.55	2.91*	5.14***	5.18***	1.05	1.14	2.07
Noise sensitivity[c]	Noise sensitive	2.42	2.29	2.29*	2.37*	1.08	1.71	0.96	2.04*
Income[d]	$50-100k	2.46	1.23	0.89	1.42	1.35	0.98	1.59	0.80
	>$100k	1.77	1.12	1.74	0.91	0.45	0.62	2.25	1.00
	Prefer not to answer	0.91	1.80	1.22	2.51	0.99	0.50	1.13	0.91

Reference categories as follows: [a]Female; [b]No hearing problems; [c]Not noise sensitive; [d]<$50k; P-values for odds ratios that are statistically significant are denoted with asterisks (*<0.05; **<0.01; ***<0.001; ****<0.0001). CI: Confidence interval.

Health

Results of the unadjusted logistic regression models for self-reported general heath and diagnosis of relevant health outcomes are presented in Table 17. With increasing nocturnal aircraft noise exposure L_{night}, respondents were significantly more likely to rate their health as worse, i.e. as fair or poor rather than good to excellent. This association was not statistically significant after adjusting for BMI, sex, age, hearing problems, noise sensitivity and income (Table 18). With increasing BMI, respondents were more likely to rate their health as worse and report a prior diagnosis of a sleep disorder, hypertension, and diabetes. With increasing age, respondents were more likely to report a prior diagnosis of a sleep disorder, hypertension, arrhythmia, heart disease and diabetes. There were no significant effects of sex or noise sensitivity on any of the measured health outcomes.

Table 17 Odds ratios and 95% confidence intervals from unadjusted logistic regression models for general health and diagnosis of different health outcomes.

Covariate	Outcome measure							
	General health†	Sleep disorder	Hypertension	Chronic headaches/ Migraine	Arrythmia	Heart disease	Stomach ulcer	Diabetes
L_{night}	1.06 [1.02-1.11]**	1.00 [0.96-1.04]	1.00 [0.97-1.04]	1.04 [0.98-1.11]	0.98 [0.92-1.04]	1.06 [0.97-1.15]	0.95 [0.86-1.05]	0.98 [0.93-1.03]

† Odds ratio of reporting health as poor or fair. P-values for odds ratios that are statistically significant are denoted with asterisks (*<0.05; **<0.01).

Table 18 Odds ratios from logistic regression models for general health and diagnosis of different health outcomes, adjusted for age, BMI, sex, hearing problems, noise sensitivity and income.

Covariate	Level	Outcome measure							
		General health†	Sleep disorder	Hypertension	Chronic headaches/ Migraine	Arrythmia	Heart disease	Stomach ulcer	Diabetes
L_{night} [95% CI]	Continuous	1.04 [1.00-1.09]	0.99 [0.95-1.03]	1.00 [0.96-1.04]	1.03 [0.96-1.10]	0.99 [0.92-1.06]	1.08 [0.98-1.18]	0.95 [0.85-1.06]	0.96 [0.90-1.01]
BMI	Continuous	1.08***	1.07**	1.13****	0.98	1.01	1.02	0.95	1.10***
Sex[a]	Male	1.33	0.85	1.04	0.51	1.14	1.86	0.52	0.82
Age	Continuous	0.99	1.03*	1.10****	0.98	1.07**	1.06*	1.03	1.06***
Hearing problems[b]	Hearing problems	2.28*	2.03	1.25	1.24	2.12	2.27	0.67	0.85
Noise sensitivity[c]	Noise sensitive	1.28	1.61	0.87	1.36	1.65	1.02	0.35	1.31
Income[d]	$50-100k	0.78	1.13	1.15	0.84	1.27	1.42	0.78	0.64
	>$100k	0.22	2.03	1.98	0.36	0.94	1.03	1.47	1.30
	Prefer not to answer	1.30	1.60	1.10	0.60	0.73	0.57	-	3.09*

† Odds ratio of reporting health as poor or fair. Reference categories as follows: [a]Female; [b]No hearing problems; [c]Not noise sensitive; [d]<$50k; P-values for odds ratios that are statistically significant are denoted with asterisks (*<0.05; **<0.01; ***<0.001; ****<0.0001). CI: Confidence interval. Among respondents who chose not to report income, none reported stomach ulcers, so the odds ratio could not be determined.

4. Subjective sleep quality, disturbance and coping strategies

From the 3600 long form versions of the postal surveys sent out, we found that residents living in regions with higher levels of nighttime aircraft noise were more likely to report poor overall sleep quality. They also reported greater difficulty falling asleep within 30 minutes and trouble sleeping at night due to waking in the middle of the night or too early in the morning. These findings are consistent with those of a recent World Health Organization (WHO) review on environmental noise and self-reported sleep outcomes [11]. While the WHO report found a statistically significant relationship between aircraft noise and disruptions to sleep only when noise was referred to in the question, in our study L_{night} was associated with poorer self-reported sleep quality without a reference to noise. However, the title of our survey referenced noise, which may have influenced respondents when answering questions about their sleep. Furthermore, the choice of classification we used for coding the dichotomous variables of reporting difficulty sleeping, i.e. we coded a sleep difficulty as present if it occurred once a week or more rather than the single highest response of three time a week or more, should moderate conclusions regarding associations between aircraft noise and subjective sleep. According to the American Psychiatric Association, a criterion for diagnosis of insomnia is that sleep difficulty occurs at least 3 times per week [86]. Few respondents reported that trouble with sleep occurred at least 3 times a week, precluding statistical analysis of this response category only, and further illustrating the need for the larger sample sizes that will be obtained in the nationwide study. Respondents were increasingly likely to report that they were very or extremely annoyed and sleep disturbed with increasing L_{night}, which are responses corresponding to the "highly annoyed" and "highly sleep disturbed" classifications used by the WHO in their estimations of the disease burden of environmental noise [87]. These level-dependent associations between aircraft noise and annoyance and sleep disturbance are consistent with what has been found previously in the literature [11, 88, 89]. Good quality sleep is important for many biological functions and overall health, and so public perception that aircraft noise is disrupting sleep is a relevant concern. However, questionnaires on self-reported sleep may not fully capture the magnitude of the effect of environmental noise on sleep. Nighttime awakenings due to noise often occur without conscious awareness, and so residents may not accurately estimate the degree to which noise affects their sleep. There is also some evidence that self-reported sleep may only be weakly associated with objective sleep measures of sleep [90]. Thus, future field studies on the physiological responses to nighttime aircraft noise are needed to elucidate the objective impact of aircraft noise on sleep.

Along with disrupting sleep, aircraft noise can be annoying to residents living near airports. It was found that residents living in regions with higher L_{night} levels were significantly more likely to report feeling highly annoyed by aircraft noise over the last twelve months. This finding is consistent with previous annoyance studies (e.g., [91-93]). A limitation of this finding is that we only examined the associations between L_{night} and annoyance to aircraft noise, and so we cannot exclude daytime noise exposure as the main source of annoyance. A high level of annoyance to aircraft noise is concerning not just because of its impact on mood, but also because of its potential to influence sleep. While we sleep, the brain continues to processes and evaluate auditory stimuli, and so noise events that have emotional relevancy may induce a nighttime arousal with a higher probability compared to those that are less emotionally relevant [94]. Addressing annoyance to aircraft noise may thus be an important component in preventing aircraft noise-induced sleep disturbance. Because of the limited sample size and low response rate, annoyance levels found in this study should not be generalized to the studied, or other, airports. A much larger national survey across a more representative selection of 20 airports was recently conducted by the FAA and is expected to provide more precise exposure-response functions for daytime and nighttime aircraft noise annoyance [95].

Residents who were sensitive to noise were more likely to report annoyance and sleep disturbance by aircraft noise, as well as worse subjective sleep overall in all measures of sleep quality. This is in line with previous findings that noise sensitive individuals report worse sleep [96-98]. As a result of noise, sensitive respondents were also were more likely to report using three of the eight measured sleep aids at least once a week, further supporting the idea they were more psychologically susceptible to, or more cognizant of, nocturnal noise. It is however unclear how sensitivity might influence the impact of noise on sleep biology, with several studies finding minimal or no physiologic effects of sensitivity [97, 99, 100], or even that their sleep was objectively better [101]. Regardless of whether or not physiologic effects of noise are moderated by an individual's sensitivity, consistent findings, both here and previously, that they report worse subjective sleep and increased annoyance and disturbance remain relevant when considering the public health implications of nocturnal aircraft noise exposure.

Those exposed to high levels of aircraft noise during sleep may try to adapt to the noise using various sleeping aids, such as putting in earplugs or closing windows. We found that L_{night} was significantly associated with an increased likelihood to frequently close windows when trying to sleep and to use alcohol, television and/or music as sleep aids because of noise. These findings suggest that residents in communities with higher L_{night} are concerned with noise affecting their sleep, and they engage in coping behaviors to adapt to the noise at night. A limitation is that our survey questions on sleep aids referenced noise in general—rather than aircraft noise specifically. It may be possible that residents living in neighborhoods with higher L_{night} use sleep aids to block out other sources of nighttime noise as well. However, given that these residents were significantly annoyed and disturbed in their sleep by aircraft noise it is plausible that aircraft noise was the primary noise source that induced these coping behaviors.

In the long-term, exposure to high levels of aircraft noise may have adverse health consequences [102, 103]. It is thought that nighttime aircraft noise exposure increases the risk of cardiovascular disease [104] and is known to disturb sleep, which—when restricted on a chronic basis—is associated with increased risk of cancer, obesity, and diabetes [105]. However, we did not find an association between L_{night} and poorer self-reported general health after adjusting for individual-level covariates and sociodemographic factors. Nor did we find an association between L_{night} and diagnosis of heart disease, hypertension or diabetes. However, we were underpowered to detect the small effect sizes expected for these health outcomes. However, the significant relationships between BMI and age with a number of the health outcomes are all positive, as would be expected for sleep disorders, hypertension, heart disease and diabetes, indicating that the questionnaire items may be suitable for capturing the prevalence of diagnosis among the sampled population.

5. Limitations

There are a number of limitations with results of the survey results, most notably that our sample size was small (268 surveys or 8.5% of the surveyed population). Our response rate was lower (10.1%) than the 46-76% response rates seen in other postal questionnaires on attitudes towards aircraft noise [58], and so survey responses may not accurately represent the attitudes and sleep patterns of the population around Atlanta airport. However, this survey was primarily aimed at recruitment for a field study, and its response rate may not be comparable to other attitudinal questionnaires). Additionally, we did not have information on noise exposure levels in the bedrooms of survey participants. Our survey study used estimated outdoor nighttime aircraft exposure levels based on flight traffic data; however, these estimates may not always reflect actual noise levels in the bedroom (see section V.F). If residents close their windows at night, noise levels can be diminished by up to 28 dB [106], and indeed in the field study we measured lower indoor levels among participants who slept with the windows closed compared to those who did not (see Figure 27 and Figure 28 in section V.F on Noise Expose Validation). Aircraft noise can also be masked by noise from air conditioning, television or white noise machines. Accurate bedroom noise levels can only be obtained with measurement, such as was performed for a small sample of survey respondents in the field study. Lastly, because of the exploratory design of this study, we decided not to correct for multiple testing, and therefore inferences drawn from these tests may not be reproducible [107]. Despite these limitations, evidence of adverse effects of aircraft noise warrants further investigation in larger subject cohorts.

C. Non-participation analysis

It is important that participants in the field study are representative of the population from which they are recruited. We therefore compared demographic data for survey respondents who participated or did not participate in the field study. The percentages of participant and non-participant race, sex, age, L_{Night}, marital status, household income, education, employment, noise sensitivity, sleep disturbance by aircraft noise at home, general health and sleep quality over the past month are presented in Appendix 5.

A number of categories for certain items had low numbers of responses, and the total sample size (n=407) was less than 1000. We therefore used Fisher's exact test of independence, rather than the more general χ^2 test, to test whether the proportion of responses for the demographic variables were different between participants and non-participants [108]. Exact p-values are reported in the figure captions. There were no statistically significant indications of differences between the participation groups for race, sex, age, L_{Night}, marital status, household income, education, employment, noise sensitivity or sleep disturbance by aircraft noise at home. There were statistically significant differences between the participation groups for self-reported general health (p=0.0004) and sleep quality over the past month (p=0.023).

The lack of differences between participation groups for the majority (10 of 12) of the variables suggests a good representativeness of the field study participants relative to the wider population. However, a greater proportion of non-participants rated their sleep quality and general health as worse than the field study participants. In other words, survey respondents with poorer health and/or sleep quality were less likely to enroll in the field study. Part of this may be explained by subjects not meeting eligibility criteria for inclusion into the study.

Fifty eight percent of respondents were black, which is a similar proportion to the 62.5% mean proportion for the sampled sampling region (calculated based on proportions and number of houses in each sampling region, see Table 3). The mean age of respondents was 53.0 years. Although this is greater than the mean age of 35.1 years for the sampled census tracts (calculated as (Σ(houses per tract × mean age per tract)) / total number of houses in all tracts), see Table 3), this is an expected result since census data includes children, whereas our survey contact letters specified that respondents must be an adult (Appendix 3). For highest level of completed education, 48.6 % of respondents had no college education, which is slightly higher, although not greatly so, than the 40.0% of the population without college education in the sampling region (calculated based on proportions and number of houses in each sampling region, see Table 3). For annual household income, 42.6% of all responses were in the 25-50k category and below, and 61.6% of all responses were in the $50-75k category. The median household income of respondents is therefore between $50-75k. The mean for the sampled census tracts is $49.0k (see Table 3 calculated based on proportions and number of houses in each tract). On one hand, if the median income was in the lower range of the $50-75k category, there would be a good agreement between survey respondents and the general population in the sampling region. On the other hand, an income in the higher range of the category would suggest that respondents earned substantially more than their counterparts. In the absence of more precise income data, no firm conclusions can therefore be drawn in this regard.

D. Field study morning questionnaires

Six participants entered questionnaire data directly into RedCap using their own computer. Twenty eight participants completed paper versions of the same questionnaire (Appendix 4).

There were a total of 165 completed questionnaires from 33 field study participants (expected N=170). One participant did not complete the morning questionnaires during the field study.

Results of the crude models are presented in Table 19, with equivalent ($L_{AEq,sleep}$) or maximum ($L_{AS,max}$) noise level during the individualized sleep period as the independent variables. The number of self-reported awakenings and sleep disturbance by aircraft noise significantly increased with increasing $L_{AS,max,sleep}$. There were no statistically significant effects for any other outcomes in the crude models.

Table 19 Effect of equivalent nighttime aircraft noise ($L_{AEq,sleep}$) or maximum aircraft noise ($L_{AS,max}$) during sleep on questionnaire outcomes. Crude noise-only model. Parameter estimates are presented as regression coefficients (β). Statistically significant (p<0.05) Type III effects are highlighted in bold typeface. df=degrees of freedom. CI=confidence interval. SSS=Stanford Sleepiness Scale.

Response	$L_{AEq,sleep}$		$L_{AS,max}$	
	p	β (95% CI)	p	β (95% CI)
Sleep latency (minutes)	0.448	-0.512 (-1.835; 0.811)	0.552	0.141 (-0.323; 0.604)
Awakenings (n)	0.075	0.031 (-0.003; 0.065)	**<0.001**	**0.037 (0.019; 0.054)**
Tiredness (0-10)	0.571	0.046 (-0.113; 0.205)	0.058	0.069 (-0.002; 0.141)
Sleepiness (dichotomous SSS)	0.792	-0.019 (-0.157; 0.120)	0.438	-0.029 (-0.102; 0.044)
Difficulty falling asleep (0-10)	0.444	-0.049 (-0.173; 0.076)	0.495	0.030 (-0.055; 0.115)
Sleep restlessness (0-10)	0.229	-0.086 (-0.226; 0.054)	0.844	0.009 (-0.083; 0.101)
Sleep quality (0-10)	0.959	0.005 (-0.189; 0.199)	0.134	0.059 (-0.018; 0.135)
Disturbance by aircraft noise (dichotomous)	0.334	0.133 (-0.137; 0.403)	**0.003**	**0.106 (0.036; 0.175)**

Results of the adjusted models are presented in Table 20. There was quasi-complete separation of the data, whereby the dichotomous sleepiness variable separated the predictor variables to a certain degree, and therefore the regression model could not estimate the maximum likelihood ratio. Where complete or quasi-complete separation occurred, the problematic predictor variable were excluded from the model.

No statistically significant effects of $L_{AEq,sleep}$ were found. With increasing $L_{AS,max}$ there were significant increases in tiredness (β=0.005, p=0.005) and , as with the crude model, self-reported awakenings (β=0.051, p<0.001). These findings provide some support to the hypothesis that nocturnal aircraft noise can have adverse effects on sleep. Physiologic awakening probability increased with the maximum noise level of a discrete aircraft noise event (see section V.E), and based on the questionnaire data the participants seem to recall at least some of these awakenings. Furthermore, recalled awakenings can have a moderate correlation with self-reported tiredness [109].

There was a significant effect of the number of airplane noise events on sleepiness in the $L_{AEq,sleep}$ model, and tiredness in $L_{AS,max}$ model, whereby participants reported lower sleepiness and lower tiredness with higher numbers of airplanes. On the one hand, this could indicate that individuals who are chronically exposed to a high number of aircraft noise events habituate to the exposure. There is evidence that physiologic habituation to single noise events occurs within nights, but not between-nights in the short-term, particularly for autonomic arousal [110]. However, in the long-term, there might be some level physiologic habituation to nocturnal noise, but this habituation does not seem to be total, i.e. at least some degree of response persists [111]. An alternative explanation for the finding of lower sleepiness and tiredness could be that individuals exposed to a high number of events over time have more impaired sleep than counterparts exposed to fewer events. Incognizant of this decreasing objective sleep quality, they may downwardly adjust their criteria for what they consider as good subjective sleep, i.e. they “get used” to this poorer sleep as the norm, such as seems to occur with aging [112]. Such a process would manifest superficially as a psychological habituation, with lower levels of sleepiness and tiredness than an individual may have reported previously even with adversely impacted sleep physiology. However, in the absence of more detailed data on the objective sleep of the participants in the current field study, both explanations of the lower tiredness and sleepiness following nights with a higher number of aircraft noise events remain speculative.

There was a significant effect of sex on tiredness in the $L_{AEq,sleep}$ model, whereby men were less tired than women. There was a similar effect in the $L_{AS,max}$ model, but the result was of borderline statistical significance (p=0.052). This result is in line with some earlier work finding that although women may have better objective sleep than men, they frequently report greater sleep disturbance [113], and are at increased risk for developing sleep disorders including insomnia and restless legs syndrome [114, 115].

There was a significant effect of age on sleep latency in the $L_{AEq,sleep}$ model. This increasing sleep latency with age is in line with typical age-related alterations in sleep [116].

There were significant effects of sleeping with open windows on sleep latency in the $L_{AEq,sleep}$ and $L_{AS,max}$ models, and on awakenings in the $L_{AS,max}$ model. With fully or partially open windows, sleep latency was shorter and there were fewer recalled awakenings. One possible explanation is that individuals who find it difficult to sleep, and therefore have longer sleep latencies and more awakenings, may be more likely to close their window to lower noise levels. Alternatively, open windows, while resulting in higher indoor noise levels, could lead to better air quality and temperature in the bedroom, which may *per se* help promote certain aspects of subjective sleep [117].

No statistically significant effects were found for any of the independent variables in either of the $L_{AEq,sleep}$ and $L_{AS,max}$ models for difficulty falling asleep, sleep restlessness, sleep quality or disturbance by aircraft noise. The absence of an effect on disturbance by aircraft noise in particular is surprising, as self-reported sleep disturbance by a particular noise source has frequently been reported in the literature [11, 118]. However, as part of eligibility for the field study, participants did not regularly use sleep medication, did not suffer from sleep disorders, and were generally free from internal and external factors that could interfere with sleep, all of which indicates they were habitually good sleepers. Taken with the fact that they generally reported better sleep quality than postal survey respondents who did not participate in the field study (see section V.C), the current study population may represent a particularly resilient subgroup who do not feel their sleep is disturbed, or the size of any disturbance effect in this group was too small to be detected with our limited sample size.

Table 20 Effect of equivalent nighttime aircraft noise ($L_{AEq,sleep}$) or maximum nighttime aircraft noise ($L_{AS,max}$) during sleep on questionnaire outcomes. Fully adjusted model. Parameter estimates are presented as regression coefficients (β). Statistically significant ($p<0.05$) Type III effects are highlighted in bold typeface. * Reference category=women. † Reference category=completely closed. ‡ Excluded from model due to quasi-complete separation. df=degrees of freedom. CI=confidence interval. SSS=Stanford Sleepiness Scale.

		$L_{AEq,sleep}$		$L_{AS,max}$	
Response	**Independent variable**	**p**	**β (95% CI)**	**p**	**β (95% CI)**
Sleep latency (minutes)	Noise ($L_{AEq,sleep}$/L_{ASmax})	0.090	-0.947 (-2.041; 0.146)	0.482	-0.181 (-0.686; 0.324)
	Sex *	0.268	5.456 (-4.206; 15.117)	0.739	-0.086 (-0.595; 0.422)
	Number of planes	0.143	0.140 (-0.047; 0.327)	0.332	0.097 (-0.099; 0.293)
	Age	**0.013**	**0.443 (0.095; 0.790)**	**0.024**	**0.457 (0.061; 0.853)**
	Windows †	**0.012**	**-11.769 (-20.904; -2.635)**	**0.013**	**-13.392 (-24.012; -2.773)**
Awakenings (n)	Noise ($L_{AEq,sleep}$/L_{ASmax})	0.079	0.040 (-0.005; 0.085)	**<0.001**	**0.051 (0.028; 0.074)**
	Sex (ref=women)	0.857	-0.039 (-0.468; 0.390)	0.467	0.161 (-0.272; 0.593)
	Number of planes	0.978	0.000 (-0.006; 0.006)	0.263	-0.004 (-0.011; 0.003)
	Age	0.074	0.014 (-0.001; 0.030)	0.067	0.011 (-0.001; 0.022)
	Windows †	0.063	-0.578 (-1.187; 0.031)	**0.016**	**-0.783 (-1.418; -0.148)**
Tiredness (0-10)	Noise ($L_{AEq,sleep}$/L_{ASmax})	0.322	0.092 (-0.090; 0.273)	**0.005**	**0.118 (0.036; 0.199)**
	Sex *	**0.008**	**-2.054 (-3.579; -0.530)**	0.052	-1.591 (-3.195; 0.014)
	Number of planes	0.070	-0.022 (-0.045; 0.002)	**0.001**	**-0.031 (-0.048; -0.013)**
	Age	0.275	0.026 (-0.021; 0.074)	0.471	0.018 (-0.031; 0.068)
	Windows †	0.373	-0.887 (-2.837; 1.063)	0.140	-1.365 (-3.177; 0.447)
Sleepiness (dichotomous SSS)	Noise ($L_{AEq,sleep}$/L_{ASmax})	0.558	0.046 (-0.108; 0.200)	0.832	0.012 (-0.097; 0.121)
	Sex *	0.131	-1.808 (-4.153; 0.537)	0.145	-1.805 (-4.233; 0.624)
	Number of planes	**0.026**	**-0.050 (-0.094; -0.006)**	0.072	-0.051 (-0.106; 0.005)
	Age	0.463	0.020 (-0.033; 0.073)	0.462	0.018 (-0.031; 0.067)
	Windows †‡		-		-
Difficulty falling asleep (0-10)	Noise ($L_{AEq,sleep}$/L_{ASmax})	0.472	-0.053 (-0.196; 0.091)	0.428	0.044 (-1.197; 0.154)
	Sex *	0.976	0.021 (-1.386; 1.429)	0.750	0.233 (-1.197; 1.663)
	Number of planes	0.874	-0.001 (-0.017; 0.015)	0.298	-0.010 (-0.028; 0.009)
	Age	0.056	0.045 (-0.001; 0.091)	0.050	0.043 (0.000; 0.085)
	Windows †	0.176	-1.045 (-2.559; 0.469)	0.083	-1.420 (-3.023; 0.183)
Sleep restlessness (0-10)	Noise ($L_{AEq,sleep}$/L_{ASmax})	0.375	-0.069 (-0.221; 0.083)	0.560	0.033 (-0.077; 0.143)
	Sex *	0.224	-0.835 (-2.181; 0.512)	0.323	-0.668 (-1.993; 0.657)
	Number of planes	0.293	-0.008 (-0.024; 0.007)	0.096	-0.016 (-0.036; 0.003)
	Age	0.115	0.032 (-0.008; 0.072)	0.105	0.031 (-0.006; 0.068)
	Windows †	0.094	-1.212 (-2.629; 0.205)	0.068	-1.559 (-3.231; 0.113)

Table 20 continued

		$L_{AEq,sleep}$		$L_{AS,max}$	
Response	**Independent variable**	**p**	**β (95% CI)**	**p**	**β (95% CI)**
Sleep quality (0-10)	Noise ($L_{AEq,sleep}$/L_{ASmax})	0.587	-0.063 (-0.292; 0.165)	0.122	0.058 (-0.016; 0.132)
	Sex *	0.507	-0.563 (-2.227; 1.101)	0.731	-0.290 (-1.943; 1.363)
	Number of planes	0.465	0.009 (-0.016; 0.034)	0.894	-0.001 (-0.023; 0.020)
	Age	0.188	-0.038 (-0.094; 0.018)	0.137	-0.040 (-0.093; 0.013)
	Windows †	0.127	1.640 (-0.465; 3.745)	0.275	1.157 (-0.921; 3.235)
Disturbance by aircraft noise (dichotomous)	Noise ($L_{AEq,sleep}$/L_{ASmax})	0.433	0.060 (-0.089; 0.208)	0.183	0.092 (-0.043; 0.226)
	Sex *	0.668	-0.456 (-2.539; 1.627)	0.962	-0.053 (-2.235; 2.129)
	Number of planes	0.378	0.016 (-0.019; 0.051)	0.464	0.015 (-0.025; 0.055)
	Age	0.378	0.023 (-0.136; 0.183)	0.859	0.015 (-0.151; 0.181)
	Windows †	0.763	-0.285 (-2.137; 1.568)	0.263	-0.765 (-2.106; 0.575)

In summary, only minimal effects of aircraft noise were found on self-reported sleep outcomes. Maximum and average nighttime aircraft sound pressure levels have previously been found to predict event-related awakenings [11]. Accordingly, even with a small sample size by questionnaire study standards, we saw a statistically significant increase in the number self-reported awakenings with increasing $L_{AS,max}$, although the effect of $L_{AEq,sleep}$ was of only borderline significance (p=0.079), which could be due to insufficient statistical power resulting from the limited sample size.

There may have been some misinterpretation of the questionnaire response scales among the participants, as response scales were sometimes in different directions relative to one another. For instance, the three items of question 8 (ease of falling asleep, sleep restlessness and sleep quality) had a 0-10 response scale, with a value of 10 indicating the worst sleep on two scales (most difficult to sleep and most restless), but the best sleep quality. In one case, a participant rated themselves as very restless (10 out of 10) and having difficulty falling asleep (8 of 10), but with very good sleep quality (9 out of 10). It is unlikely, albeit not impossible, that this rating of sleep quality is accurate, but instead results from the inversion of the response scale. Rather than taking what would be an unethical and unscientific approach of trying to guess what we believed the respondent intended, we always used the actual responses. In future field studies it will be important to minimize the possibility of confusion or misinterpretation of any questionnaire items, improving data quality.

E. Event-related analysis

This section describes the physiologic event-related response to aircraft noise events during sleep.

1. Study participants and data loss

Thirty-four subjects consented to participate in the study, and provided at least some data (a single subject consented but did not participate in the measurements nor returned the equipment). Of the 34 subjects, the acoustical calibration before the equipment was sent out and after it was returned differed by >2 dBA and was considered invalid in 10 subjects. These differences were caused by an unprotected gain controller that, likely unwillingly, was moved by research staff or study participants after initial calibration (see Figure 22).

Figure 22: Sound recorder gain controller issues. The left pane shows two gain controllers behind a metal bar. As the protection of these controllers is minimal, the position of the controllers was changed in N=10 study participants after initial calibration. A 3D-printed gain control stabilizer (middle pane) was used for all remaining measurements. In the final approach, which will be used in the future National Sleep Study (but was not implemented around ATL), the gain controller was fixed in one position with a hot glue gun before calibration (right pane).

Of the remaining 24 subjects, one subject contributed only one valid night of physiological data, and only a single aircraft noise event was recorded in this period. In another subject, not a single aircraft noise event was recorded during the measurement nights. These two participants were thus excluded from data analysis. Therefore, 22 subjects (8 male; mean ± SD age 50.0 ± 14.0 years; mean ± SD BMI 27.8 ± 3.3 kgm^{-2}) contributed to the final analysis. A total of 1,900 aircraft noise events were recorded in the bedroom. In 154 aircraft noise events (8.1%), no physiological data were available. Finally, in 79 aircraft noise events (4.2%), an awakening reaction started prior to the start of the aircraft noise event, and so were excluded. A total of 1,667 aircraft noise events (87.7% of 1,900) therefore contributed to the data analysis.

2. Aircraft noise levels

The distribution of indoor maximum noise levels for the 1,667 aircraft events within participant’s homes that contributed to the data analysis is shown in Figure 23. The average $L_{AS,max}$ of aircraft events was 40.1 dB (median 39.4 dB, range 28.9 dB-63.4 dB). A distribution of average noise levels in the minute preceding the start of each aircraft noise event is also shown in Figure 23 (average 30.9 dB, median 29.8 dB, range 22.4 dB-56.5 dB).

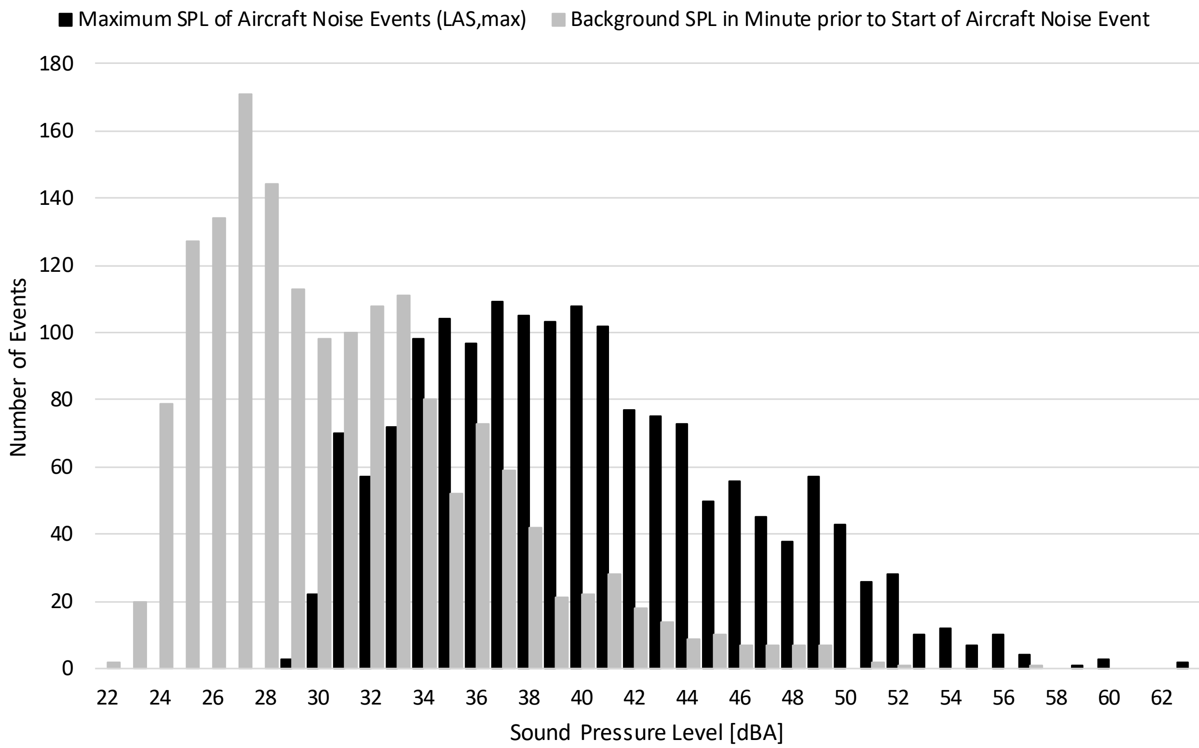


Figure 23. Indoor noise levels for participants near the airport. Black: L_{ASmax} of aircraft events; Gray: L_{AEq} one minute before the start of each aircraft event.

The number of events per night per subject who lived near the airport is shown in Figure 24. Out of the 22 participants that contributed to data analysis, the median number of aircraft noise events experienced across the 5 study nights was 43 (range 5-297).

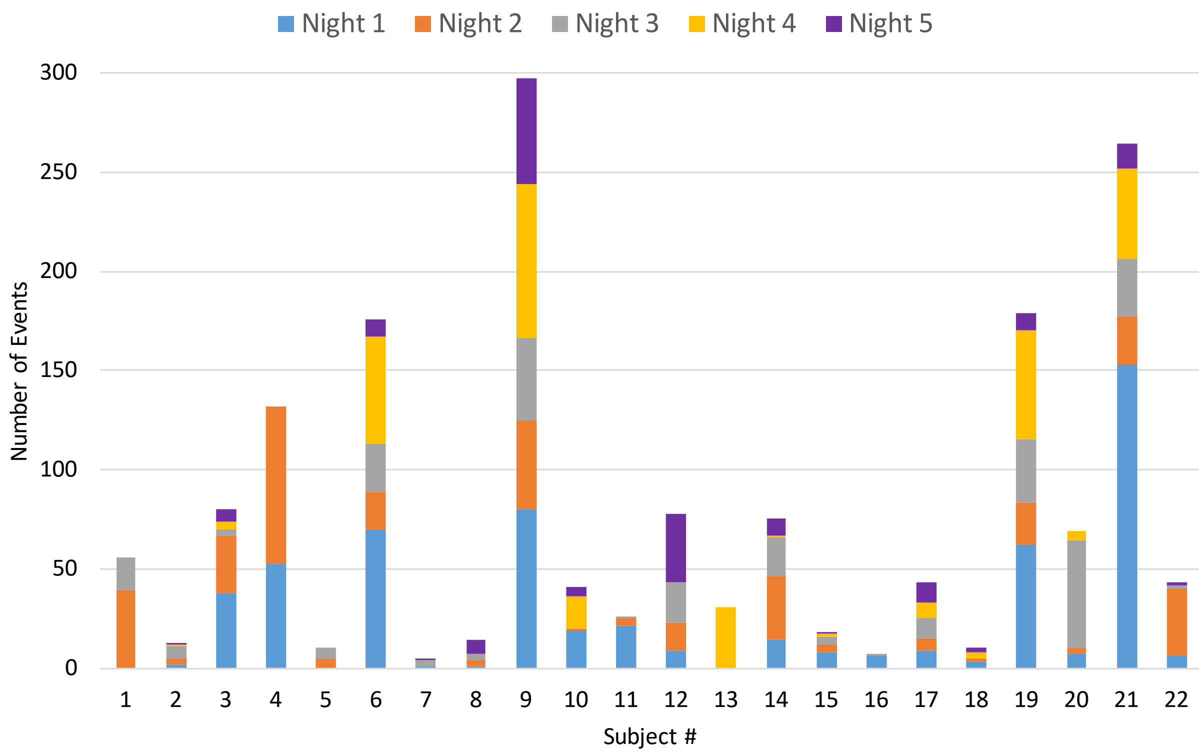


Figure 24. Number of aircraft noise events per subject near ATL airport for each of the 5 study nights. The colors indicate study nights.

3. Single event awakening analysis

Random intercept logistic regression models were calculated for the probability of awakening to an aircraft. Model 1 contained only the indoor maximum noise level, Model 2 was adjusted for age, sex, BMI, and time from sleep onset (Table 21). A total of 1,667 aircraft noise events contributed to the analysis. In both models the coefficient for L_{ASmax} was positive (i.e., awakening probability increased statistically significantly with increasing L_{ASmax}) but not statistically significant, likely due to the low sample size and power of the study. In Model 2, adjustment had little influence on the estimate of the coefficient for L_{ASmax} (0.0288 in Model 1 vs. 0.0254 in Model 2, respectively). None of the investigated confounders (age, sex, BMI, and time from sleep onset) had a statistically significant influence on awakening probability.

Table 21. Random effect logistic regression models for the probability of awakening

	Model 1			**Model 2**		
	Estimate	**SE**	**p-value**	**Estimate**	**SE**	**p-value**
L_{ASmax} [dB]	0.0288	0.0148	0.0647	0.0254	0.0126	0.0572
Age [years]				-0.0054	0.0052	0.3159
Male				-0.1359	0.2910	0.6454
BMI				-0.0021	0.0304	0.9450
Time [min]				-0.0005	0.0005	0.3346

SE: Standard Error

The exposure-response relationship for additional awakenings due to aircraft events (P_{noise}-$P_{spontaneous}$), based on unadjusted Model 1 above, is shown in Figure 25. To account for spontaneous awakenings in the exposure-response function [119], an estimate statement was used in NLMIXED to subtract awakening probability at 29 dB from the awakening probability at the maximum SPL of interest. The threshold of 29 dB was based on the median background noise level one minute prior to the start of the aircraft noise events in this study (29.8 dB). Due to the relatively low number of subjects and aircraft noise events per subject, the 95% confidence interval of the exposure-response function is relatively wide. As the p-value for the $L_{AS,max}$ estimate was ≥0.05, the 95% confidence intervals in Figure 25 includes 0% for higher noise levels (negative estimates were converted to 0%).

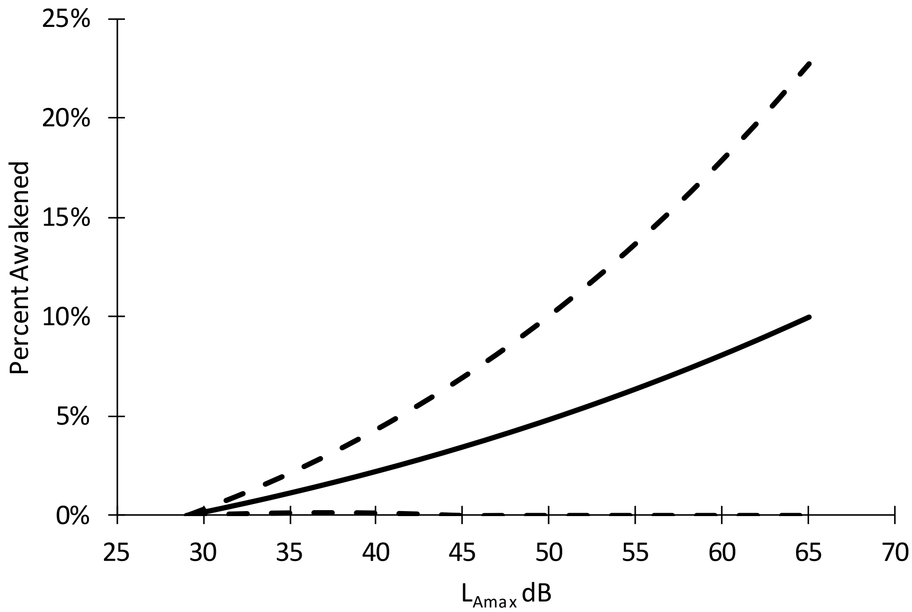


Figure 25. The unadjusted probability of an additional awakening induced by aircraft noise depending on indoor maximum SPL L_{Amax} (slow time weighting) for ATL International Airport. Dashed lines indicate 95% confidence intervals.

F. Noise exposure validation

Selection of geographical areas from which to recruit field study participants and analysis of survey response data were both based on modelled outdoor, rather than measured indoor, aircraft noise data. Effects of noise on sleep depend on noise levels during sleep in the bedroom, and therefore are affected by the accuracy of the modelling and GIS coding, sound insulation (including window opening and closing), position of the bedroom in the dwelling (for instance facing towards or away from flight paths), and sleep times of the occupants (for instance if they sleep during low- or high- air traffic volume times The modelled outdoor aircraft noise exposure L_{night} correlated significantly with $L_{AEq,sleep}$ (r=0.63, p=0.001) and $L_{AS,max,sleep}$ (r=0.57, p=0.004), shown in Figure 26. There were lower measured aircraft noise levels in the bedroom when participants closed their windows (mean ± SD level 27.2±0.4 dB $L_{AEq,sleep}$; 44.8±0.9 dB $L_{AS,max,sleep}$) compared to when it was partially or completely open (30.2±1.2 dB $L_{AEq,sleep}$; 52.7±2.5 dB $L_{AS,max,sleep}$). These data stratified by window closing were averaged across all noise exposure categories, and we do not have outdoor noise measurements, so the difference between the window closed and partially/completely open groups does not reflect the noise reduction effect of closing a window. For instance, individuals with higher outdoor aircraft noise levels were more likely to close their windows [51], which may also be the case in this field study. Given the number of factors that can influence indoor noise levensl compared to L_{Night}, the correlation coefficient indicates a rather good capability of the noise modelling procedure to predict average aircraft noise levels in the bedroom during sleep.

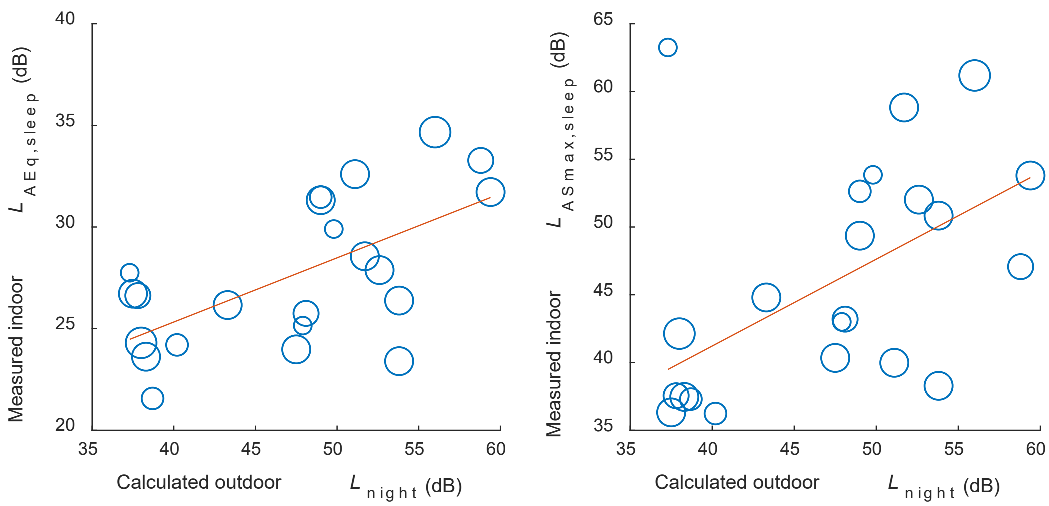


Figure 26 Scatter plot of between calculated outdoor L_{night} (abcissa) and mean measured indoor aircraft noise level during sleep for each participant (ordinate; left pane $L_{Aeq,sleep}$; right pane $L_{ASmax,sleep}$). The number of observations (nights) for each participant is indicated by the circle radius. The least squares regression line, calculated with weighted data, is shown in red..

Cross-sectional information on the influence of window closing/opening on indoor noise level can be determined by stratifying measured aircraft noise levels by the morning questionnaire item on window position. As anticipated, there were generally lower noise levels, both sleep period average (Figure 27) and maximum (Figure 28), when the window was closed compared to when it was partially or completely open. Note that these data were averaged across all noise exposure categories, and we do not have outdoor noise measurements, so the difference between the window closed and partially/completely open groups does not necessarily reflect the noise reduction effect of closing a window. For instance, as found in the postal surveys, individuals with higher outdoor aircraft noise levels were more likely to close their windows (see section V.B.3), which may also be the case in the field study.

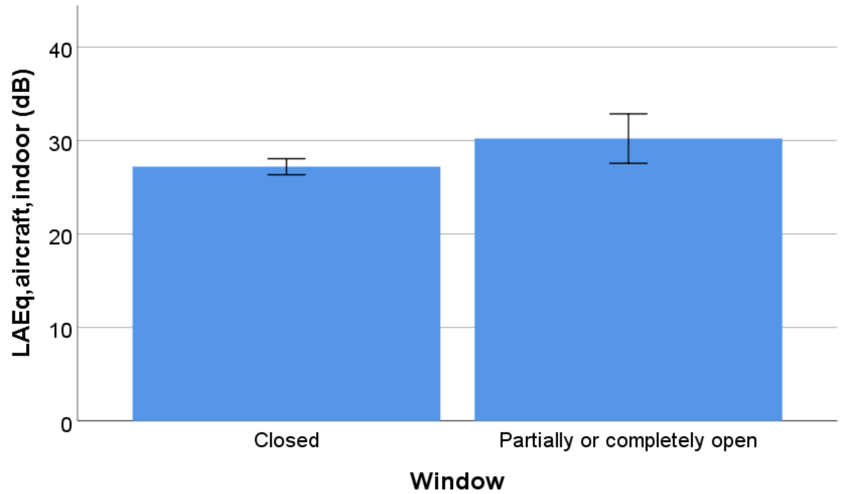


Figure 27 Mean L_{AEq} during sleep stratified by window position during the night. Error bars indicate 95% confidence intervals.

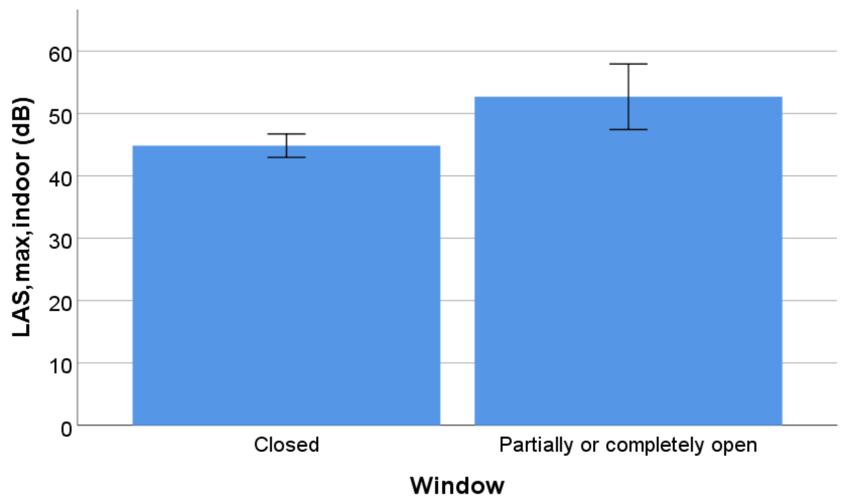


Figure 28 Mean $L_{AS,max}$ during sleep stratified by window position during the night. Error bars indicate 95% confidence intervals.

G. Feasibility assessment

The following sections describe whether the approach adopted in the study presented in this report is feasible to implement on a larger scale in the National Sleep Study.

1. Cost effectiveness of postal surveys for study recruitment

In rounds 1-5, the gift card amount was randomized among respondents, so we used the mean cost of the possible $2, $5 and $10 amounts ($5.67) in the cost calculations. In rounds 6-17, 12.4% of initial survey waves were non-deliverable and returned to us with the $2 cash incentive still included. For each individual survey that was delivered, an average of $0.248 (i.e. 12.4% of $2) was recouped from these non-deliverable initial waves, and accounted for in the cost calculations. The costs for each individual survey and follow-up wave mailed out, the total cost per individual and the resulting total cost to receive a single completed survey are presented in Table 22, stratified by the different survey sampling protocols. The number of surveys sent out to receive a single response are the reciprocals of the response probabilities in Table 10. These data do not account for any associated personnel costs.

Table 22 Survey sampling cost effectiveness, ordered from the most to least cost effective method to receive a single completed survey.

Sampling protocol			**Surveys needed to receive 1 response (n)***	**Surveys sent to recruit 1 participant (n)*#**	**Costs ($)**							
Follow-up waves (n)	Survey length	Survey incentive			Initial wave	Follow-up wave 1	Follow-up wave 2	Follow-up wave 3	Total per mailed individual	Per response received*	Total to receive 1 response †	Recruit 1 participant†#
3	Short	$2	4.61	50.7	3.01	0.70	1.01	1.01	5.74	26.44	28.89	317.51
3	Medium	$2	4.88	53.6	3.09	0.70	1.09	1.09	5.96	29.09	31.84	349.88
0	Long	$2	12.20	134.1	3.09	-	-	-	3.09	37.65	39.54	434.48
3	Long	$2	6.13	67.4	3.09	0.70	1.09	1.09	5.96	36.59	39.99	439.50
2	Long	$2	8.33	91.5	3.09	0.70	1.09	-	4.88	40.64	44.01	483.66
0	Long	Gift card	32.26	354.5	1.09	-	-	-	1.09	40.83‡	46.81‡	503.38

*Assumes 100% delivery rate
†Assumes 87.6% delivery rate and, if applicable, $0.248 recouped from non-deliverable initial survey waves.
‡Includes a mean gift card cost of $5.67
#Assumes 9.1% participation rate from completed surveys across all survey mailing rounds, independent of mailing protocol. Does not include cost for actual participation in the field study ($150 or $200).

The most cost effective approach was the short survey with a $2 cash incentive and 3 follow-up waves, whereby on average 50.7 surveys were sent, with a total associated cost of $317.51, to recruit one participant into the field study. A slightly higher number of medium length surveys were sent to recruit one participant (n=53.6), which when combined with the slightly higher cost of mailing each individual survey resulted in a total associated cost of $349.88, to recruit one participant into the field study. The long surveys were the least cost effective approaches, due to the lower response rates.

The most inexpensive sampling protocol had the lowest response rate, with the consequence that it the least effective approach in terms of the financial cost to receive one completed survey. Conversely, the three sampling protocols with three follow-up waves were the most expensive, but when using the short and medium length survey were the most cost effective approaches owing to their increased response rates. The short survey was the most cost effective in terms of materials due to a slightly lower cost and a higher response rate. We required additional telephone contact with the short survey respondents to obtain further information regarding field study eligibility, but since personnel costs were not included, this approach may not truly be the most cost effective approach overall for field study recruitment.

Three follow-up waves approximately doubled the response rate compared to sending no follow-up. The additional cost of those follow-up waves ($2.88 for long surveys) was comparable with the cost of mailing a new long survey to a new household with no follow-ups ($3.09), hence both approaches could be anticipated to yield similar response rates at similar costs. This is consistent with findings reported by Mayfield et al. [120]. It would be preferable to increase response rate from initial non-responders to minimize bias and increase the representativeness of the sample.

2. Study attrition

An overview of attrition of recruitment of study participants is given in Figure 29. Of 237 survey respondents interested in the field study, only 79 met the inclusion criteria. Of those 79 who were eligible, 64 were contacted and sent consent forms for review. The main reason for not sending consent forms was being unable to reach survey respondents by telephone, typically because they did not respond to voicemail messages left by the research team, who were therefore unable to confirm their interest and eligibility. Of the 64 respondents who were sent consent forms, 45 consented and signed and returned the forms. Of those 45 who consented, 37 were enrolled into the field study and sent the equipment. Of those who were enrolled, 3 dropped out before the start of their study period, resulting in a total of 34 participants who completed (or partially completed) the study.

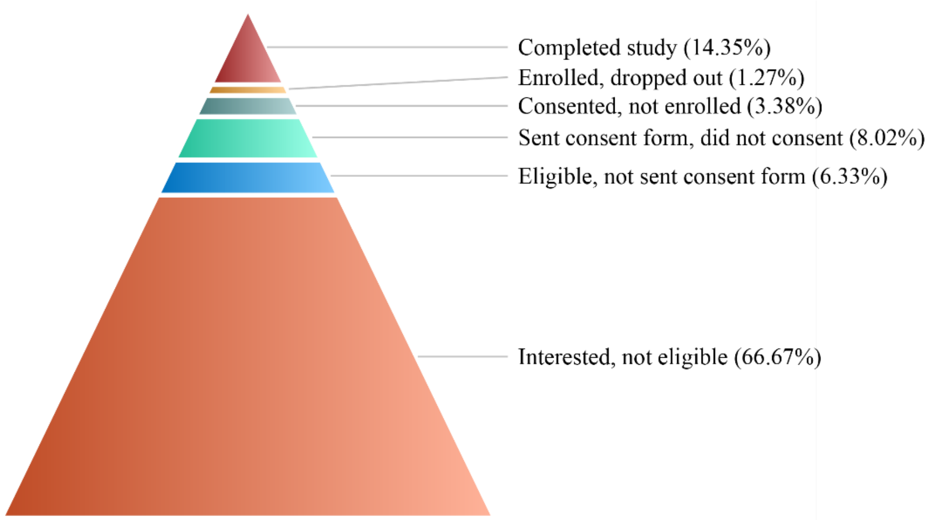


Figure 29 Graphical illustration of attrition at progressive stages of field study recruitment and implementation. Percentages are relative to the total number of survey respondents who indicated an interest in participating in the field study (n=237).

3. Study compliance and data loss

The purpose of this pilot study was to determine feasibility for a large-scale national field study. Thirty-four participants recorded acoustic and physiological data using equipment sent directly to their homes for 5 nights. In total, we received 160 nights of ECG data out of an anticipated 170 nights (94.1% successful data collection; Table 23). Participants also recorded 153 nights of acoustic data (90.0%), and completed 165 morning questionnaires (97.1%). Two participants accidentally began data collection a day ahead of schedule on the Sunday evening, but both agreed to record an extra night of physiological and acoustic data so that we had data from the same days of the week as other participants. These extra two days of data were included in both physiologic and morning questionnaire analysis. One participant found the Faros 90 device somewhat uncomfortable and collected only 3 nights of ECG data, but continued to collect acoustic data and complete the morning questionnaires. The most common reason for missing acoustic data was failure of the participants to initialize correctly the noise recorder. For one participant, only three nights of acoustic data were collected due to a technical error prior to sending them the equipment.

In order to perform analysis of study data, measurements of physiological and acoustic data must occur concurrently. Participants did not always record physiological and acoustic data in the same evenings. In total, we received 149 nights of overlapping acoustic and physiological data (87.6% data matching). For 9 subjects, acoustic data were excluded from the analysis due to large discrepancies in the sound recorder calibration value prior to and after completing study measurements. During transit and while participants are handling equipment, the calibration dials of the sound recorder can shift. When this happens, it is no longer possible to convert the mp3 recording to sound pressure levels, which is necessary to determine the $L_{AS,max}$ of aircraft noise events. After removing a total of 38 acoustic data files from the analysis due to calibration errors, our data collection rate was 65.3%. However, this problem was remedied for future subjects by securing the dials on the sound recorder in a fixed position with adhesive prior to shipment.

Table 23 A total of 170 overlapping nights of acoustic and ECG data were expected for 34 study participants. Amount of data collected and percent of usable data received is detailed below.

	ECG	Acoustic	Morning surveys	Matched ECG and acoustic data	Matched ECG and acoustic data included in analysis
Total Nights of Data Collected (n)	160	153	165	149	111
Proportion of anticipated data successfully collected (%)	94.1	90.0	97.1	87.6	65.3

4. Equipment loss

One set of equipment was lost, whereby after enrollment in the study one participant did not complete study measurements and could not be reached after repeated attempts at contact via phone and mail. Equipment was returned undamaged and in a timely manner by all other study participants enrolled in the study.

5. Summary

The study design is feasible to implement on a larger scale in the National Sleep Study. Thirty-four out of 37 enrolled participants recruited by postal surveys were able to receive and set-up study equipment, record measurements, and return equipment with minimal assistance from staff. In total, participants recorded 87.6% of requested data. Data loss as a result of calibration errors was remedied during the study and is not anticipated to be a continued problem in future studies.

VI. Acknowledgement

This research was funded by the U.S. Federal Aviation Administration Office of Environment and Energy through ASCENT, the FAA Center of Excellence for Alternative Jet Fuels and the Environment, project 017 through FAA Award Number 13-C-AJFE-UPENN-004 under the supervision of Sean Doyle and Natalia Sizov. Any opinions, findings, conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the FAA.

We would like to thank Sarah McGuire for her invaluable contributions during the design and data gathering stage of to this study. We would also like to thank Uwe Müller for spending 3 months at the University of Pennsylvania and for developing software critical for the success of the project. We would like to thank Emanuel Hermosillo and Jad Nasrini for assisting in the scoring of acoustic events. We are grateful to students working at the Unit for Experimental Psychiatry who assisted with survey mailings and data entry. Finally, we would like to thank colleagues form the German Aerospace Center (DLR) for their continued collaboration on projects on the effects of aircraft noise on sleep.

Major Accomplishments

Prepaid cash incentives and sending follow-up reminder and survey waves were an effective method of improving response rates to postal questionnaires. Although no factors of the different sampling protocols improved the probability of a respondent participating in the field study *per se*, using a pre-issued cash incentive and sending more follow-up waves, and subsequently improving response rates and achieving higher numbers of people from which to recruit, may be an effective strategy for improving recruitment into field studies.

Among postal survey respondents, calculated outdoor nighttime air traffic noise was significantly associated with self-reports of worse overall sleep quality, trouble falling asleep within 30 minutes, annoyance, and sleep disturbance. Residents in areas exposed to higher levels of aircraft noise coped by closing the windows at night. After adjustment for sociodemographic factors, we did not find a significant effect of nocturnal aircraft noise exposure on any of the investigated self-reported health outcomes. The low sample size and response rate are limitations of this study warranting a replication of the findings in larger, representative subject cohorts.

Postal survey respondents were, based on available census data, representative of their geographical region. The respondents who eventually participated in the field study were in many, but not all, ways similar to survey respondents who either did not wish to or were not eligible to take part in the field study. Recruitment by postal questionnaire is therefore a feasible approach in obtaining a large, representative sample for future studies around multiple airports.

Two thirds of survey respondents who were interested in the field study did not meet the eligibility criteria. Among the interested and eligible respondents there was some attrition at each stage of the study enrollment process, with 34 individuals (43% of interested and eligible respondents) eventually completing the study. Based on lessons learned during this pilot study, a lower attrition rate could be expected in future studies.

Data of sufficient quality and quantity to investigate the effects of aircraft noise on sleep were obtained, despite some data loss in the field study due to technical issues with the equipment and non-compliance among the participants. The technical issues were the main cause of data loss, and a number of approaches to minimize data loss during the field study were identified. Non-compliance was low, with both physiologic and acoustic data collected by the participants in 87.6% of all study nights. The study therefore demonstrates the feasibility of mailing equipment to participants to obtain unattended physiologic and acoustic measurement data.

The current study was an investigation among a sample population of limited size, living close to a single airport. The findings of physiologic and self-reported effects of aircraft noise on sleep may not be representative of response among a demographically diverse national study population exposed to different patterns of nocturnal aircraft noise. A larger-scale study on the effects of aircraft noise on sleep around a representative sample of US airports is needed to provide up-to-date exposure–response functions. The approach used in the present pilot study has been demonstrated to be feasible for the purpose of this National Sleep Study.

A. Recommendations for National Sleep Study

1. Methodological approaches demonstrated as feasible in the current ATL study

In the pilot study presented in this report, we demonstrated the feasibility of a number of key methodological approaches for the National Sleep Study. These include:

- Recruiting participants who are representative of their geographical area for a field study via postal questionnaires.
- Aside from field study eligibility, the postal surveys for recruitment are useful for collecting community response data in their own right.
- A recruitment strategy using a random sampling stratified by noise exposure strata ensured a broad range of measured noise level in the bedroom, allowing for a wide range of exposure in the physiologic exposure-response awakening curve.
- Using the measurement equipment deployed around ATL, to collect unattended noise and physiological data of sufficient quality over five consecutive nights.
- Collecting questionnaire data allowing for non-response analysis and non-participation analysis.
- Telephone contact on the first and last day of the field study, as well as offering 24-hour support should the participants require assistance, was effective at mitigating data loss.

2. Updates to methodology

Based on the findings of the pilot study presented in this report, we would make the following recommendations for changes in the study methodology when implementing the study on a national scale. The reasons for these recommendations can be found in the appropriate section of this report.

Questionnaires, including postal survey and field study morning survey

- Use a consistent number of levels in response scales; 5-point Likert.
- Ensure the direction of the response scales is consistent; leftmost is the most positive rating, rightmost is the most negative rating.
- Collect data allowing for non-response analysis and non-participation analysis.

Postal survey mailing strategy

- Send three follow-up waves: one reminder postcard after 7 days, a second paper copy of the survey after 28 days, and a third paper copy of the survey after 48 days.
- Include a $2 pre-paid cash incentive with the initial mailing.
- Use a medium length survey; around 26 questions.
- Include all field study eligibility questions in the survey.
- Offer mail response mode only with initial mailing, and offer both mail and online response modes with follow-up mailings.
- Offer a $150 incentive for volunteering for the field study.
- Omit "Department of Psychiatry" and "Unit for Experimental Psychiatry" from the recruitment survey envelope

Eligibility criteria for field study

- Change "children in the household under 5 years of age" to "Any individuals in the household requiring care during the night"

Field study

- Hot glue the gain control dials on the noise recorder firmly in place.
- Use the data analysis software developed in this project
- Consider time synchronicity issues between measurement devices, and correct deviations in the data streams accordingly

- Scoring all acoustic events in Akustikview in a given night is cumbersome, can take 2 hours or more, and is likely not feasible for the National Sleep Study . Efforts will be made to minimize manual effort in identifying aircraft noise events, which may include integrating flight rack radar data into the Akustikview software, or using scheduled flight operations data to identify periods in which to score acoustic events.

Publications

Peer Reviewed Journal Publications

Basner, M., Witte, M., & McGuire, S. (2019). Aircraft noise effects on sleep – Results of a pilot study near Philadelphia International Airport. International Journal of Environmental Research and Public Health, 16(17): 3178.

Rocha, S., Smith, M., Witte, M., & Basner, M. (2019). Survey results of a pilot sleep study near Atlanta International Airport. International Journal of Environmental Research and Public Health, 16(22): 432.

Smith, M., Rocha, S., Witte, M., & Basner, M. (2020). On the feasibility of measuring physiologic and self-reported sleep disturbance by aircraft noise on a national scale: A pilot study around Atlanta airport. Science of the Total Environment, 718: 137368

Smith, M., Witte, M., Rocha, S., & Basner, M. (2019). Effectiveness of incentives and follow-up on increasing survey response rates and participation in field studies. BMC Medical Research Methodology, 19: 230.

Published Conference Proceedings

Basner, M., Smith, M., Rocha, S., & Witte, M. (2019). Pilot field study on the effects of aircraft noise on sleep around Atlanta International Airport. Presentation and conference paper 23rd International Congress on Acoustics, Aachen, Germany.

Outreach Efforts

Conference presentations

Basner, M., Smith, M., Rocha, S. and Witte, M. Pilot Field Study on the Effects of Aircraft Noise on Sleep Around Atlanta International Airport. Conference paper to be presented as an oral presentation at 23rd International Congress on Acoustics, 2019, Aachen, Germany

Smith, M., Rocha, S., Witte, M. and Basner, M. Self-reported sleep disturbance by aircraft noise around Atlanta airport. Abstract only, presented as a poster at SLEEP 2019, San Antonio, TX

Awards

None

Student Involvement

None

Plans for Next Period

This task is compete.

References

1. Xie L, Kang H, Xu Q, Chen MJ, Liao Y, Thiyagarajan M, et al. Sleep drives metabolite clearance from the adult brain. Science. 2013;342(6156):373-7. doi: 10.1126/science.1241224. PubMed PMID: 24136970.
2. Naidoo N. Cellular stress/the unfolded protein response: relevance to sleep and sleep disorders. Sleep Med Rev. 2009;13(3):195-204. doi: 10.1016/j.smrv.2009.01.001. PubMed PMID: 19329340; PubMed Central PMCID: PMCPMC2964262.
3. Mackiewicz M, Shockley KR, Romer MA, Galante RJ, Zimmerman JE, Naidoo N, et al. Macromolecule biosynthesis: a key function of sleep. Physiol Genomics. 2007;31(3):441-57. doi: 10.1152/physiolgenomics.00275.2006. PubMed PMID: 17698924.
4. Rasch B, Born J. About Sleep's Role in Memory. Physiol Rev. 2013;93(2):681-766. doi: 10.1152/physrev.00032.2012. PubMed PMID: WOS:000317599400006.

5. Banks S, Van Dongen HP, Maislin G, Dinges DF. Neurobehavioral dynamics following chronic sleep restriction: dose-response effects of one night for recovery. Sleep. 2010;33(8):1013-26. PubMed PMID: 20815182; PubMed Central PMCID: PMCPMC2910531.
6. Cappuccio FP, Taggart FM, Kandala NB, Currie A, Peile E, Stranges S, et al. Meta-analysis of short sleep duration and obesity in children and adults. Sleep. 2008;31(5):619-26. PubMed PMID: 18517032; PubMed Central PMCID: PMCPMC2398753.
7. Knutson KL, Spiegel K, Penev P, Van Cauter E. The metabolic consequences of sleep deprivation. Sleep Med Rev. 2007;11(3):163-78. doi: 10.1016/j.smrv.2007.01.002. PubMed PMID: 17442599; PubMed Central PMCID: PMCPMC1991337.
8. Gottlieb DJ, Redline S, Nieto FJ, Baldwin CM, Newman AB, Resnick HE, et al. Association of usual sleep duration with hypertension: the Sleep Heart Health Study. Sleep. 2006;29(8):1009-14. PubMed PMID: 16944668.
9. Badran M, Yassin BA, Fox N, Laher I, Ayas N. Epidemiology of Sleep Disturbances and Cardiovascular Consequences. Can J Cardiol. 2015;31(7):873-9. doi: 10.1016/j.cjca.2015.03.011. PubMed PMID: 26037823.
10. Cappuccio FP, D'Elia L, Strazzullo P, Miller MA. Sleep duration and all-cause mortality: a systematic review and meta-analysis of prospective studies. Sleep. 2010;33(5):585-92. PubMed PMID: 20469800; PubMed Central PMCID: PMCPMC2864873.
11. Basner M, McGuire S. WHO Environmental Noise Guidelines for the European Region: A Systematic Review on Environmental Noise and Effects on Sleep. Int J Env Res Pub He. 2018;15(3). doi: 10.3390/ijerph15030519.
12. Fidell S, Pearsons K, Tabachnick BG, Howe R. Effects on sleep disturbance of changes in aircraft noise near three airports. Journal of the Acoustical Society of America. 2000;107(5):2535-47. doi: Doi 10.1121/1.428641. PubMed PMID: WOS:000086967900025.
13. Basner M. Design for a US field study on the effects of aircraft noise on sleep. Cambridge, MA: Partnership for Air Transportation Noise and Emissions Reduction (PARTNER), 2012 PARTNER-COE-2012-003 Contract No.: PARTNER-COE-2012-003.
14. Aasvang GM, Overland B, Ursin R, Moum T. A field study of effects of road traffic and railway noise on polysomnographic sleep parameters. J Acoust Soc Am. 2011;129(6):3716-26. Epub 2011/06/21. doi: 10.1121/1.3583547. PubMed PMID: 21682396.
15. Basner M, Samel A, Isermann U. Aircraft noise effects on sleep: application of the results of a large polysomnographic field study. J Acoust Soc Am. 2006;119(5 Pt 1):2772-84. Epub 2006/05/20. PubMed PMID: 16708935.
16. Elmenhorst EM, Pennig S, Rolny V, Quehl J, Müller U, Maass H, et al. Examining nocturnal railway noise and aircraft noise in the field: sleep, psychomotor performance, and annoyance. Sci Total Environ. 2012;424:48-56. Epub 2012/03/27. doi: S0048-9697(12)00234-3 [pii]

10.1016/j.scitotenv.2012.02.024. PubMed PMID: 22444069.
17. Flindell IH, Bullmore AJ, Robertson KA, Wright NA, Turner C, Birch CL, et al. Aircraft Noise and Sleep, 1999 UK Trial Methodology Study. Southampton, UK: Institute of Sound and Vibration Research, University of Southampton, 2000.
18. Danker-Hopfe H, Anderer P, Zeitlhofer J, Boeck M, Dorn H, Gruber G, et al. Interrater reliability for sleep scoring according to the Rechtschaffen & Kales and the new AASM standard. J Sleep Res. 2009;18(1):74-84. Epub 2009/03/03. doi: 10.1111/j.1365-2869.2008.00700.x. PubMed PMID: 19250176.
19. Basner M, Griefahn B, Penzel T. Inter-rater agreement in sleep stage classification between centers with different backgrounds. Somnologie. 2008;12(1):75-84.
20. Agnew HW, Jr., Webb WB, Williams RL. The first night effect: an EEG study of sleep. Psychophysiology. 1966;2(3):263-6. Epub 1966/01/01. PubMed PMID: 5903579.
21. Basner M, Griefahn B, Muller U, Plath G, Samel A. An ECG-based algorithm for the automatic identification of autonomic activations associated with cortical arousal. Sleep. 2007;30(10):1349-61. Epub 2007/11/01. PubMed PMID: 17969469; PubMed Central PMCID: PMC2266280.
22. Basner M, Müller U, Elmenhorst EM, Kluge G, Griefahn B. Aircraft noise effects on sleep: a systematic comparison of EEG awakenings and automatically detected cardiac activations. PhysiolMeas. 2008;29(9):1089-103.
23. McGuire S, Müller U, Plath G, Basner M. Refinement and validation of an ECG based algorithm for detecting awakenings. 11th International Congress on Noise as a Public Health Problem (ICBEN); Nara, Japan2014.
24. Basner M, Brink M, Elmenhorst EM. Critical appraisal of methods for the assessment of noise effects on sleep. Noise & health. 2012;14(61):321-9. doi: 10.4103/1463-1741.104902. PubMed PMID: WOS:000313350600011.
25. Schreckenberg D, Eikmann T, Faulbaum F, Haufe C, Herr C, Klatte M, et al. NORAH – study on Noise Related Annoyance, Cognition and Health: a transportation noise effects monitoring program in Germany. ICBEN 2011; London, UK2011.

26. Müller U, Elmenhorst EM, Mendolia F, Quehl J, Basner M, McGuire S, et al. A comparison of the effects of night time air traffic noise on sleep at Cologne/Bonn and Frankfurt Airport after the night flight ban. ICBEN; 18-22 June; Zürich2017.
27. NORAH study: Publications [cited 2019 March 18th]. Available from: https://www.norah-studie.de//en/alle.html.
28. Müller U, Elmenhorst EM, Mendolia F, Quehl J, Basner M, McGuire S, et al. The NORAH-sleep study: Effects of the night flight ban at Frankfurt airport. Internoise 2016; August 21-24, ; Hamburg, Germany2016.
29. Griefahn B, Bröde P, Marks A, Basner M. Autonomic arousals related to traffic noise during sleep. Sleep. 2008;31(4):569-77. Epub 2008/05/07. PubMed PMID: 18457245; PubMed Central PMCID: PMC2279756.
30. Basner M, Isermann U, Samel A. Aircraft noise effects on sleep: Application of the results of a large polysomnographic field study. JAcoustSocAm. 2006;119(5):2772-84.
31. Brink M, Schreckenberg D, Thomann G, Basner M. Aircraft Noise Indexes for Effect Oriented Noise Assessment. Acta Acustica United with Acustica. 2010;96(6):1012-25. doi: Doi 10.3813/Aaa.918364. PubMed PMID: ISI:000285082200003.
32. Tetreault LF, Plante C, Perron S, Goudreau S, King N, Smargiassi A. Risk Assessment of Aircraft Noise on Sleep in Montreal. Can J Public Health. 2012;103(4):E293-E6. PubMed PMID: WOS:000208937500012.
33. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33(1):159-74. Epub 1977/03/01. PubMed PMID: 843571.
34. Rechtschaffen A, Kales A. A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects. Los Angeles: UCLA; 1968.
35. McGuire S, Witte M, Kallarackal A, Basner M. Pilot Study Examining the Effects of Aircraft Noise on Sleep in Communities near Philadelphia International Airport. Sleep. 2017;40:A60-A1. doi: DOI 10.1093/sleepj/zsx050.161. PubMed PMID: WOS:000433175000163.
36. International Electrotechnical Commission. IEC 61672-1 Electroacoustics - Sound level meters. 2013.
37. Federal Aviation Administration. Integrated Noise Model (INM) [cited 2019 November]. Available from: https://www.faa.gov/about/office_org/headquarters_offices/apl/research/models/inm_model/.
38. Federal Aviation Administration. Performance Data Analysis and Reporting System (PDARS) [cited 2019 November]. Available from: https://www.faa.gov/about/office_org/headquarters_offices/ato/service_units/systemops/perf_analysis/perf_tools/.
39. Czajka JL, Beyler A. Declining response rates in federal surveys: Trends and implications. Final report - Volume I. Mathematica Policy Research, 2016 June 15th. Report No.: 40146.D4C.
40. Armstrong JS, Overton TS. Estimating Nonresponse Bias in Mail Surveys. J Marketing Res. 1977;14(3):396-402. doi: Doi 10.2307/3150783. PubMed PMID: WOS:A1977DS66100020.
41. Edwards PJ, Roberts I, Clarke MJ, DiGuiseppi C, Wentz R, Kwan I, et al. Methods to increase response to postal and electronic questionnaires. Cochrane Db Syst Rev. 2009;(3). doi: ARTN MR000008

10.1002/14651858.MR000008.pub4. PubMed PMID: WOS:000268037800036.
42. Nakash RA, Hutton JL, Jorstad-Stein EC, Gates S, Lamb SE. Maximising response to postal questionnaires--a systematic review of randomised trials in health research. Bmc Med Res Methodol. 2006;6:5. Epub 2006/03/01. doi: 10.1186/1471-2288-6-5. PubMed PMID: 16504090; PubMed Central PMCID: PMCPMC1421421.
43. Knoll M, Soller L, Ben-Shoshan M, Harrington D, Fragapane J, Joseph L, et al. The use of incentives in vulnerable populations for a telephone survey: a randomized controlled trial. BMC Res Notes. 2012;5:572. Epub 2012/10/23. doi: 10.1186/1756-0500-5-572. PubMed PMID: 23083313; PubMed Central PMCID: PMCPMC3503563.
44. Chandler JJ, Paolacci G. Lie for a Dime: When Most Prescreening Responses Are Honest but Most Study Participants Are Impostors. Soc Psychol Pers Sci. 2017;8(5):500-8. doi: 10.1177/1948550617698203. PubMed PMID: WOS:000408630200003.
45. Buysse DJ, Reynolds CF, 3rd, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res. 1989;28(2):193-213. Epub 1989/05/01. doi: 0165-1781(89)90047-4 [pii]. PubMed PMID: 2748771.
46. Weinstein ND. Individual differences in reactions to noise: a longitudinal study in a college dormitory. J Appl Psychol. 1978;63(4):458-66. Epub 1978/08/01. PubMed PMID: 701213.
47. Fields JM, De Jong RG, Gjestland T, Flindell IH, Job RFS, Kurra S, et al. Standardized general-purpose noise reaction questions for community noise surveys: Research and a recommendation. Journal of Sound and Vibration. 2001;242(4):641-79. doi: DOI 10.1006/jsvi.2000.3384. PubMed PMID: WOS:000168482600006.
48. Barba C, Cavalli-Sforza T, Cutter J, Darnton-Hill I, Deurenberg P, Deurenberg-Yap M, et al. Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. Lancet. 2004;363(9403):157-63. PubMed PMID: WOS:000187939200021.

49. Basner M, Brink M, Elmenhorst EM. Critical appraisal of methods for the assessment of noise effects on sleep. Noise Health. 2012;14(61):321-9. Epub 2012/12/22. doi: 10.4103/1463-1741.104902. PubMed PMID: 23257586.
50. McGuire S, Müller U, Plath G, Basner M, editors. Refinement and validation of an ECG based algorithm for detecting awakenings. 11th International Congress on Noise as a Public Health Problem; 2014; Nara, Japan: International Commission of Biological Effects of Noise (ICBEN).
51. Rocha S, Smith MG, Witte M, Basner M. Survey Results of a Pilot Sleep Study Near Atlanta International Airport. Int J Env Res Pub He. 2019;16(22):4321. PubMed PMID: doi:10.3390/ijerph16224321.
52. Textor J, van der Zander B, Gilthorpe MS, Liskiewicz M, Ellison GTH. Robust causal inference using directed acyclic graphs: the R package 'dagitty'. Int J Epidemiol. 2016;45(6):1887-94. doi: 10.1093/ije/dyw341. PubMed PMID: WOS:000398261100024.
53. Hoddes E, Zarcone V, Smythe H, Phillips R, Dement WC. Quantification of sleepiness: a new approach. Psychophysiology. 1973;10(4):431-6. PubMed PMID: 4719486.
54. Scott P, Edwards P. Personally addressed hand-signed letters increase questionnaire response: a meta-analysis of randomised controlled trials. Bmc Health Serv Res. 2006;6. doi: Artn 111

10.1186/1472-6963-6-111. PubMed PMID: WOS:000240734000001.
55. Sahlqvist S, Song Y, Bull F, Adams E, Preston J, Ogilvie D, et al. Effect of questionnaire length, personalisation and reminder type on response rate to a complex postal survey: randomised controlled trial. Bmc Med Res Methodol. 2011;11. doi: Artn 62

10.1186/1471-2288-11-62. PubMed PMID: WOS:000291362600001.
56. Möller-Leimkühler AM, Möller H-J, Maier W, Gaebel W, Falkai P. EPA guidance on improving the image of psychiatry. European Archives of Psychiatry and Clinical Neuroscience. 2016;266(2):139-54. doi: 10.1007/s00406-016-0678-5.
57. Halpern SD, Ubel PA, Berlin JA, Asch DA. Randomized Trial of $5 Versus $10 Monetary Incentives, Envelope Size, and Candy to Increase Physician Response Rates to Mailed Questionnaires. Medical Care. 2002;40(9):834-9. PubMed PMID: 00005650-200209000-00012.
58. Guski R, Schreckenberg D, Schuemer R. WHO Environmental Noise Guidelines for the European Region: A Systematic Review on Environmental Noise and Annoyance. Int J Env Res Pub He. 2017;14(12). doi: ARTN 1539

10.3390/ijerph14121539. PubMed PMID: WOS:000423699400102.
59. Sinclair M, O'Toole J, Malawaraarachchi M, Leder K. Comparison of response rates and cost-effectiveness for a community-based survey: postal, internet and telephone modes with generic or personalised recruitment approaches. Bmc Med Res Methodol. 2012;12. doi: Artn 132

10.1186/1471-2288-12-132. PubMed PMID: WOS:000311347600001.
60. Singer E, Van Hoewyk J, Neugebauer RJ. Attitudes and behavior - The impact of privacy and confidentiality concerns on participation in the 2000 Census. Public Opin Quart. 2003;67(3):368-84. doi: Doi 10.1086/377465. PubMed PMID: WOS:000186082100003.
61. Bowling A. Mode of questionnaire administration can have serious effects on data quality. J Public Health-Uk. 2005;27(3):281-91. doi: 10.1093/pubmed/fdi031. PubMed PMID: WOS:000231762400005.
62. Groves RM, Singer E, Corning A. Leverage-saliency theory of survey participation - Description and an illustration. Public Opin Quart. 2000;64(3):299-308. doi: Doi 10.1086/317990. PubMed PMID: WOS:000166024000004.
63. Ryan C. Language Use in the United States: 2011. U.S. Department of Commerce. Economics and Statistics Administration. U.S. Census Bureau., 2013 ACS-22 Contract No.: ACS-22.
64. Brick JM, Montaquila JM, Han DF, Williams D. Improving Response Rates for Spanish Speakers in Two-Phase Mail Surveys. Public Opin Quart. 2012;76(4):721-32. doi: 10.1093/poq/nfs050. PubMed PMID: WOS:000311670600005.
65. Guo YM, Kopec JA, Cibere J, Li LC, Goldsmith CH. Population Survey Features and Response Rates: A Randomized Experiment. Am J Public Health. 2016;106(8):1422-6. doi: 10.2105/Ajph.2016.303198. PubMed PMID: WOS:000384981200034.
66. Ryan JM, Corry JR, Attewell R, Smithson MJ. A comparison of an electronic version of the SF-36 General Health Questionnaire to the standard paper version. Qual Life Res. 2002;11(1):19-26. doi: Doi 10.1023/A:1014415709997. PubMed PMID: WOS:000175137200003.
67. Dillman DA, Smyth JD, Christian LM. Internet, phone, mail, and mixed-mode surveys : the tailored design method. 4th edition. ed. Hoboken: Wiley; 2014. xvii, 509 pages p.
68. Patrick ME, Couper MP, Laetz VB, Schulenberg JE, O'Malley PM, Johnston LD, et al. A Sequential Mixed-Mode Experiment in the U.S. National Monitoring the Future Study. J Surv Stat Methodol. 2018;6(1):72-97. doi: 10.1093/jssam/smx011. PubMed PMID: WOS:000433601600004.

69. Medway RL, Fulton J. When More Gets You Less: A Meta-Analysis of the Effect of Concurrent Web Options on Mail Survey Response Rates. Public Opin Quart. 2012;76(4):733-46. doi: 10.1093/poq/nfs047. PubMed PMID: WOS:000311670600006.
70. McMaster HS, LeardMann CA, Speigle S, Dillman DA, Tea MCFS. An experimental comparison of web-push vs. paper-only survey procedures for conducting an in-depth health survey of military spouses. Bmc Med Res Methodol. 2017;17. doi: ARTN 73

10.1186/s12874-017-0337-1. PubMed PMID: WOS:000400239100001.
71. Groves RM, Peytcheva E. The impact of nonresponse rates on nonresponse bias - A meta-analysis. Public Opin Quart. 2008;72(2):167-89. doi: 10.1093/poq/nfn011. PubMed PMID: WOS:000256523900001.
72. American Association for Public Opinion Research. Response rates - An overview 2019 [cited 2019 November]. Available from: https://www.aapor.org/Education-Resources/For-Researchers/Poll-Survey-FAQ/Response-Rates-An-Overview.aspx.
73. Cook C, Heath F, Thompson RL. A meta-analysis of response rates in Web- or internet-based surveys. Educ Psychol Meas. 2000;60(6):821-36. doi: Doi 10.1177/00131640021970934. PubMed PMID: WOS:000165284000001.
74. Johnson TP, Wislar JS. Response Rates and Nonresponse Errors in Surveys. Jama-J Am Med Assoc. 2012;307(17):1805-6. doi: 10.1001/jama.2012.3532. PubMed PMID: WOS:000303386800018.
75. Halbesleben JRB, Whitman MV. Evaluating Survey Quality in Health Services Research: A Decision Framework for Assessing Nonresponse Bias. Health Serv Res. 2013;48(3):913-30. doi: 10.1111/1475-6773.12002. PubMed PMID: WOS:000318700900002.
76. United States Census Bureau. 2010 Dicennial census of population and housing 2010 [cited 2019 July]. Available from: https://www.census.gov/programs-surveys/decennial-census/decade.2010.html.
77. United States Census Bureau. American Community Survey (ACS) 2019 [cited 2019 November]. Available from: https://www.census.gov/programs-surveys/acs.
78. Saleh A, Bista K. Examining Factors Impacting Online Survey Response Rates in Educational Research: Perceptions of graduate students. J MultiDisciniplinary Evaluation. 2017;13(29):63-74.
79. Funkhouser E, Vellala K, Baltuck C, Cacciato R, Durand E, McEdward D, et al. Survey Methods to Optimize Response Rate in the National Dental Practice-Based Research Network. Eval Health Prof. 2017;40(3):332-58. doi: 10.1177/0163278715625738. PubMed PMID: WOS:000406071500003.
80. Church AH. Estimating the Effect of Incentives on Mail Survey Response Rates - a Metaanalysis. Public Opin Quart. 1993;57(1):62-79. doi: Doi 10.1086/269355. PubMed PMID: WOS:A1993KY41200004.
81. Edwards P, Roberts I, Clarke M, DiGuiseppi C, Pratap S, Wentz R, et al. Increasing response rates to postal questionnaires: systematic review. British medical journal. 2002;324(7347):1183-5. doi: DOI 10.1136/bmj.324.7347.1183. PubMed PMID: WOS:000175763200015.
82. Mercer A, Caporaso A, Cantor D, Townsend R. How Much Gets You How Much? Monetary Incentives and Response Rates in Household Surveys. Public Opin Quart. 2015;79(1):105-29. doi: 10.1093/poq/nfu059.
83. Han DF, Montaquila JM, Brick JM. An Evaluation of Incentive Experiments in a Two-Phase Address-Based Sample Mail Survey. Surv Res Methods-Ger. 2013;7(3):207-18. PubMed PMID: WOS:000328484700006.
84. Weil J, Mendoza AN, McGavin E. Recruiting older adults as participants in applied social research: Applying and evaluating approaches from clinical studies. Educ Gerontol. 2017;43(12):662-73. doi: 10.1080/03601277.2017.1386406. PubMed PMID: WOS:000418681700008.
85. Sharp EC, Pelletier LG, Levesque C. The Double-Edged Sword of Rewards for Participation in Psychology Experiments. Can J Behav Sci. 2006;38(3):269-77. doi: 10.1037/cjbs2006014. PubMed PMID: WOS:000207230700007.
86. American Psychiatric Association. DSM-5 Task Force. Diagnostic and statistical manual of mental disorders : DSM-5. 5th ed. Washington, D.C.: American Psychiatric Association; 2013. xliv, 947 p. p.
87. Fritschi L, Brown AL, Kim R, Schwela DH, Kephalopoulos S. Burden of disease from environmental noise. Fritschi L, editor. Bonn: World Health Organization; 2011.
88. Miedema HME, Vos H. Associations Between Self-Reported Sleep Disturbance and Environmental Noise Based on Reanalyses of Pooled Data From 24 Studies. Behavioral Sleep Medicine. 2007;5(1):1-20. doi: 10.1080/15402000709336723.
89. Kim SJ, Lee KW, Park JB, Min KB, Kil HG, Lee C, et al. Exposure–Response Relationship Between Aircraft Noise and Sleep Quality: A Community-based Cross-sectional Study. Osong Public Health and Research Perspectives. 2014;5(2):108-14.
90. Jackowska M, Ronaldson A, Brown J, Steptoe A. Biological and psychological correlates of self-reported and objective sleep measures. Journal of psychosomatic research. 2016;84:52-5. doi: 10.1016/j.jpsychores.2016.03.017. PubMed PMID: WOS:000375237000008.

91. Miedema HM, Oudshoorn CG. Annoyance from transportation noise: relationships with exposure metrics DNL and DENL and their confidence intervals. Environmental Health Perspectives. 2001;109(4):409-16. PubMed PMID: PMC1240282.
92. Quehl J, Müller U, Mendolia F. Short-term annoyance from nocturnal aircraft noise exposure: results of the NORAH and STRAIN sleep studies. International archives of occupational and environmental health. 2017;90(8):765-78. doi: 10.1007/s00420-017-1238-7. PubMed PMID: WOS:000412896400005.
93. Babisch W, Houthuijs D, Pershagen G, Cadum E, Katsouyanni K, Velonakis M, et al. Annoyance due to aircraft noise has increased over the years-Results of the HYENA study. Environ Int. 2009;35(8):1169-76. doi: 10.1016/j.envint.2009.07.012. PubMed PMID: WOS:000271360900010.
94. Portas CM, Krakow K, Allen P, Josephs O, Armony JL, Frith CD. Auditory processing across the sleep-wake cycle: simultaneous EEG and fMRI monitoring in humans. Neuron. 2000;28(3):991-9. Epub 2001/02/13. doi: S0896-6273(00)00169-0 [pii]. PubMed PMID: 11163282.
95. Cointin R, Sizov N, Hileman JI. U.S. Civil Aircraft Noise Annoyance Survey Design. Hamburg2016.
96. Miedema HME, Vos H. Demographic and attitudinal factors that modify annoyance from transportation noise. Journal of the Acoustical Society of America. 1999;105(6):3336-44. doi: Doi 10.1121/1.424662. PubMed PMID: ISI:000080778100034.
97. Marks A, Griefahn B. Associations between noise sensitivity and sleep, subjectively evaluated sleep quality, annoyance, and performance after exposure to nocturnal traffic noise. Noise Health. 2007;9(34):1-7. Epub 2007/09/14. PubMed PMID: 17851221.
98. Park J, Chung S, Lee J, Sung JH, Cho SW, Sim CS. Noise sensitivity, rather than noise level, predicts the non-auditory effects of noise in community samples: a population-based survey. Bmc Public Health. 2017;17. doi: ARTN 315

10.1186/s12889-017-4244-5. PubMed PMID: WOS:000400798400001.
99. Smith MG, Croy I, Ögren M, Hammar O, Lindberg E, Persson Waye K. Physiological effects of railway vibration and noise on sleep. J Acoust Soc Am. 2017;141(5):3262-9.
100. Di Nisi J, Muzet A, Ehrhart J, Libert JP. Comparison of cardiovascular responses to noise during waking and sleeping in humans. Sleep. 1990;13(2):108-20. Epub 1990/04/01. PubMed PMID: 2330471.
101. Smith MG, Ogren M, Ageborg Morsing J, Persson Waye K. Effects of ground-borne noise from railway tunnels on sleep: A polysomnographic study. Build Environ. 2019;149:288-96. doi: 10.1016/j.buildenv.2018.12.009. PubMed PMID: WOS:000457118300025.
102. Basner M, Babisch W, Davis A, Brink M, Clark C, Janssen S, et al. Auditory and non-auditory effects of noise on health. Lancet. 2014;383(9925):1325-32. doi: 10.1016/S0140-6736(13)61613-X. PubMed PMID: 24183105.
103. Basner M, Clark C, Hansell A, Hileman JI, Janssen S, Shepherd K, et al. Aviation Noise Impacts: State of the Science. Noise & health. 2017;19(87):41-50. doi: 10.4103/nah.NAH_104_16. PubMed PMID: WOS:000404119700001.
104. Babisch W. Cardiovascular effects of noise. Noise Health. 2011;13(52):201-4. doi: 10.4103/1463-1741.80148. PubMed PMID: WOS:000290490700001.
105. Watson NF, Badr MS, Belenky G, Bliwise DL, Buxton OM, Buysse D, et al. Joint Consensus Statement of the American Academy of Sleep Medicine and Sleep Research Society on the Recommended Amount of Sleep for a Healthy Adult: Methodology and Discussion. Sleep. 2015;38(8):1161-83. doi: PII sp-00367-15

10.5665/sleep.4886. PubMed PMID: WOS:000358838100006.
106. Locher B, Piquerez A, Habermacher M, Ragettli M, Roosli M, Brink M, et al. Differences between Outdoor and Indoor Sound Levels for Open, Tilted, and Closed Windows. Int J Env Res Pub He. 2018;15(1). doi: ARTN 149

10.3390/ijerph15010149. PubMed PMID: WOS:000424121200148.
107. Perneger TV. What's wrong with Bonferroni adjustments. British medical journal. 1998;316(7139):1236-8. doi: DOI 10.1136/bmj.316.7139.1236. PubMed PMID: WOS:000073224100034.
108. McDonald JH. Handbook of Biological Statistics. 3rd ed. Baltimore, Maryland: Sparky House Publishing; 2014.
109. Croy I, Smith MG, Gidlöf-Gunnarsson A, Persson Waye K. Optimal questions for sleep in epidemiological studies: Comparisons of subjective and objective measures in laboratory and field studies. Behav Sleep Med. 2017;15(6):466-82. doi: 10.1080/15402002.2016.1163700. PubMed PMID: 27159152.
110. Basner M, Müller U, Elmenhorst EM. Single and combined effects of air, road, and rail traffic noise on sleep and recuperation. Sleep. 2011;34(1):11-23. Epub 2011/01/05. PubMed PMID: 21203365; PubMed Central PMCID: PMC3001788.
111. Tassi P, Rohmer O, Schimchowitsch S, Eschenlauer A, Bonnefond A, Margiocchi F, et al. Living alongside railway tracks: Long-term effects of nocturnal noise on sleep and cardiovascular reactivity as a function of age. Environ Int. 2010;36(7):683-9. Epub 2010/06/24. doi: 10.1016/j.envint.2010.05.001

S0160-4120(10)00087-5 [pii]. PubMed PMID: 20569986.
112. Åkerstedt T, Schwarz J, Gruber G, Lindberg E, Theorell-Haglöw J. The relation between polysomnography and subjective sleep and its dependence on age - poor sleep may become good sleep. J Sleep Res. 2016;25(5):565-70. doi: 10.1111/jsr.12407. PubMed PMID: 27122391.
113. Vitiello MV, Larsen LH, Moe KE. Age-related sleep change: Gender and estrogen effects on the subjective-objective sleep quality relationships of healthy, noncomplaining older men and women. Journal of psychosomatic research. 2004;56(5):503-10. doi: 10.1016/S0022-3999(04)00023-6. PubMed PMID: 15172206.
114. Mong JA, Cusmano DM. Sex differences in sleep: impact of biological sex and sex steroids. Philos Trans R Soc Lond B Biol Sci. 2016;371(1688):20150110. doi: 10.1098/rstb.2015.0110. PubMed PMID: 26833831; PubMed Central PMCID: PMCPMC4785896.
115. Mallampalli MP, Carter CL. Exploring sex and gender differences in sleep health: a Society for Women's Health Research Report. J Womens Health (Larchmt). 2014;23(7):553-62. doi: 10.1089/jwh.2014.4816. PubMed PMID: 24956068; PubMed Central PMCID: PMCPMC4089020.
116. Ohayon MM, Carskadon MA, Guilleminault C, Vitiello MV. Meta-analysis of quantitative sleep parameters from childhood to old age in healthy individuals: developing normative sleep values across the human lifespan. Sleep. 2004;27(7):1255-73. PubMed PMID: 15586779.
117. Mishra AK, van Ruitenbeek AM, Loomans MGLC, Kort HSM. Window/door opening-mediated bedroom ventilation and its impact on sleep quality of healthy, young adults. Indoor Air. 2018;28(2):339-51. doi: 10.1111/ina.12435. PubMed PMID: WOS:000425083500013.
118. Miedema HM, Vos H. Associations between self-reported sleep disturbance and environmental noise based on reanalyses of pooled data from 24 studies. Behav Sleep Med. 2007;5(1):1-20. Epub 2007/02/23. doi: 10.1207/s15402010bsm0501_1. PubMed PMID: 17313321.
119. Brink M, Basner M, Schierz C, Spreng M, Scheuch K, Bauer G, et al. Determining physiological reaction probabilities to noise events during sleep. Somnologie. 2009;13(4):236-43.
120. Mayfield A, Amaya A, Carris K. A matter of time: The value and optimal timing of follow-up questionnaire mailings in a multimode survey 68th Annual Conference of the American Association for Public Opinion Research May 16-19; Boston, MA. Proceedings of the Survey Research Methods Section: American Statistical Association; 2013.

VII. Appendices

Appendix 1. Field study equipment

Table 24 List of all field study equipment and associated quantities and costs

Equipment category	Item	Quantity per box	Cost per item(s) ($)
Sound	Zoom H5 Handy Recorder	1	216.39
Sound	Earthworks M23 Measurement Microphone	1	376.92
Sound	SM Series XLR Microphone Cable	1	2.99
Sound	Remote Control for Zoom H5 Handy Recorder	1	18.71
Sound	Rechargeable AA NiMH Batteries	2	4.89
Sound	Foam Windscreens for 3/8" Diameter Microphones	1	2.23
Sound	Multi-Function Ball Head with Removable Bottom Shoe Mount	1	19.99
Sound	Hot Shoe Post Adapter	1	5.21
Sound	4" Cold Shoe Extension	1	14.21
Sound	Transcend 32 GB microSDHC	1	16.99
Sound	USB 2.0 Digital Camera Cable	1	2.44
Sound	USB Wall Plug	1	8.70
Sound	Reversible Thread Adapter (Steel)	1	3.71
Physiology	Faros 90 Sensor Kit (includes eMotion Faros 90 sensor, cable set, eMotion LAB software, docking station)	1	527.00
Physiology	VELCRO(R) Brand Dots	9	2.60
Physiology	Slim Micro USB Charger Cable	1	3.23
Physiology	Ambu BlueSensor VLC Electrodes	16	8.08
Shipping	Pick and Pack Foam Sheet	1	5.75
Shipping	Convoluted Foam Set	1	2.39
Shipping	Soft Foam Charcoal Sheet 2" Thickness	1	2.87
Shipping	Soft Foam Charcoal 1" Thickness	1	1.81
Shipping	Corrugated Shipment Box	1	0.98
Shipping	Gusseted Polyester Bag	1	0.46
Shipping	Packing Tape Sheets	5	3.18
Medical	Alcohol Prep Pads Wipes	4	0.08
Medical	Durapore Medical Tape	1	0.55
Medical	Hydrocortisone 1% Anti-Itch Cream 1 Oz Tube	1	2.39
Miscellaneous	Ziploc(R) 1 Quart Storage Bags	5	0.36
Miscellaneous	Office Depot(R) Brand File Folder	1	0.46
Miscellaneous	Brother(R) Black-On-White Tape Labels	7	5.69
		Total	**1261.24**

Appendix 2. Postal questionnaires

A. Short questionnaire

Q1. During the **past month**, how would you rate your sleep quality overall?

Very Good	Fairly Good	Fairly Bad	Very Bad
☐	☐	☐	☐

Q2. In general, would you say your health is…?

Poor	Fair	Good	Very Good	Excellent
☐	☐	☐	☐	☐

Q3. Rate how strongly you agree or disagree to the statement: I am sensitive to noise

Strongly Disagree 1	2	3	4	5	Strongly Agree 6
☐	☐	☐	☐	☐	☐

Q4. Thinking about the **last 12 months** or so, when you were here at home, how much was your sleep disturbed by noise from the following sources?

	Not at all	Slightly	Moderately	Very	Extremely
Q4a. Road Traffic	☐	☐	☐	☐	☐
Q4b. Trains	☐	☐	☐	☐	☐
Q4c. Aircraft	☐	☐	☐	☐	☐
Q4d. Neighbors	☐	☐	☐	☐	☐

Q6. What race do you consider yourself to be? (mark all that apply)

- ☐ American Indian or Alaska Native
- ☐ Asian
- ☐ Black or African American
- ☐ Prefer Not to Answer
- ☐ Native Hawaiian or Other Pacific Islander
- ☐ White
- ☐ Other (please specify):_________________________

Q7. Gender: ☐ Male ☐ Female

Q8. Age: _________________________(years)

B. Medium questionnaire

Q1. During the **past month**, how would you rate your sleep quality overall?

Very Good	Fairly Good	Fairly Bad	Very Bad
☐	☐	☐	☐

Q2. How often have you have taken medicine (prescribed or "over the counter") to help you sleep in the **past month**?

Not during the past month	Less than once a week	Once or twice a week	Three or more times a week
☐	☐	☐	☐

Q3. How often in the **past month** have you done the following because of noise when trying to sleep at home?

	Never 1	Rarely 2	Sometimes 3	Often 4	Always 5
Q3a. Wear earplugs or headphones	☐	☐	☐	☐	☐
Q3b. Turn on the TV	☐	☐	☐	☐	☐
Q3c. Turn on music	☐	☐	☐	☐	☐
Q3d. Close windows	☐	☐	☐	☐	☐
Q3e. Use a sound machine	☐	☐	☐	☐	☐
Q3f. Turn on a fan	☐	☐	☐	☐	☐

Q4. Rate how strongly you agree or disagree to the statement: I am sensitive to noise

Strongly Disagree 1	2	3	4	5	Strongly Agree 6
☐	☐	☐	☐	☐	☐

Q5. Thinking about the **last 12 months** or so, when you were here at home, how much was your sleep disturbed by noise from the following sources?

	Not at all	Slightly	Moderately	Very	Extremely
Q5a. Road Traffic	☐	☐	☐	☐	☐
Q5b. Trains	☐	☐	☐	☐	☐
Q5c. Aircraft	☐	☐	☐	☐	☐
Q5d. Neighbors	☐	☐	☐	☐	☐

Q6. In general, would you say your health is…?

Poor	Fair	Good	Very Good	Excellent
☐	☐	☐	☐	☐

Q7. Have you ever been diagnosed by a health professional with any of the following sleep disorders?

☐ Sleep Apnea ☐ Narcolepsy ☐ Restless Leg Syndrome
☐ Periodic Limb Movement Syndrome ☐ Insomnia ☐ None
☐ Other (please specify): ______________________________

Q8. Do you have any problems or difficulties with your sense of hearing? ☐ Yes ☐ No

Q9. Have you ever been diagnosed by a health professional with the following conditions (mark all that apply)?

☐ Hypertension/ High blood pressure ☐ Arrhythmia/ Irregular heartbeat ☐ None

Q10. If currently employed, does your job require overnight shift work?

(Overnight shift work refers to work for at least 4 hours between 12 a.m. midnight to 6 a.m. in the morning) ☐ Yes ☐ No

Q11. How long have you lived at your current residence

Less than 1 year	1-5 years	5-10 years	More than 10 years
☐	☐	☐	☐

Q12. Are there children in this household under the age of 5? ☐ Yes ☐ No

Q13. Are you Hispanic or Latino? ☐ Yes ☐ No ☐ Prefer Not to Answer

Q14 What race do you consider yourself to be? (mark all that apply)

☐ American Indian or Alaska Native ☐ Native Hawaiian or Other Pacific Islander
☐ Asian ☐ White
☐ Black or African American ☐ Other (please specify):______________________
☐ Prefer Not to Answer

Q15. Gender:	☐ Male ☐ Female	**Q16.** Age:	______________________(years)

Q17. What is your height?	__________feet__________inches
Q18. What is your weight?	__________________________lbs

C. Long questionnaire

Q1a. During the **past month**, at what time have you usually gone to bed on weekdays or workdays?	____________________(HH:MM AM/PM)
Q1b. During the **past month**, at what time have you usually woken up on weekdays or workdays?	____________________(HH:MM AM/PM)
Q1c. During the **past month**, how much sleep did you usually get on weekdays or workdays?	____________________(Hours)

Q2. During the **past month**, how would you rate your sleep quality overall?

Very Good	Fairly Good	Fairly Bad	Very Bad
☐	☐	☐	☐

Q3. For the following questions, select the response that best reflects how often the following occurred during the **past month.**

	Not during the past month	Less than once a week	Once or twice a week	Three or more times a week
Q3a. You had trouble sleeping because you cannot get to sleep within **30 minutes**?	☐	☐	☐	☐
Q3b. You had trouble sleeping because you wake up in the middle of the night or early morning?	☐	☐	☐	☐
Q3c. You have taken medicine (prescribed or "over the counter") to help you sleep?	☐	☐	☐	☐
Q3d. You had trouble staying awake while driving, eating meals, or engaging in social activity?	☐	☐	☐	☐

Q4. How often in the **past month** have you done the following because of noise when trying to sleep at home?

	Never 1	Rarely 2	Sometimes 3	Often 4	Always 5
Q4a. Wear earplugs or headphones	☐	☐	☐	☐	☐
Q4b. Use alcohol	☐	☐	☐	☐	☐
Q4c. Use medication	☐	☐	☐	☐	☐
Q4d. Turn on the TV	☐	☐	☐	☐	☐
Q4e. Turn on music	☐	☐	☐	☐	☐
Q4f. Close windows	☐	☐	☐	☐	☐
Q4g. Use a sound machine	☐	☐	☐	☐	☐
Q4h. Turn on a fan	☐	☐	☐	☐	☐

Q5. For the following statements respond how strongly you agree or disagree.

	Strongly Disagree 1	2	3	4	5	Strongly Agree 6
Q5a. I am easily awakened by noise	☐	☐	☐	☐	☐	☐
Q5b. I get used to most noises without much difficulty	☐	☐	☐	☐	☐	☐
Q5c. I find it hard to relax in a place that is noisy	☐	☐	☐	☐	☐	☐
Q5d. I am good at concentrating no matter what is going on around me	☐	☐	☐	☐	☐	☐
Q5e. I get mad at people who make noise that keeps me from falling asleep or getting work done	☐	☐	☐	☐	☐	☐
Q5f. I am sensitive to noise	☐	☐	☐	☐	☐	☐

Q6. Thinking about the **last 12 months** or so, when you were here at home, how much was your sleep disturbed by noise from the following sources?

	Not at all	Slightly	Moderately	Very	Extremely
Q6a. Road Traffic	☐	☐	☐	☐	☐
Q6b. Trains	☐	☐	☐	☐	☐
Q6c. Aircraft	☐	☐	☐	☐	☐
Q6d. Industries/Factories	☐	☐	☐	☐	☐
Q6e. Construction	☐	☐	☐	☐	☐
Q6f. Neighbors	☐	☐	☐	☐	☐
Q6g. Air Conditioner	☐	☐	☐	☐	☐

Q7. Thinking about the **last 12 months** or so, when you are here at home, how much does noise from each of the following bother, disturb, or annoy you?

	Not at all	Slightly	Moderately	Very	Extremely
Q7a. Road Traffic	☐	☐	☐	☐	☐
Q7b. Trains	☐	☐	☐	☐	☐
Q7c. Aircraft	☐	☐	☐	☐	☐
Q7d. Industries/Factories	☐	☐	☐	☐	☐
Q7e. Construction	☐	☐	☐	☐	☐
Q7f. Neighbors	☐	☐	☐	☐	☐
Q7g. Air Conditioner	☐	☐	☐	☐	☐

Q8. In general, would you say your health is…?

Poor	Fair	Good	Very Good	Excellent
☐	☐	☐	☐	☐

Q8. In general, would you say your health is…?

Poor	Fair	Good	Very Good	Excellent
☐	☐	☐	☐	☐

Q9. Have you ever been diagnosed by a health professional with any of the following sleep disorders?

☐ Sleep Apnea ☐ Narcolepsy ☐ Restless Leg Syndrome
☐ Periodic Limb Movement Syndrome ☐ Insomnia ☐ None
☐ Other (please specify): ______________________________

Q10. Do you have any problems or difficulties with your sense of hearing? ☐ Yes ☐ No

Q11a. Have you ever been diagnosed by a health professional with the following conditions (mark all that apply)?	**Q11b**. If you have been diagnosed with a condition listed under Q11a, have you been treated for the condition in the **past month?**	
☐ Hypertension/High blood pressure	☐ Yes	☐ No
☐ Chronic headaches/Migraines	☐ Yes	☐ No
☐ Arrhythmia/Irregular heartbeat	☐ Yes	☐ No
☐ Heart disease	☐ Yes	☐ No
☐ Stomach ulcer	☐ Yes	☐ No
☐ Diabetes	☐ Yes	☐ No
☐ None of the above		

Q12. What is your marital status?

☐ Single
☐ Married
☐ Widowed
☐ Separated
☐ Divorced
☐ Domestic Partners

Q13. What was your total household income last year?

☐ < $25,000
☐ $25,000-$50,000
☐ $50,000-$75,000
☐ $75,000-$100,000
☐ $100,000-$150,000
☐ >$150,000
☐ Prefer Not to Answer

Q14. What is the highest level of education you have completed?

☐ Less than High School
☐ High School Graduate
☐ College Graduate or Higher

Q15. What is your current employment status?

☐ Working
☐ Unemployed
☐ Student
☐ Retired
☐ Homemaker

Q16. If currently employed, does your job require overnight shift work?

(Overnight shift work refers to work for at least 4 hours between 12 am midnight to 6 am in the morning)

☐ Yes ☐ No

Q17. Are you Hispanic or Latino? ☐ Yes ☐ No ☐ Prefer Not to Answer

Q18. What race do you consider yourself to be? (mark all that apply)

☐ American Indian or Alaska Native
☐ Native Hawaiian or Other Pacific Islander
☐ Asian
☐ White
☐ Black or African American
☐ Other (please specify):________________________
☐ Prefer Not to Answer

Q20. How long have you lived at your current residence?

☐ Less than 1 year
☐ 1-5 years
☐ 5-10 years
☐ More than 10 years

Q21. Do you have an air conditioner in your bedroom?

☐ No Unit
☐ Central Air Conditioner
☐ Window Unit

Q22. How many people (including yourself) reside in this household? ____________________

Q23. How many children in this household are under the age of 5? ____________________

Q24. Gender: ☐ Male ☐ Female

Q25. Age: _________________________(years)

Q26. What is your height? __________feet__________inches

Q27. What is your weight? ___________________________lbs

Appendix 3. Initial contact letters

Text highlighted in yellow indicates text that was changed based on recipient and mailing round.

Community Noise Study

Sponsored by the Federal Aviation Administration

Forename Surname
or Current Resident
Street
City, GA Zip code

Dear Forename Lastname or Current Resident,
Your household has been selected to take part in an important study on the effect of noise in your community on sleep which is sponsored by the Federal Aviation Administration. We encourage 1 adult in the household to complete the attached brief survey. The information you provide will be used to develop and revise nighttime noise policies.

Your participation is voluntary. However, your participation is essential to inform us about your neighborhood. Your answers will be treated as confidential. We have enclosed $2.00 as a token of our appreciation for your participation.

In addition to the survey, we are conducting a 5 night in home study which includes measurements of heart rate and body movement and the indoor noise levels in the bedroom at night. Participants of this additional study will receive $20/30/40.00 per night, for a total of $100/150/200.00. For information on how to participate in this optional study please refer to the last page of the attached survey booklet.

If you have any questions about this study:
Call: 215-573-3815
Email: noise@mail.med.upenn.edu
Visit: https://www.med.upenn.edu/uep/projects_pcns.html

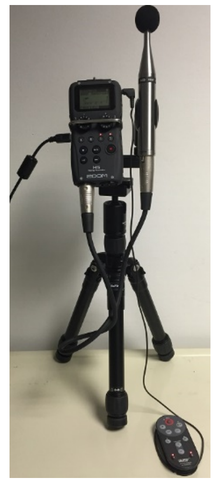
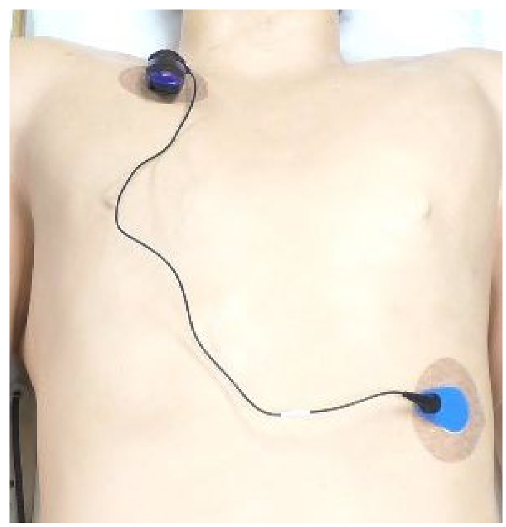
Thank you in advance for your participation!

Sincerely,

Basner

Mathias Basner, MD, PhD
Associate Professor, University of Pennsylvania

Community Noise and Sleep Study

Indoor Noise Measurements Indoor sound recordings will be made during the sleep period. The microphone and sound recorder should be placed near the sleeping position on a dresser or nightstand. Participants will need to start/stop the sound recorder each night/morning.	
Heart Rate and Body Movement Measurements During the night both heart rate and movement will be measured. The device used is battery operated. There are two electrodes for measuring heartrate there are two electrodes. One electrode will go just below the right clavicle; the other electrode will go on the left side of the chest below the pectoral muscle/breast. There is a button on the device for starting and stopping the measurements each night/morning.	

Appendix 4. Morning questionnaire

Morning Questionnaire

Instructions

- Please mark all answers clearly
- If the question is multiple choice, mark your answer by placing an x in the box: ☒
- If there are no response alternatives listed, write in your response in the provided space

1. **Current Date**: ___________ **Current Time**:______________

2. **Last night did you sleep with the windows...**
   - ☐ Closed
   - ☐ Partially Open
   - ☐ Completely Open

3. **At what time did you...**

   go to bed and switch off the light last night? ______________ (Hour: Minute)

   wake up this morning? ______________ (Hour: Minute)

   get out of bed this morning? ______________ (Hour: Minute)

4. **How long did it take you to fall asleep after you turned the lights off?**

   _______________(minutes)

5. **Did you wake up during the night?**
   - ☐ Yes
   - ☐ No

   If so, how many times? ________________

   What were the reasons, please describe:__________________________________

   ____

6. How do you feel right now?

	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	
awake, active, refreshed	0	1	2	3	4	5	6	7	8	9	10	tired, dull, sleepy

7. Please check the box next to the statement that best describes how sleepy you feel right now...

- ☐ Feeling active, vital, alert, or wide awake
- ☐ Functioning at high levels, but not at peak; able to concentrate
- ☐ Awake, but relaxed; responsive but not fully alert
- ☐ Somewhat foggy, let down
- ☐ Foggy; losing interest in remaining awake; slowed down
- ☐ Sleepy, woozy, fighting sleep; prefer to lie down
- ☐ No longer fighting sleep, sleep onset soon; having dream-like thoughts

8. Please evaluate last night's sleep:

Falling asleep was:

☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
0: very easy	1	2	3	4	5	6	7	8	9	10: very difficult

My sleep was:

☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
0: very calm	1	2	3	4	5	6	7	8	9	10: very restless

Overall Sleep Quality:

☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
0: low	1	2	3	4	5	6	7	8	9	10: high

9a. How bothered, disturbed, or annoyed do you feel by last night's Aircraft noise?

☐ Not at all ☐ Slightly ☐ Moderately ☐ Very ☐ Extremely

10. How bothered, disturbed, or annoyed do you feel by last night's Road Traffic noise?

☐ Not at all ☐ Slightly ☐ Moderately ☐ Very ☐ Extremely

11. How bothered, disturbed, or annoyed do you feel by last night's Train noise?

☐ Not at all ☐ Slightly ☐ Moderately ☐ Very ☐ Extremely

12. How bothered, disturbed, or annoyed do you feel by noise in general last night?

☐ Not at all ☐ Slightly ☐ Moderately ☐ Very ☐ Extremely

13. Other comments?

....

....

....

Appendix 5. Non-participation figures

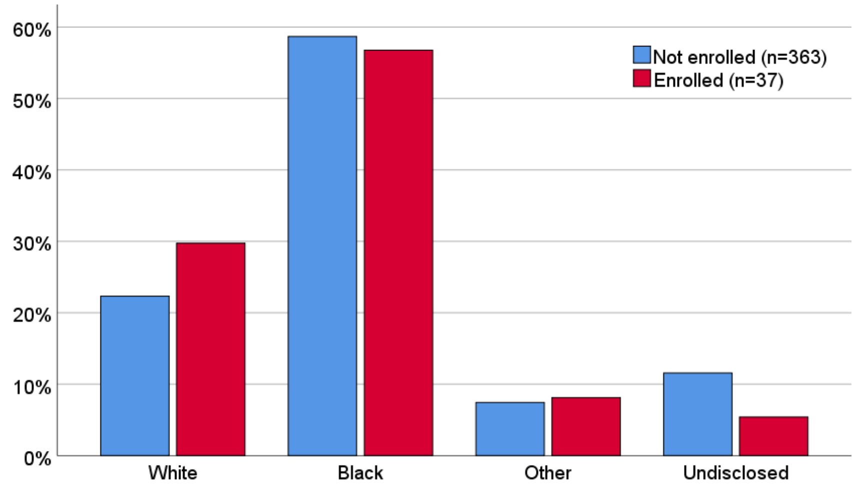


Figure 30 Race. Respondents indicating multiple ethnicities are classified as "Other". p=0.557.

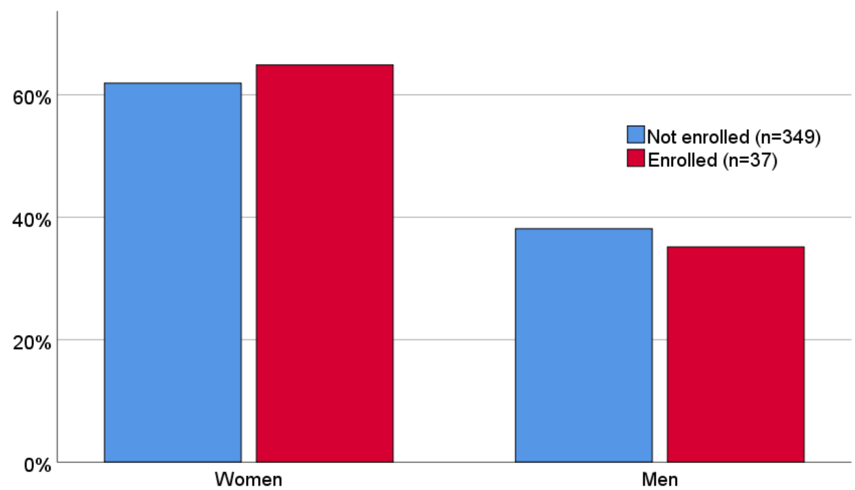


Figure 31 Sex. p=0.859.

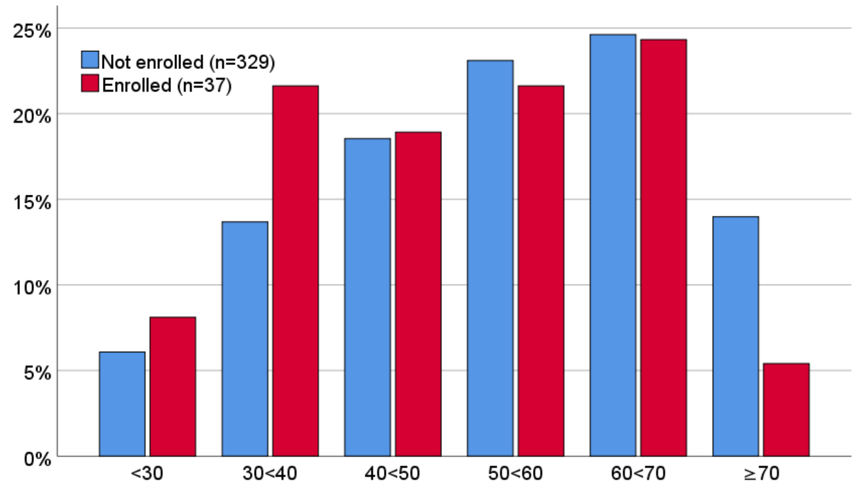


Figure 32 Categorical age. Excludes one non-participant respondent listing an age of 4 years. p=0.580

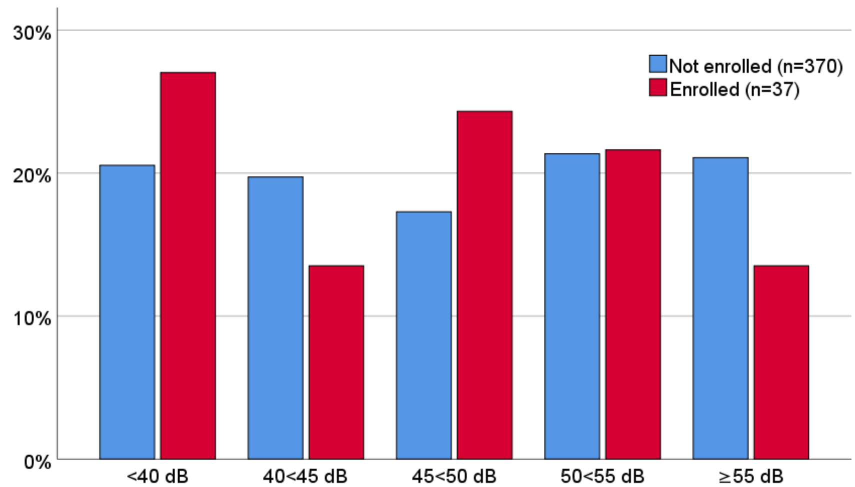


Figure 33 Categorical L_{Night}. p=0.527.

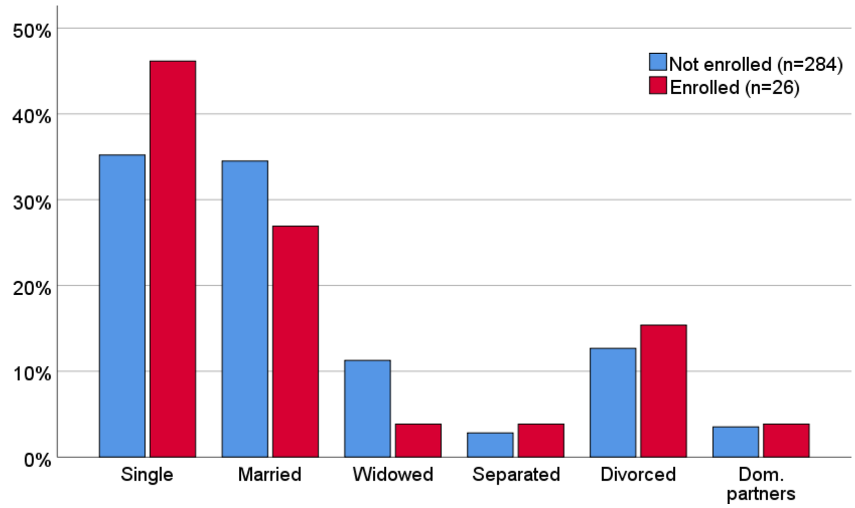


Figure 34 Marital status. p=0.649.

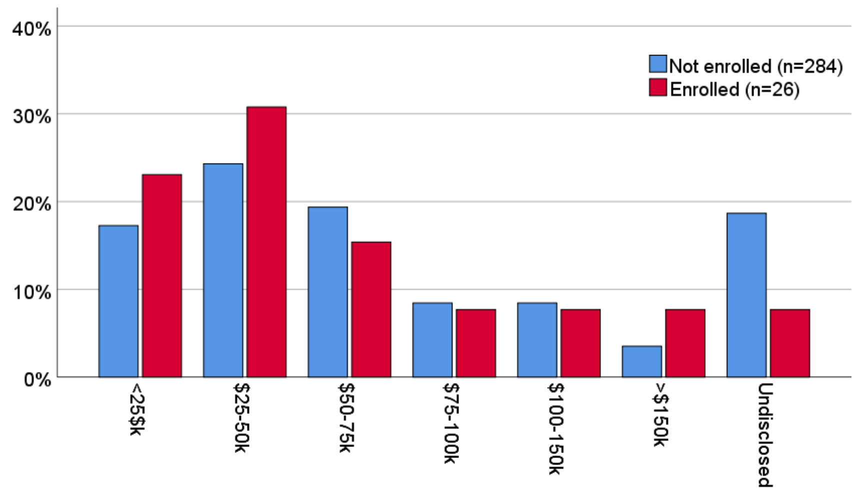


Figure 35 Annual household income. p=0.634.

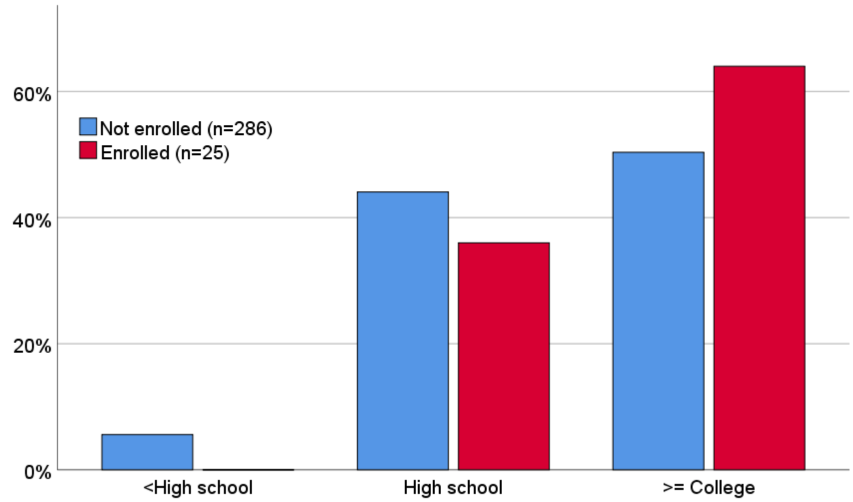


Figure 36 Highest education level completed. p=0.374.

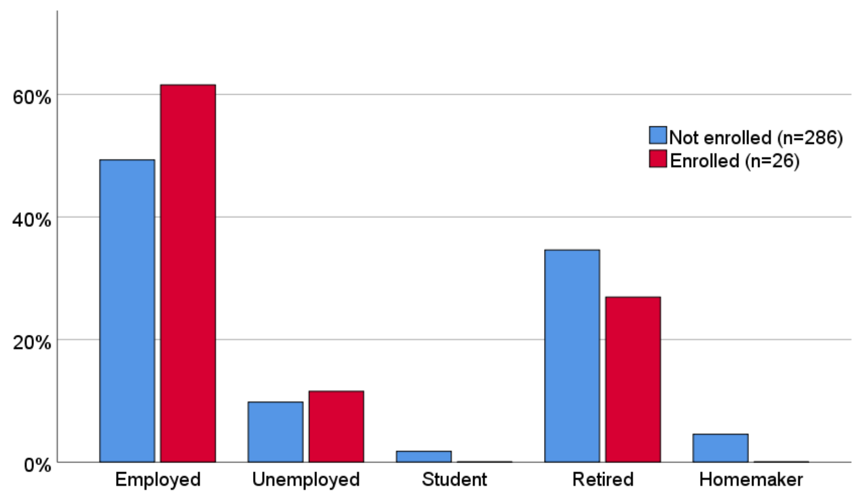


Figure 37 Employment status. p=0.733.

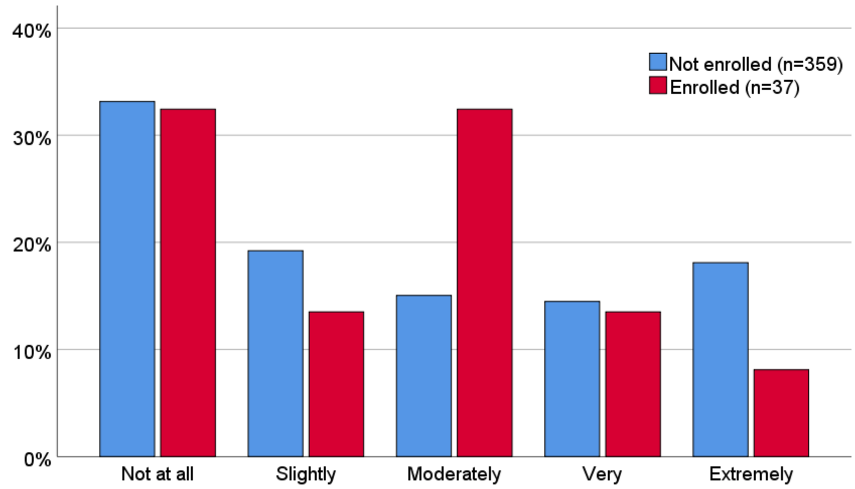


Figure 38 Sleep disturbance by aircraft noise over past 12 months. p=0.100.

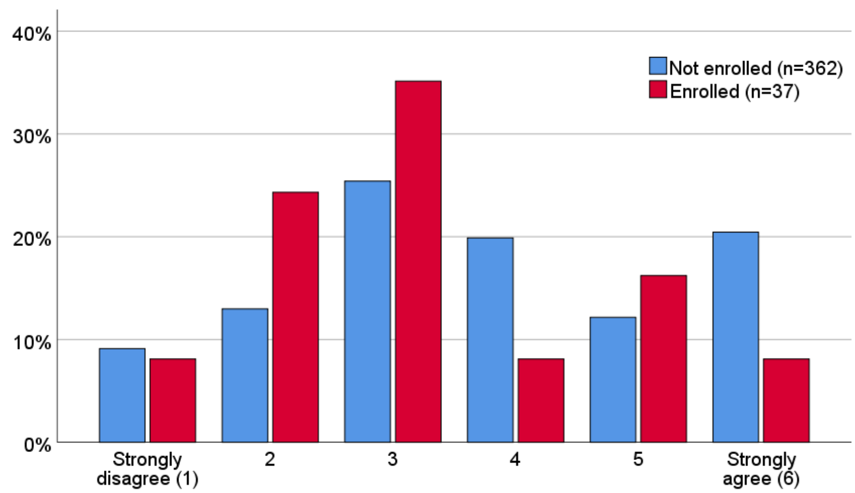


Figure 39 I am sensitive to noise. p=0.065.

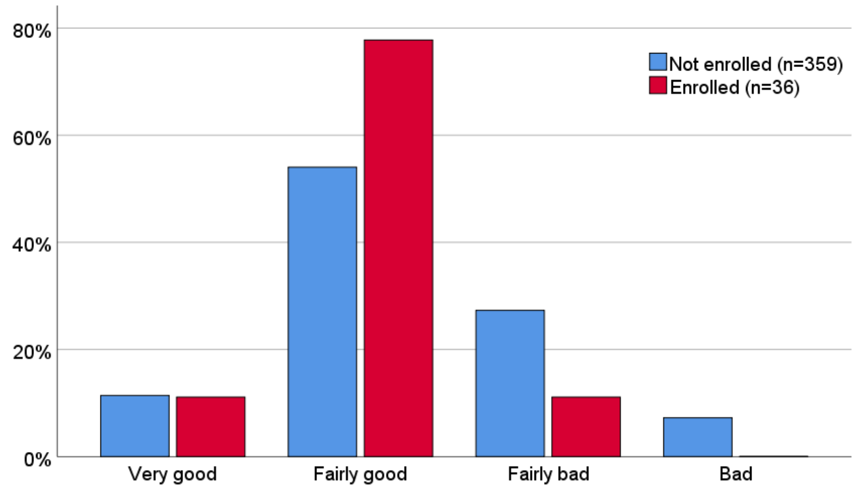


Figure 40 Overall sleep quality during past month. p=0.023.

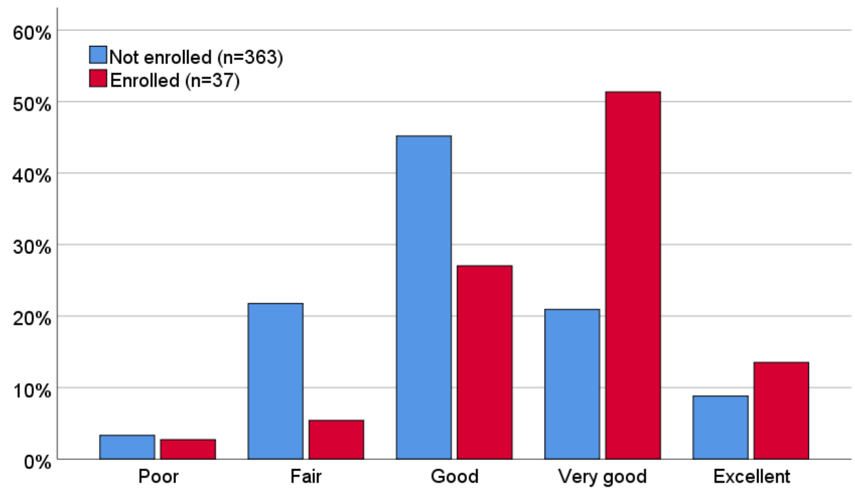


Figure 41 Self-rated general health. p=0.0004.