

183rd Meeting of the Acoustical Society of America

Nashville, Tennessee
5-9 December 2022

Noise: Paper 2pNS3

Simulating effects of noise dose error and sample size on dose-response models for sonic booms

William Jeffrey Doebler and Aaron B. Vaughn

*Structural Acoustics Branch, NASA Langley Research Center, Hampton, VA, 23517, USA;
william.j.doebler@nasa.gov; aaron.b.vaughn@nasa.gov*

Error in noise dose estimates can degrade results of community noise surveys. Accurate dose-response models are essential for informing regulators in their noise policy decisions. Increasing the number of survey participants, i.e., sample size, can improve precision but does not alleviate the effect of dose error. During the Quesst mission, NASA will survey community annoyance response as the X-59 aircraft flies supersonically over communities in the USA. To determine the practical impact of dose error on community noise survey results, this paper simulates dose-response data with varying levels of error and sample size. Improvements to precision are found to be minimal for a sample size greater than 2000 participants. In the absence of dose error, a sampled population of 2000 results in a precision of ± 2 dB in the dose that elicits 3% high annoyance. Additionally, this study determines that when the dose error exceeds 2 dB, the estimated dose-response curve deviates outside the confidence interval of the true dose-response model.

1. INTRODUCTION

As part of its Quesst mission,¹ NASA will administer community noise surveys using X-59 low-noise supersonic aircraft as a noise source. Survey participants will respond to noise from supersonic flyovers of the X-59 up to six times per day for about one month. Noise monitors sparsely placed throughout the communities will measure the noise levels of flyovers, and these measurements in combination with physics-based predictions will estimate the noise level at participant locations.² One goal of the Quesst mission is to provide regulators with noise-annoyance data from community studies to inform future supersonic aircraft noise certification standards. The community noise survey data can be summarized with dose-response models, where noise level is the dose and percent highly annoyed is the response.

Misestimating the dose, i.e., dose error, is known to cause attenuation bias in dose-response models, which manifests as a flattening of the summary curve slope.³⁻⁷ In community noise surveys, dose is commonly estimated because it can be impractical to directly measure the actual noise dose each participant experienced, i.e., the true dose. Estimated doses are prone to error from measurement and prediction uncertainties. The attenuation bias from dose error is not alleviated by increasing the number of survey participants, i.e., the sample size. Increased sample size improves precision but does not improve accuracy. On the other hand, having too few participants can cause high sampling variability, where the dose-response curve of the sampled participants is non-representative of the overall population. Results of one small sample of participants can significantly differ from another sample.

This paper explores simulated community noise survey data to inform Quesst mission planning by analyzing the practical impacts of dose error and sample size on dose-response curves. The work focuses on single-event dose-response data and provides an estimate of the dose error magnitude that causes attenuation of practical significance. The number of participants to obtain a reasonably precise dose-response curve with minimal sampling variability is also identified.

First, descriptions are provided of methods to generate dose-response data and fit a dose-response model, followed by assessments of dose error and sample size. Results include analyses of the effects of dose error and sample size on accuracy, precision, and sampling variability. Simulation limitations are given, followed by concluding remarks. The Appendices present methods to account for dose error in dose-response modeling and discuss results from additional simulations focused on cumulative daily noise doses rather than single-event doses from individual aircraft flyovers.

2. METHODS

This section describes dose-response data simulation and dose-response modeling methods. Simulated data enable exact knowledge of modeling parameters that describe the underlying population, which is otherwise unattainable. Simulated participants described in Sec. 2.A are subjected to known, true doses described in Sec. 2.B. They respond probabilistically as highly annoyed or not to the true dose. The true doses are perturbed to simulate estimated doses with error (also described in Sec. 2.B), and a multilevel model is fit to each dataset (Sec. 2.C). A flow diagram (Figure 2) describing the data generation and modeling process is provided at the end of Sec. 2.C.

A. POPULATION AND PARTICIPANT MODEL

Models describing the behavior of a community noise survey population and its individual participants are used to generate simulated dose-response datasets. To produce the realistic datasets for this study, a population model⁸ from a previous NASA sonic boom community noise survey, Quiet Supersonic Flights 2018 (QSF18),⁹ was utilized. The model assumes that participants provide binary responses (highly annoyed or not) to noise doses based on probability defined by a logistic function:

$$p_{ij} = \frac{1}{1 + e^{-(\beta_{0i} + \beta_1 \text{Dose}_{ij})}}. \quad (1)$$

Here, Dose_{ij} is the noise level of the j^{th} sonic boom that the i^{th} participant experienced. Each participant is described by an intercept of annoyance onset, β_{0i} , a random effect (RE) that is normally distributed about the mean intercept, β_0 , with standard deviation, σ_{RE} , i.e., $\beta_{0i} \sim N(\beta_0, \sigma_{RE})$. The participants share a common slope of increasing annoyance with dose, β_1 . The specific population parameter values are given in Table 1. Additional realism is simulated with survey withdrawal and (non)response rates that were observed in QSF18. These rates are described in greater detail in the next section.

Table 1. Parameters describing the QSF18-based simulation population.

Population Model Parameters		
$\beta_0 = -19.0$	$\beta_1 = 0.153$	$\sigma_{RE} = 2.451$

B. SIMULATED DATASET GENERATION

Simulated dose-response datasets were generated with varied number of participants and dose error strengths. This section describes the process for generating the datasets.

The first dataset generation step is to draw a random sample of participants from the population. For this study, the number of participants, $n_{participants}$, in each dataset was varied from 50 to 10,000. Each participant is assigned an ID number and defined by their β_{0i} , which is drawn from a normal distribution as described previously in Section 2.A. An example of individual dose-response curves is shown in Fig. 1b as a collection of solid gray lines. These individual dose-response curves are normally distributed about the average individual's dose-response curve noted by the dashed black line.

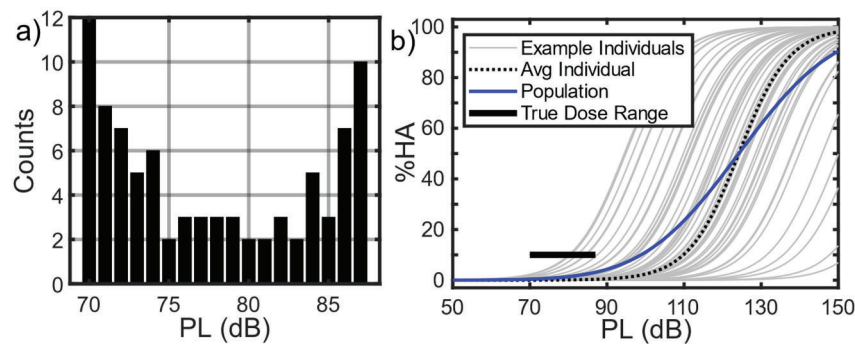


Figure 1. a) 86 Single event ($N_{booms} = 86$) dose design and b) single event true population response in percent highly annoyed (%HA) with the true single event dose range shown as the horizontal black line.

The next step is to expose the participants to known noise doses, referred to as true doses, and observe their probabilistic binary responses to each true dose. Each simulated participant was exposed to 86 true doses over the Perceived Level (PL) range of 70 to 87 dB as shown in Fig. 1a.¹⁰ This exposure design was an initial candidate design for the X-59 community tests.¹¹ The range of true doses is also illustrated as the black horizontal line in Fig. 1b. The corresponding binary response is simulated at each true dose based on the given probability of high annoyance to the dose, p_{ij} from Eq. 1. For each dose, a random draw from a uniform distribution from 0 to 1, u_{ij} , is compared to p_{ij} . If $u_{ij} < p_{ij}$, then the binary response is recorded as a 1 for highly annoyed and a 0 for not highly annoyed otherwise.

Simulated dose error is now inserted into the data. A random draw from a mean-zero normal distribution is added to each true dose resulting in a set of “estimated” doses. For this study, datasets were generated with no dose error and with dose error of standard deviation, σ_D , from 2 to 8 dB. For reference, two previous NASA sonic boom community tests estimated σ_D to be 3.7 and 4.9 dB.^{12,9}

At this point, the datasets have 100% participation, i.e., every participant provided an annoyance response to every noise exposure. To produce more realistic results, participants and responses are eliminated from the datasets to account for anticipated withdrawal and (non)response rates. A withdrawn participant is one who signed up to participate in the survey but never provided any responses. Non-withdrawn participants provide some responses but may not respond to every available survey.

Withdrawal and response rates observed in QSF18 are used in this study, with a probability of withdrawal of 0.258 and a probability of 0.469 that a non-withdrawn participant responds to any given survey. To model survey withdrawal, all responses from a random selection of participants are removed, corresponding to the withdraw rate. This dataset still has a 100% response rate, but now only contains non-withdrawn participants. It is referred to as the “realistic withdraw with 100% response rate” dataset in Fig. 2. Subsequently, a random selection of responses is removed, corresponding to 1 minus the response rate. This results in a dataset that is missing responses from some participants entirely and has a varying number of responses for each of the non-

withdrawn participants. It is referred to as the realistic participation dataset. The 100% participation, realistic withdraw with 100% rate, and realistic participation rate datasets were stored and analyzed separately, but the realistic participation datasets are the focus of this paper.

To demonstrate sampling variability, 30 replicate datasets were generated for each combination of factor and level (i.e., 7 levels of number of participants, 3 levels of participation, and 5 levels of dose error) noted in Table 2. The dose-response curves for replicate datasets should match when the sample size is sufficiently large. A total of 3,150 datasets were generated and analyzed for this study.

Table 2. Dose-response dataset factors and levels.

Factor	Levels
Number of participants	50, 500, 2500, 3000, 5000, 7500, 10000
Level of participation	0% withdrawal and 100% response rate (100% participation), 25.8% withdrawal and 100% response rate (realistic withdraw with 100% response rate), 25.8% withdrawal and 46.9% response rate (realistic participation)
Dose error standard deviation σ_D	0, 2, 4, 6, 8 dB
Replicate datasets	30

C. REGRESSION MODEL

This section describes the dose-response regression model. Using a model with the same mathematical form as the participant population model allows comparison between the known population parameters (Table 1) and those estimated by the model.

A multilevel logistic regression (MLR) model was fit to each simulated dose-response dataset. In the R programming language,¹³ the `glmer` function from the `lme4` package¹⁴ was used to fit the models as follows:

```
mod=lme4::glmer(response~dose+(1|participant_id),data=dat,family=binomial,nAGQ = 25)
```

The input data `dat` contains the participant's ID number `participant_id`, the true or estimated dose to which they responded `dose`, and their response (highly annoyed=1 or not=0) `response`. Following model accuracy and stability recommendations from Cruze et al.,¹⁵ 25 quadrature nodes `nAGQ` are used. The function returns the MLR model `mod`, which includes the β_0 , β_1 , and σ_{RE} fit parameters, standard error of β_0 and β_1 (notated as σ_{β_0} and σ_{β_1}), and the correlation between β_0 and β_1 (notated as ρ). The runtimes are reasonable to fit the model to the datasets using a modern laptop computer, ranging from less than a second for the 50 participant datasets to about 30 minutes for the 10,000 participant datasets.

The model parameters β_0 and β_1 describe the average participant and not the overall population. An additional step is required to achieve a population summary dose-response relationship, as depicted by the blue line in Fig. 1b. The integration method of Pavlou et al.¹⁶ is used and modified to develop confidence intervals and account for the error in and correlation between β_0 and β_1 .

The Pavlou method integrates over the β_{0i} random effect, and the integrand is the product of the **logistic probability** of observing high annoyance from participant with intercept β_{0i} to a given dose and the **normal probability** of observing a participant with intercept β_{0i} . This results in the simulation population's average probability of high annoyance to a dose:

$$p(\text{Dose}) = \int_{-\infty}^{\infty} \frac{1}{1 + e^{-(\beta_{0i} + \beta_1 \text{Dose})}} \frac{1}{\sigma_{RE}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\beta_{0i} - \beta_0}{\sigma_{RE}}\right)^2} d\beta_{0i} . \quad (2)$$

This integral does not have an analytical solution, so it must be solved numerically. For this work, Eq. 2 was approximated by brute-force, by averaging a large sample of **logistic** functions with **normally** distributed β_{0i} random intercepts based on σ_{RE} over a dose range of interest. The brute-force approximation converged using 50,000 normally distributed participant β_{0i} draws, resulting in the average of 50,000 corresponding logistic functions per solve of Eq. 2. For this paper, the evaluated dose range was 50 to 150 dB, which encompasses 20 dB below the true dose range and about 30 dB above the dose that elicits 50% high annoyance in the average participant. The choice is somewhat arbitrary but was selected so the full S-shape character of the curve could be visualized as in Fig. 1b.

The dose-response curve from a single solve of Eq. 2 ignores the fact that there is error in and correlation between model fit parameters. This paper extends Pavlou's method to account for these using a bootstrap method that yields a more robust population summary curve estimate and a confidence interval. The bootstrap method converged using a set of 1,000 correlated pairs of β_0 and β_1 , which are normally distributed based on their standard error values of σ_{β_0} and σ_{β_1} and whose correlation is ρ . Then, Eq. 2 is solved numerically as previously described for each of the 1,000 correlated pairs, resulting in 1,000 population summary curve estimates. The final population dose-response relationship for a given dataset replicate is the mean of the 1,000 curve estimates. These final curves are shown for each replicate in the sampling variability Sec. 3.C below. The confidence intervals for a replicate dataset are the 2.5th and 97.5th percentile of the 1,000 estimates. The mean of 30 replicates and the mean confidence interval are shown for the various combinations of participant number and dose error in Sec. 3.A.

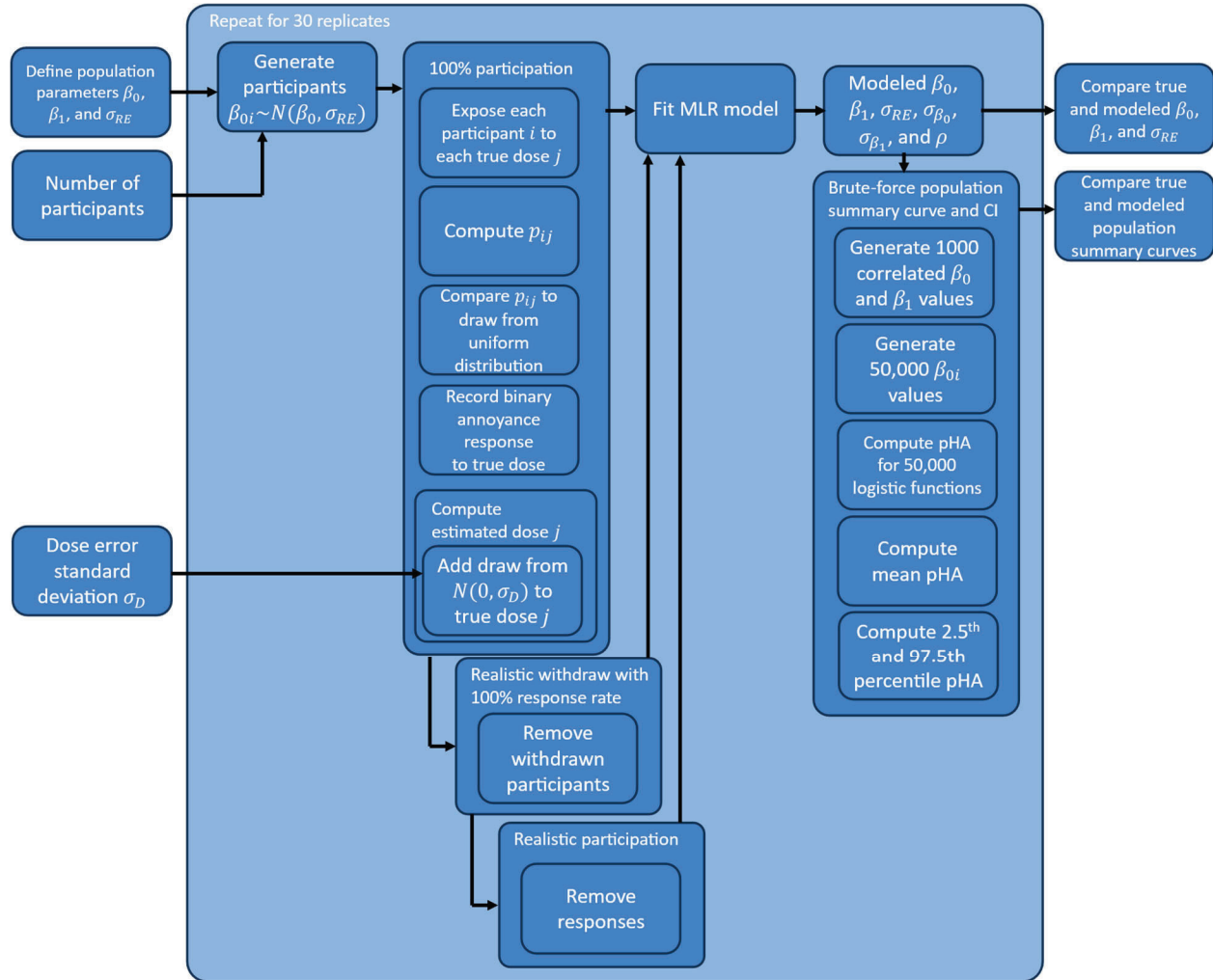


Figure 2. Flow diagram describing the simulation dataset generation and regression processes.

3. RESULTS

Results of the MLR model fit are given to determine the practical impact of dose error and sample size on dose-response curves. First, impacts on model parameter accuracy are shown, then effects on precision. Sampling variability is demonstrated, followed by a discussion of limitations of the simulations and results.

A. ACCURACY

Dose error causes misestimation of model parameters that describe the population. Box and whisker plots in Figure 3 show the impact of increasing amounts of dose error on model parameters, with β_0 in Fig. 3a, β_1 in

Fig. 3b, and σ_{RE} in Fig. 3c. Color is used to depict the dose error strength from 0 to 8 dB. Each box and whisker depicts the range and quartile values of parameter estimates for the 30 replicate datasets. The true parameter value is noted by a horizontal dashed black line. Results are shown for several different number of (non-withdrawn) participants. For these datasets, the dose error impact on model estimates has distinct linear trends,^a where $\beta_0 \propto \sigma_D$ and $\beta_1 \propto -\sigma_D$. The trend for σ_{RE} appears to be proportional to $-\sigma_D$ but is less distinct.

The impacts of sample size are also observed in Figure 3. Moving along the abscissa, increasing the number of participants reduces the box and whisker range, indicating lower sampling variability because the model estimates of the parameter are more similar between replicates. The reduction in whisker range becomes minuscule beyond a sample size of 2,000 participants, which suggests there is little benefit to sampling more than 2,000 participants. However, inaccurate parameter estimation persists in the presence of dose error regardless of the number of participants, illustrated by deviations of the parameters from the black dashed line.

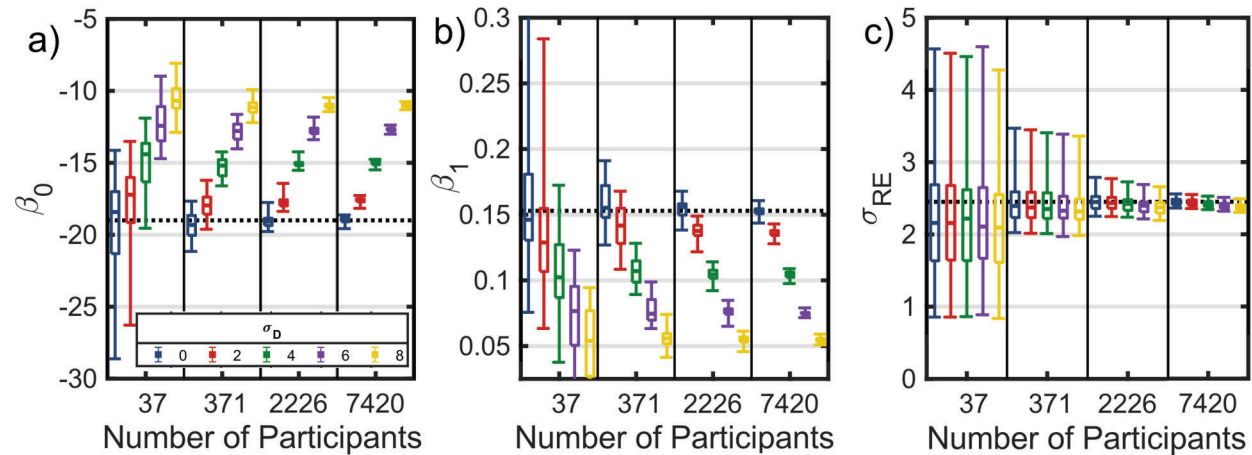


Figure 3. Model fit parameters as a function of sample size and dose error standard deviation, σ_D . Box and whiskers depict percentile values of the parameter fits from 30 replicate datasets. The dashed black line shows the true value.

The practical impact of the population parameter values on the estimated dose-response curve is not immediately intuitive.^b To complement the discussion of parameter estimation results above, Figure 4 depicts the impact of dose error on population summary curve accuracy for a sample size of 7240 participants. The blue curve shows the modeled population summary curve relative to the true population in red. When σ_D reaches or exceeds 2 dB, the modeled dose-response curve begins to differ from the true population.^c Based on these results and given the limitations of the simulation approach (described later in Sec. 3.D), models that account for dose error should be considered for use during X-59 community tests if the estimated dose error is greater than 2 dB. Potential dose error correction methods that could be used are summarized in the Appendix.

^a Proportional relationships of model parameter misestimation as a function of increasing dose error is the basis to correct for the effects of dose error via Simulation Extrapolation (SIMEX), which is discussed in the Appendix.

^b It should be mentioned that the ratio $-\beta_0/\beta_1$, which corresponds to the dose that elicits 50% high annoyance (in both the population and average participant), and the dose D that elicits $P\%$ high annoyance (in the average participant only) are related by $D = (-\beta_0 - \ln(100\%/P - 1))/\beta_1$. Note that Eq. 2 must be used to find the high annoyance of the population. This nuance is shown in Fig. 1b, where the population and average participant curves intersect and match at 50% high annoyance but otherwise differ.

^c There appears to be a “pivot point” around 80 dB where the attenuated model curve intersects the true population curve. The model curve is accurate at this dose regardless of dose error and it separates the modeled curve between over and underestimation of the true population. One might be able to analytically derive the pivot point dose value based on the functional form of the true population curve and the true dose distribution. Further understanding of this concept may provide greater confidence in the point estimate of the model curve at the pivot point dose. It may also be interesting to explore artificially increasing the dose error strength and observing the pivot point in previous experimental data that is known to contain dose error such as QSF18.

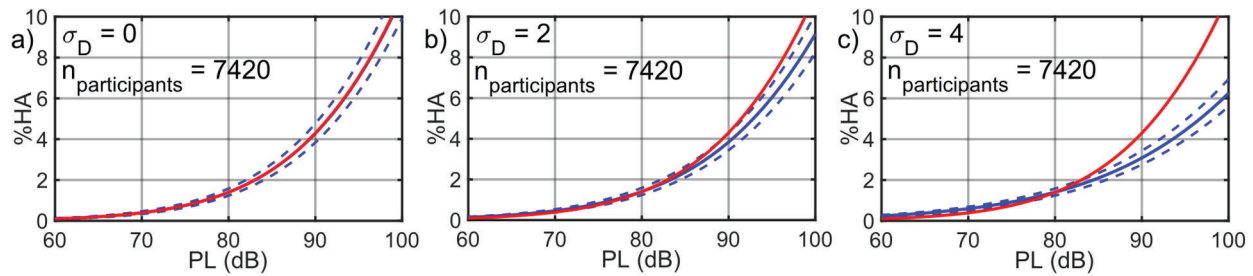


Figure 4. Model estimated (blue) dose-response curves with increasing dose error standard deviation compared to the true (red) population.

B. PRECISION

The sample size of the dataset determines the precision in the dose-response model. Having too few participants or responses could produce large confidence intervals about the dose-response curve. Results with large confidence intervals may not be useful for developing noise regulations. However, there are diminishing returns with increasing sample size, as precision is inversely proportional to the square root of sample size. For the X-59 tests, budget also factors into sample size consideration as participants will receive monetary compensation for their participation in the community noise surveys.

While the precision required for future supersonic noise regulations is unknown, trends and analysis of sample size and precision offer valuable insight. An illustration of precision as a function of sample size is given in Figure 5. The confidence intervals (blue dashed lines) decrease in width with more participants. One may reasonably assume that regulations could consider the dose that elicits X% high annoyance in the population. As an example, the dose that elicits 3% high annoyance (%HA) is examined here. For this simulated population, the population summary curve intersects 3%HA at 86.7 dB. Figure 6 shows the precision in the model estimate of this dose. Each blue dot and whisker represent the confidence interval width about the 86.7 dB point estimate from the 30 replicates. With increased number of participants, the precision plateaus, which shows the limited return on investment. With 1855 participants, the precision in the dose that elicits 3%HA is about ± 2 dB. A four-fold increase in participants to 7420 improves precision only marginally to ± 1 dB.

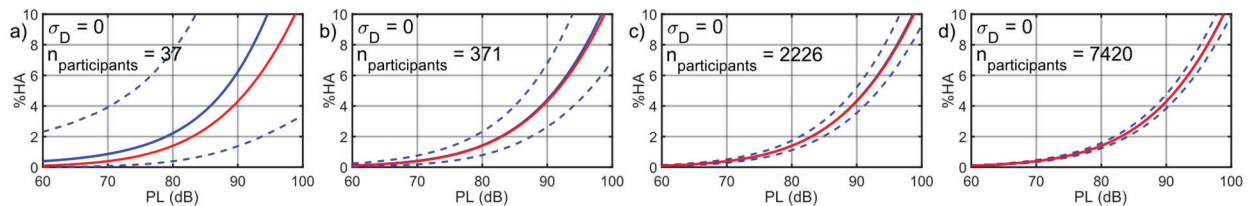


Figure 5. Precision in modeled dose-response curves as a function of sample size.

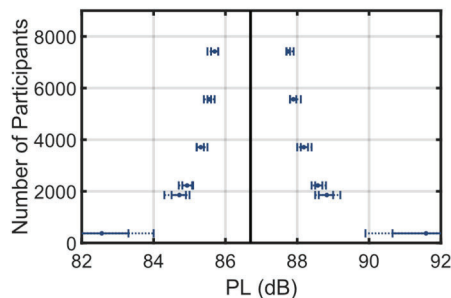


Figure 6. Confidence interval widths showing precision in the dose (86.7 dB) that elicits 3% HA as a function of sample size.

C. SAMPLING VARIABILITY

Sampling variability occurs when the sample size is small, resulting in considerable variability in dose-response models between samples. Low sampling variability is desirable for the X-59 community tests because each community will likely only be sampled once. Figure 7 illustrates sampling variability with no dose error and varying numbers of participants in Fig. 7a to 7c and 4 dB dose error and 7420 participants in Fig. 7d. Each of the blue curves corresponds to one dataset replicate. There is high sampling variability with 37 participants and low sampling variability with 7420 participants. With dose error present, Fig. 7d depicts similar sampling variability as in Fig. 7c, which has the same number of participants. The slope attenuation observed for the blue curves in the Fig. 7d also reinforces that increasing the number of participants does not alleviate the dose error impact.

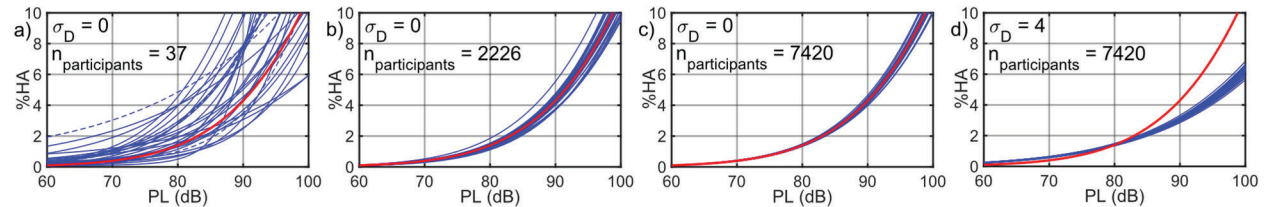


Figure 7. Demonstration of sampling variability and that the dose error impact is not alleviated by increased sample size.

D. LIMITATIONS

As with any experiment, simulated or otherwise, there are limitations that should be discussed. The representativeness of the population model is limited as it was based on the QSF18 modeling results, which is not nationally or globally representative. The population response of X-59 test communities is unknown and may differ. Furthermore, this work assumed withdrawal and participation rates that were observed in the QSF18 test, and that randomly removed responses were not correlated to other quantities such as sound level. Care should be taken when interpreting results, and ongoing work aims to reduce the dependence on the QSF18-based population.¹⁷ The dose design utilized in this study is subject to change, and additional real-world effects on dose, such as geographic and atmospheric propagation-induced variability, are not included here. Additionally, the dose error was considered to be mean-zero and normally distributed with fixed variance at all doses. Dose error is likely to vary with level, with greater variability at low levels due to poor signal-to-noise ratio of sonic booms relative to ambient conditions. Present limitations are left for future work.

4. CONCLUSIONS

This simulation study explored the effects of dose error and sample size on single-event dose-response curves to better understand their practical impact on future X-59 community test data. The simulation population was based on population model parameters from the QSF18 survey data, and noise dose values came from an initial X-59 dose design. Model parameters and resulting dose-response curves begin to differ from the true population when the standard deviation of the dose error exceeds 2 dB. In X-59 community tests, it is recommended to utilize a modeling method that accounts for dose error if the dose error is estimated to exceed 2 dB. A plateau in both precision and sampling variability occurs above about 2000 participants, which suggests the sample size of the X-59 community tests need not exceed 2000. This sample size leads to a precision of about ± 2 dB PL for the (single event) dose that elicits high annoyance in 3% of the population. While not shown explicitly for brevity, this same sample size results in about ± 4 dB PLDNL for the daily cumulative dose (Appendix B has more details on the cumulative dose simulations).

ACKNOWLEDGMENTS

Work conducted in support of the NASA Commercial Supersonic Technology Project. The authors thank those who contributed to the conception of this work, namely Nathan Cruze, Kathryn Ballard, Jonathan Rathsam, Pete Parker, Jacob Klos, Andy Christian, and Ran Cabell of NASA Langley Research Center.

APPENDIX A: SUMMARY OF DOSE ERROR MITIGATION METHODS

This study demonstrates the need to consider measurement error in the Quesst mission. Various methods for mitigating the impact of dose error on dose-response relationships are being explored¹⁸ and are described briefly in this Appendix. Also of note is that methods to account for dose error are prone to error themselves. Applying a method to correct the dose error may still result in a poorly estimated dose-response relationship due to inaccurate or insufficient dose uncertainty quantification methods.

The Simulation Extrapolation (SIMEX)^{19–21} method was alluded to in a Sec. 3.A. The SIMEX process requires an estimate of the dose error magnitude of the estimated doses, for example $\sigma_D = 2$ dB, and relies on the fact that model parameter attenuation is functionally related to σ_D , e.g., $\beta_0 = f(\sigma_D)$. First, a model is fit to the estimated dose-response data. Next, the estimated doses are further perturbed with additional, simulated error (e.g., $\sigma'_D = \sigma_D + 1$ dB), and a model is fit to the further-perturbed data whose fit parameters are further attenuated. This process is repeated many times for increasing dose perturbation σ'_D . A function can then be fit to the increasingly attenuated model parameters, and extrapolated backwards to the no error, $\sigma'_D = 0$, case.

Another method, Regression Calibration,²² has recently been explored by Horonjeff^{23–25} for community noise annoyance survey data. This method requires an estimate of the distribution of true doses, the dose error distribution, and an estimated functional form of the responses, e.g., a logistic function. The response function is transformed by the distributions of true doses and dose error, and then the transformed function is fit to the estimated dose-response data. The transformed function fit still includes the parameters of the assumed functional form of the responses, so the final dose-response relationship uses these parameters.

Bayesian methods are also available for accounting for dose error.^{26,27} These are probabilistic models that consider the estimated dose and estimated dose error to determine the true doses. These could be paired with the data cloning method, which aims to reduce the impact of the selection of a prior distribution.²⁸

APPENDIX B: CUMULATIVE DOSE-RESPONSE DATA

While the main body focused on datasets representing a population responding to individual aircraft flyovers, i.e., single events, X-59 community tests will also include end-of-day summary surveys. These responses of participants' overall annoyance to the current day's flights constitutes cumulative response. A cumulative noise dose in Perceived Day-Night Level (PLDNL) is computed from the day's single-event levels using the Day-Night Level calculation procedure:

$$PLDNL = 10 \log_{10} \left(\sum_k 10^{\frac{PL_k}{10}} \right) - 49.4 \quad (3)$$

where PL_k is the Perceived Level of the k^{th} single event of the day.

An additional 3,150 cumulative event datasets were generated and analyzed for a QSF18-based cumulative event population defined by $\beta_0 = -10.3$, $\beta_1 = 0.109$, and $\sigma_{RE} = 2.1117$. The initial X-59 community test cumulative noise exposure design shown in Figure 8 was used with units in PL Day Night Level (PLDNL), based on the single event levels from Fig. 1a. Additionally, the withdraw rate is 0.15, and the participation rate 0.6616. The cumulative results are omitted from the main body for brevity as trends are similar to the single-event data. For a sample size of 2000 participants, the dose that elicits 3%HA has a precision of ± 4 dB PLDNL.

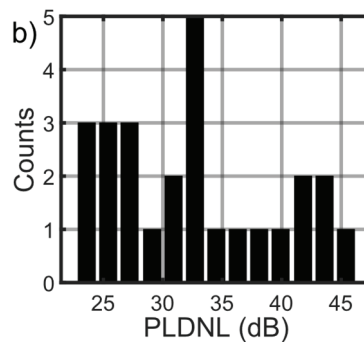


Figure 8. Cumulative dose design ($N_{\text{flight days}} = 26$).

REFERENCES

- ¹ NASA Quesst mission: <https://nasa.gov/quesst>
- ² S. V. Lympany and J. A. Page, “Estimating sonic boom metrics across a community using a Kalman filter,” *J. Acoust. Soc. Am.* **152**, (2022); <https://ntrs.nasa.gov/citations/20220017380>
- ³ R. Carroll, D. Ruppert, L. Stefanski, and C. Crainiceanu, *Measurement Error in Nonlinear Models: A Modern Perspective* (2nd ed.), Chapman and Hall/CRC, (2006); <https://doi.org/10.1201/9781420010138>
- ⁴ L. Stefanski and R. Carroll, “Covariate measurement error in logistic regression,” *The Annals of Statistics* **13**(4), p. 1335-1351, (1985); <https://doi.org/10.1214/aos/1176349741>
- ⁵ C. Kirisits, C. Lechner, and H. Kirisits, “Impact of Uncertainties Related to Noise Indicator Determination on Observed Exposure–Effect Relationship,” *Noise Health* **20**(96), p. 212-216, (2018); <https://ncbi.nlm.nih.gov/pmc/articles/PMC6301088/>
- ⁶ S. Rabe-Hesketh, A. Pickles, and A. Skrondal, “Correcting for covariate measurement error in logistic regression using nonparametric maximum likelihood estimation,” *Stat. Modelling* **3**(3), p. 215-232, (2003); <https://doi.org/10.1191/1471082X03st056oa>
- ⁷ K. S. Crump, “The effect of random error in exposure measurement upon the shape of the exposure response,” *Dose-Response* **3**(4), p. 456-464, (2005); <https://doi.org/10.2203/dose-response.003.04.002>
- ⁸ N. P. Miller, J. J. Czech, K. M. Hellauer, B. L. Bradley, S. Lohr, E. Jodts, P. Broene, D. Morganstein, J. Kali, X. Zhu, D. Cantor, J. Hudnall, K. Melia, “Analysis of the neighborhood environmental survey,” DOT/FAA/TC-21/4, (2021); <https://www.airporttech.tc.faa.gov/Products/Airport-Safety-Papers-Publications/Airport-Safety-Detail/ArtMID/3682/ArticleID/2845/Analysis-of-NES>
- ⁹ J. A. Page, K. K. Hodgdon, R. P. Hunte, D. E. Davis, T. A. Gaugle, R. Downs, R. A. Cowart, D. J. Maglieri, C. Hobbs, G. Baker, M. Collmar, K. A. Bradley, B. Sonak, D. Crom, and C. Cutler, “Quiet Supersonic Flights 2018 (QSF18) Test: Galveston, Texas Risk Reduction for Future Community Testing with a Low-Boom Flight Demonstration Vehicle,” NASA/CR-2020-220589/Volume I, (2020); <https://ntrs.nasa.gov/citations/20200003223>
- ¹⁰ W. J. Doebler, and A. Loubeau, “Updated noise dose range of NASA’s X-59 aircraft estimated from propagation simulations,” *J. Acoust. Soc. Am.* **152**, (2022); <https://ntrs.nasa.gov/citations/20220017665>
- ¹¹ J. A. Page, S. Lympany, and B. Harker, “DRD NE-01 Noise Exposure Planning Report (Preliminary),” Harris Miller Miller and Hanson (HMMH) Report No. 312430.001.001, (2021).
- ¹² J. A. Page, K. K. Hodgdon, P. Kreckler, R. Cowart, C. Hobbs, C. Wilmer, C. Koenig, T. Holmes, T. Gaugler, D. L. Shumway, J. L. Rosenberger, and D. Philips, “Waveforms and sonic boom perception and response (WSPR): low-boom community response program pilot test design, execution, and analysis,” NASA/CR-2014-218180, (2014); <https://ntrs.nasa.gov/citations/20140002785>
- ¹³ R Core Team, “R: A language and environment for statistical computing,” R Foundation for Statistical Computing, Vienna, Austria, (2021); <https://www.R-project.org/>
- ¹⁴ D. Bates, M. Maechler, B. Bolker, S. Walker, R. H. B. Christensen, H. Singmann, B. Dai, F. Scheipl, G. Grothendieck, P. Green, J. Fox, A. Bauer, and P. N. Krivitsky, “Package ‘lme4’: Linear mixed-effects models using ‘Eigen’ and S4,” (2022); <https://cran.r-project.org/web/packages/lme4/lme4.pdf>
- ¹⁵ N. B. Cruze, K. M. Ballard, A. B. Vaughn, W. J. Doebler, J. Rathsam, and P. A. Parker, “Comparison of Likelihood Methods for Generalized Linear Mixed Models with Application to Quiet Supersonic Flights 2018 Data,” NASA/TM-20220014998, (2022); <https://ntrs.nasa.gov/citations/20220014998>
- ¹⁶ M. Pavlou, G. Ambler, S. Seaman, and R. Z. Omar, “A note on obtaining correct marginal predictions from a random intercepts model for binary outcomes,” *BMC Med. Res. Meth.* **15**, (2015); <https://doi.org/10.1186/s12874-015-0046-6>
- ¹⁷ W. J. Doebler, K. M. Ballard, A. B. Vaughn, and P. A. Parker, “Dose error impacts on a collection of realistic dose-response curves based on a NASA sonic boom community noise survey,” *INTER-NOISE and NOISE-CON Congress and Conference Proceedings* **268**(4), p. 4353-4364, (2023); <https://ntrs.nasa.gov/citations/20230006770>
- ¹⁸ N. B. Cruze, “Summary and Annotated Bibliography of Measurement Error Corrections with Potential Application in Future Quesst Mission Community Noise Studies,” NASA/TM-20240015301, (2024); <https://ntrs.nasa.gov/citations/20240015301>
- ¹⁹ A. B. Vaughn, N. B. Cruze, M. A. Boucher, and W. J. Doebler, “Dose error correction using simulation extrapolation for modeling community noise dose-response relationships,” *Proc. Mtgs. Acoust.* **54**, 04002 (2024); <https://doi.org/10.1121/2.0001938>

-
- ²⁰ Sevilimedu, V. and Yu, L., “Simulation extrapolation method for measurement error: A review,” *Stat. Methods Med. Res.* 31(8), 1617–1636, (2022); <https://doi.org/10.1177/09622802221102619>
- ²¹ Cook, J. R. and Stefanski, L. A., “Simulation-extrapolation estimation in parametric measurement error models,” *J. Stat. Soc. Am.* 89, 1314–1328 (1994); <https://doi.org/10.1080/01621459.1994.10476871>
- ²² R. Carroll, D. Ruppert, L. Stefanski, and C. Crainiceanu, “Chapter 4: Regression Calibration,” in *Measurement Error in Nonlinear Models: A Modern Perspective* (2nd ed.), Chapman and Hall/CRC, p. 65-96, (2006); <https://doi.org/10.1201/9781420010138>
- ²³ R. D. Horonjeff, “An examination of dose uncertainty and dose distribution effects on community noise attitudinal survey outcomes,” *J. Acoust. Soc. Am.* **150**, p. 1691-1701, (2021); <https://asa.scitation.org/doi/full/10.1121/10.0005949>
- ²⁴ R. D. Horonjeff, “Mathematical characterization of dose uncertainty effects on functions summarizing findings of community noise attitudinal surveys,” *J. Acoust. Soc. Am.* **151**, p. 2739-2750, (2022); <https://doi.org/10.1121/10.0010311>
- ²⁵ R. D. Horonjeff, “Exposure uncertainty effects on regression parameter values,” *J. Acoust. Soc. Am.* **152**, (2022); <https://doi.org/10.1121/10.0015770>
- ²⁶ R. Carroll, D. Ruppert, L. Stefanski, and C. Crainiceanu, “Chapter 9: Bayesian Methods,” in *Measurement Error in Nonlinear Models: A Modern Perspective* (2nd ed.), Chapman and Hall/CRC, p. 205-242, (2006); <https://doi.org/10.1201/9781420010138>
- ²⁷ N. Xiang, “Model-Based Bayesian analysis in acoustics - a tutorial,” *J. Acoust. Soc. Am.* **148**(2), p. 1101–1120, (2020); <https://doi.org/10.1121/10.0001731>
- ²⁸ M. Torabi, “Likelihood Inference in Generalized Linear Mixed Measurement Error Models,” *Computational Statistics and Data Analysis* **57**(1), p. 549–557, (2013); <https://doi.org/10.1016/j.csda.2012.07.018>