

Cumulative noise metric design considerations for the NASA Quesst community test campaign with the X-59 aircraft

Aaron B. Vaughn,¹ William J. Doebler, and Andrew W. Christian
NASA Langley Research Center
2 N Dryden St., MS 463, Hampton, VA 23681 USA

ABSTRACT

The National Aeronautics and Space Administration (NASA) plans to collect X-59 aircraft dose-response data to inform international regulators in their efforts to establish acceptable noise standards to replace the current ban on overland supersonic commercial flight. For this upcoming community test campaign, the sound level from a supersonic overflight of the X-59 constitutes a single-event dose, and an end-of-day survey after multiple flight events constitutes a cumulative response. This study considers how X-59 community tests could be designed to explore the relationship between multiple single events and cumulative annoyance response. Presently, a day-night level summation of single-event doses formulates the cumulative dose; however, a generalization method may improve the understanding of the linkage between multiple single events and cumulative response. This work utilizes an analysis method that employs a parameter to vary the cumulative dose metric such that it can represent the loudest single event, remain as the day-night level summation, or be more responsive to the number of events. The analysis is demonstrated on simulated dose-response datasets derived from previous NASA field studies and from expectations of X-59 community tests. Results demonstrate that certain dose designs are more efficient for exploring the linkage between multiple single events and cumulative response.

1. INTRODUCTION

Community annoyance to sonic booms is the greatest hindrance to commercial supersonic flight [1]. This issue culminated in the ban of overland commercial supersonic flight in 1973 by the United States Federal Aviation Administration. Utilizing shaped-boom technology and as part of its Quesst mission [2], the National Aeronautics and Space Administration (NASA) is developing the X-59 aircraft to demonstrate quiet supersonic flight [3]. Community tests with the X-59 aircraft provide a unique opportunity to collect dose-response data to inform regulators in their efforts to consider replacing the supersonic speed limit with permissible noise standards.

A key to ensure the sustainability of future commercial supersonic flight is understanding how people respond to multiple supersonic overflights in a day. At its essence, community test data for supersonic overflights consist of noise doses and annoyance responses to single events and cumulative summaries, which are the end-of-day response to multiple events. While single-event data may be informative for setting aircraft certification requirements, understanding the cumulative response to multiple supersonic overflights will be important for determining the scalability of supersonic operations and therefore, their economic viability.

Presently, a cumulative dose metric has not yet been chosen for supersonic aircraft operations, though recent studies of supersonic aircraft have used day-night average sound level (DNL) for cumulative dose [4,5]. This is a common metric in community noise studies of subsonic aircraft and

¹ aaron.b.vaughn@nasa.gov

other transportation noise as it is simply a single value that describes sound exposure over time [6,7]. While DNL is well-established for regulatory purposes, it may not fully capture the complexity of the cumulative response to discrete supersonic overflight events, as it is based on the “Equal Energy Hypothesis” (EEH) where level, duration, and number of events have a prescribed tradeoff as determinants of annoyance response [8]. What if instead of DNL there was a better predictor of the annoyance to multiple quiet supersonic overflight events in a day? For instance, it is possible that people may be more sensitive to the loudest event or the number of events [9].

An analysis technique that compares differing cumulative dose metrics as predictors of annoyance can improve understanding of the cumulative dose-response relationship. In the context of novel vertical lift vehicles and previous quiet supersonic flight community data, Christian proposed an analysis method that smoothly varies a cumulative dose metric from representing the loudest event, to an equal energy summation, to a value indicative of the number of events [10,11]. This proposal includes statistical methods that can be used to determine the single-event and cumulative dose-response relationship from community noise survey data and show the level of confidence in that determination. If that response is shown to be equivalent to the EEH, then this method can verify the applicability of DNL as a cumulative metric for the given noise source. If the response appears to be more nuanced, the method can represent different behaviors – from a response driven by the loudest event that occurred over a period of time, to one determined by the number of events that occurred over the given period.

This paper evaluates dose designs for X-59 community tests as though their primary goal is to improve understanding of the linkage between single-event doses and cumulative responses. Background is provided regarding dose design considerations in previous quiet supersonic flight community studies. The methodology is subsequently outlined for the proposed analysis method to determine the relationship between individual noise events and cumulative perceptual response. A data simulation process is then explained and applied to four considered dose designs. Given the current dose designs for X-59 community tests, potential results from applying the present analysis are explored. Lastly, the ability to leverage the present analysis method in developing dose design considerations for future X-59 community tests is discussed.

2. DOSE DESIGN CONSIDERATIONS FOR PREVIOUS NASA FIELD STUDIES

Two preliminary quiet supersonic flight community studies were conducted by NASA in preparation for upcoming X-59 community tests: Waveforms and Sonic Boom Perception and Response (WSPR) [4], which were conducted in 2011 at Edwards Air Force Base (EAFB) in California, and Quiet Supersonic Flights 2018 (QSF18) [5,12], which were conducted in 2018 at Galveston, Texas. In each study, single-event and cumulative dose-response data were collected as participants responded to quiet supersonic signatures produced by an F-18 performing a low-boom dive maneuver. In analyzing response data from these tests, the candidate metric Perceived Level (PL) [13] was used for single events and its DNL counterpart, Day-Night Average Perceived Level (PLDNL), for cumulative dose. Lessons learned from each study have been applied to subsequent studies and have largely informed future X-59 community test plans.

For WSPR, the primary controlling variable in the experimental design was the cumulative daily noise exposure. The study consisted of 10 test days with five distinct targeted cumulative doses, each one to be repeated once. Given the small test area (~1 sq mi), each participant effectively received the same cumulative dose. Cumulative dose values were achieved through varying combinations of up to 14 planned single events from three noise target levels: Low, Medium, and High. The dose design attempted to balance cumulative dose across test days, the number of the three noise target levels across the design, the time separation of events, and the order of events among the sequences; however, this design was not achieved because additional conventional sonic booms occurred during the test due to non-WSPR supersonic flyovers near EAFB.

In addition to the targeted cumulative dose values, operational constraints factored into WSPR dose design and execution. Flights were limited to daytime hours with random separation times between events. Flexibility was included to allow for daily modifications of the noise exposure

schedule given the weather and instrumentation parameters. Certain atmospheric conditions are more favorable for low booms, therefore, the flight-test director was able to decide the order of the events to capitalize on ideal weather conditions.

For QSF18, the goal was to gradually acclimate the community to quiet supersonic flights over 9 flight days. Four to eight low-boom events per day were planned and achieved, as future quiet commercial supersonic flyovers in the United States are projected to occur up to 10 times per day [14]. A single event PL threshold of 80 dB was set for the first three to four days, 85 dB for the next two days, and further increased in 5 dB increments to 90 dB and 95 dB for the final two days if the community feedback suggested acceptable tolerance. However, due to negative community feedback, the highest levels (>90 dB) were not achieved. Table 1 shows the planned and achieved cumulative doses in PLDNL for these two tests.

In each study, statistical analyses were applied to the cumulative dose-response data. In WSPR, a coefficient of determination (R^2) of a linear piecewise regression suggested CDNL as the metric that best explains the variation in the percent highly annoyed. In QSF18, a statistically significant association between cumulative noise dose and annoyance response was not revealed. Essentially, there were not enough highly annoyed responses in QSF18 given the dose range for meaningful interpretation of cumulative data results. The cumulative dose range for QSF18 was lower than initially planned and substantially lower than the as-flown values for WSPR (see Table 1). Even though WSPR tested a community acclimated to hearing sonic booms, more discernable results can in part be attributed to the higher cumulative dose values. The cumulative dose range for upcoming X-59 community tests is limited by expected achievable single-event levels and flight plans: the minimum corresponds to two supersonic overflights at 70 dB and the maximum corresponds to six events at 87 dB. Therefore, similarly uninformative cumulative dose-response data as the QSF18 test are a possibility for future X-59 tests.

Table 1: Cumulative dose range in PLDNL for quiet supersonic community studies.

Study	Planned Minimum (dB)	As-flown Minimum (dB)	Planned Maximum (dB)	As-flown Maximum (dB)
WSPR	24.0	28.0	50.0	65.6
QSF18	32.0	7.3	48.0	41.1
X-59	23.6	NA	45.4	NA

There have been previous efforts to look at supersonic noise signatures from a multiple-exposure point of view. Fidell investigated the relationship between single-event and end-of-day judgements in WSPR data [15]. He concluded that it is difficult to tease out the effects of level, number, and duration of events from each other as noise metrics are generally highly correlated. One of the interesting questions posed (but not pursued here) is whether subjects were integrating their own judgments of the sounds rather than the sounds themselves (i.e., using single-event responses instead of single-event doses to create a cumulative response predictor).

3.2 ANALYSIS METHODOLOGY

The proposed analysis method to determine cumulative dose utilizes a parameter $\beth$, the Hebrew letter “bet,” and consists of the following steps: for each $\beth$ value from 0 to 1 in small increments, calculate a set of cumulative $\beth$ doses, fit the $\beth$ dose with response data to a statistical model, and evaluate the likelihood of the fit. Then a posterior distribution is computed from the likelihood values, and results are ready for interpretation.

3.1. Cumulative $\beth$ Dose Calculation

The starting point for generating a set of cumulative $\beth$ doses is the classical day-night average sound level (DNL) equation:

$$L_{dn} = 10 \log_{10} \left[\sum_{i=1}^n 10^{\frac{L_i}{10}} \right] - 49.4. \quad (1)$$

The units of L_{dn} are decibels, and in this formulation, L_i is the single-event dose level of the i^{th} of n total events in a given day and the 49.4 constant arises from integrating over the number of seconds in 24 hours.

Utilizing the generalized vector p -norm (also known as an L^p -norm) [16], Equation 1 can be reformulated into the following expression:

$$L_{dn,\beth} = 20 \log_{10} \left[\left(\sum_{i=1}^n \left(10^{\frac{L_i}{20}} \right)^{\beth^{-1}} \right)^{\beth} \right] - 49.4. \quad (2)$$

The newly introduced parameter $\beth$ is the reciprocal of p . Ignoring the time constant, if $\beth$ is set to 0, Equation 2 returns the greatest L_i value, which is the loudest single-event dose. When $\beth$ is 0.5, Equation 2 is equivalent to Equation 1, meaning the original DNL is returned. As $\beth$ approaches 1, Equation 2 coherently sums the vector of L_i , which is more responsive to the number of events.

Applying Equation 2, a new cumulative dose, or $\beth$ dose, is computed at a given $\beth$ value for each participant for each day from their single-event doses. A set of $\beth$ doses is generated per each input $\beth$ value from 0 to 1 in 0.001 increments to generate 1000 $\beth$ dose sets, which will allow for smoothly varying trends. Note that the survey responses are unchanged by $\beth$.

An example of the utility of the $\beth$ parameter is given in Figure 1. Consider three different combinations of single events, all of which have the same DNL value: one event at 80 dB, two events at 77 dB, and four events at 78, 74, 70, and 65 dB. If individuals happened to be differently annoyed by these three cases, DNL would have no utility for resolving the mechanism causing the annoyance as shown by the intersecting lines at $\beth = 0.5$ in Figure 1. However, as a function of $\beth$, there is an opportunity to consider the loudest event at $\beth = 0$, the number of events at $\beth = 1$, or some intermediate $\beth$ value to explain the annoyance as a function of a dose metric.

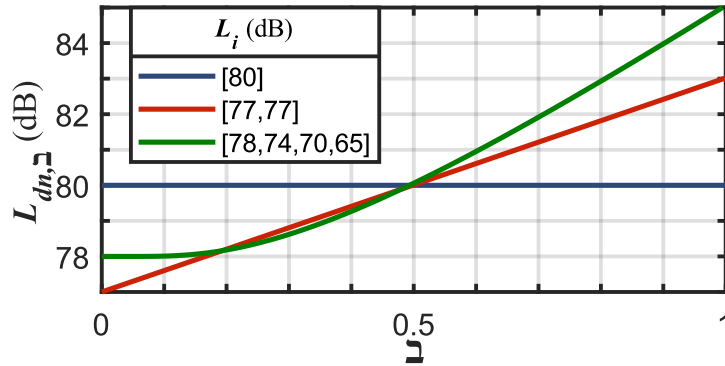


Figure 1: Example of $L_{dn,\beth}$ values (without the -49.4 term for simplicity of comparison) as a function of $\beth$ for three combinations of single events with equivalent DNL.

3.2. Statistical Dose-Response Modeling

A simple logistic regression is used to model a cumulative dose-response relationship. In a simple logistic regression, data are fully pooled and treated as independent observations. Due to the longitudinal nature of previous and future quiet supersonic flight studies, the data may contain order effects [17] and within-subject correlation, which can be accounted for with a more complex model such as a multilevel model [18]. However, simple and multilevel logistic regression produced similar single-event dose-response curve results for community data from QSF18 [19]. While it is quite feasible that individuals may be annoyed differently than the trend of the overall population, for simplicity, a simple logistic regression is considered in the present analysis to observe the overall population trends.

The logistic form produces a probability of high annoyance p as a function of j cumulative dose and estimated β parameters, as follows:

$$p_{j,\mathfrak{D}} = \text{logit}^{-1}(\beta_0 + \beta_1 \times L_{dn,\mathfrak{D},j}) = 1 / (1 + e^{-(\beta_0 + \beta_1 \times L_{dn,\mathfrak{D},j})}). \quad (3)$$

The output probability, p , is bounded from 0 to 1 but can be scaled to percent highly annoyed (%HA) from 0 to 100% when depicting population summary dose-response curves. The simple logistic regression is performed via the MATLAB `fitglm` function with the logit link and binomial distribution options. The function takes in binary “not highly annoyed” and “highly annoyed” (HA) response data paired with the $\mathfrak{D}$ dose values and outputs β_0 and β_1 that are respectively intercept and slope parameters.

3.3. Likelihood Evaluation

The output of interest of the regression is the likelihood function $\mathcal{L}$ of $\mathfrak{D}$. The likelihood function is a central concept in statistical inference and contains the information needed for probabilistic reasoning [20]. The likelihood is essentially a relative goodness of fit metric useful for differentiating between different settings of parameters (like $\mathfrak{D}$). The likelihood can be contrasted with R^2 , which is an absolute goodness of fit metric that describes how well the model is fitting to the data but cannot easily be used to distinguish between parameter settings. The likelihood statistic is computed in the log domain for computational stability purposes and describes the likelihood of the data given the fit β parameters and input $\mathfrak{D}$ value. A Bernoulli likelihood computation is used for logistic regression:

$$\mathcal{L}(\text{Data}|\mathfrak{D}) = \prod_{j=1}^N (p_{j,\mathfrak{D}})^{h_j} [1 - p_{j,\mathfrak{D}}]^{1-h_j} \quad (4)$$

For response j , h_j takes either a 1 or 0 based on whether the response was respectively HA or not.

Performing this computation for the best-fit β parameters as a function of $\mathfrak{D}$ results in a “profile likelihood” [21]. This is a one-dimensional function of only $\mathfrak{D}$ for which all other model parameters were maximized over. Thus, the β parameters are regarded as “nuisance” parameters and are ignored for inference once maximized out. Pathologies that can sometimes get in the way of implementing such a scheme (e.g., jaggedness of the likelihood function, see Ref. [20]) have not been observed in the simulations below. Therefore, it is assumed to be more expedient to compute the profile likelihood for $\mathfrak{D}$ using commonly available logistic regression software, instead of computing a representative of the full high-dimensional likelihood function (i.e., with a Monte Carlo approach) and then integrating out the nuisance parameters.

3.4. Posterior Distribution Formulation

After evaluating the likelihood for each dose-response fit at each $\mathfrak{D}$ value, a posterior distribution can be formulated. This analysis method tends toward a Bayesian interpretation, and thus the posterior distribution conceptually encodes the probability that $\mathfrak{D}$ had to be a particular value in order to give rise to the data that were observed (i.e., “the probability of $\mathfrak{D}$ given the data,” $Po(\mathfrak{D}|\text{Data})$). This is computed using Bayes’ rule:

$$Po(\mathfrak{D}|\text{Data}) = \frac{\mathcal{L}(\text{Data}|\mathfrak{D}) \cdot Pr(\mathfrak{D})}{P(\text{Data})} \quad (5)$$

where $Pr(\mathfrak{D})$ is the prior distribution and $P(\text{Data})$ is a normalizing factor [22]. The prior is a vehicle for adding preconceptions as to what $\mathfrak{D}$ might be before performing the analysis. In order to give no predisposition to any value of $\mathfrak{D}$, this study utilizes a “noninformative” prior, which is equal to a standard uniform distribution over the unit interval [0, 1] as $\mathfrak{D}$ values outside of this interval are discounted as unphysical or nonsensical. The step between the likelihood function and posterior is therefore trivial: a multiplication by 1 and a division to normalize the area under the $\mathcal{L}$ curve to 1.

3.5. Interpretation

Two descriptors of the posterior distribution results for $\mathfrak{D}$ are used here: a point estimate of the single best guess of the $\mathfrak{D}$ parameter and an interval estimate about the point estimate that indicates the

precision based on observed data. The point estimate is simply the maximum or peak of the posterior distribution, corresponding to the most probable value of α given the data. This value of α could be considered as a dominant, single descriptor of the general behavior of respondents.

The peak point estimate is complemented by an interval estimate. Noting that the area encapsulated by the posterior distribution equals 1, a critical value Po^* of $Po(\alpha|Data)$ can be found for which the area under the curve is equal to a predetermined criterion. For instance, a criterion of 0.95 would be used to determine the set of α that contains 95% of the probability of that parameter given the data by computing Po^* and taking all α for which $Po(\alpha|Data) > Po^*$ are within this interval. This technique originates from Bayesian inference and generates a credible interval (CI) [22]. This interval is similar in concept to the frequentist confidence interval; however, it is distinct. The credible interval denotes where 95% of the probability of the true α value is contained given the data as opposed to the frequentist notion of determining an interval where 95% of the results of other identical tests will fall if certain assumptions hold.

The CI will necessarily contain the maximum and some neighborhood around it (provided the function is continuous). The width of this interval around the maximum is an indication of how precise the data are in determining the best α . A broad interval indicates that the data might not strongly support one α over another, and a narrow interval provides more assurance in the estimate of the best α . Further, the CI width typically relates to the curvature of the posterior distribution (and likelihood function) near the maximum. Statistical “information” (in the Fisherian sense, perhaps) is often defined as some measure of the curvature of the likelihood function near its maximum, where high information corresponds to high curvature or a narrow peak [23].

Not all data are equally informative for confidently determining α and hence, understanding the drivers behind cumulative response behavior. It is possible to have two equally sized sets of data with one being informative and the other not informative for α . This notion of information-density is important for development of test plans, as one will try to maximize the expected information gathered for parameters of interest by modifying the test design. Generally, experimental designers seek a test design that creates contrasting outcomes for different settings of the parameter of interest. The higher this contrast, the higher the statistical power of the test, the more possible information one will gain about a parameter (cf. introduction of Ref. [24]). In this context, potential dose designs for community tests will be examined with the α analysis to determine their relative informativeness.

4. SIMULATED DOSE-RESPONSE DATA

Dose-response data from a hypothetical population are simulated for four dose designs and 11 response cases, each with 30 replicates. Each dataset has 1,000 simulated participants, similar to a single X-59 community test, and the simulated participants have a 100% response rate for the 26 flight days. A brief description of the single-event dose simulation is provided, followed by an explanation of the cumulative response generation process. Then, the four presently considered dose designs are described.

4.1 Dose Simulation Description

For each simulation, single-event doses are assigned to each participant based on the given dose design. These planned single-event doses in the dose designs are the inner carpet mean (ICM), which is the average level across the lateral extent of the boom carpet. Since participants will be located throughout the community during a supersonic overflight event, single-event dose levels are expected to vary due to geographical location. To introduce the effect of geographical variation of participants in the present simulation, single-event doses are varied based on the expected inner carpet range (ICR) of PL as determined in a separate simulation study [27]. From that study, the ICR was found to be 5 dB for quieter booms at 70 dB and 11 dB for louder booms at 87 dB. These correspond to a spread about the ICM of 70 ± 2.5 dB and 87 ± 5.5 dB. The spread is linearly interpolated here such that the spread S as a function of planned dose PL_p is $S(PL_p) = 0.176 \times PL_p - 9.853$, where PL_p can take values between 70 and 87 inclusive. A participant’s realistic single-event dose PL_r is a draw from a uniform distribution based on the spread S and the planned dose PL_p given as

$PL_r \sim \text{Uniform}[PL_p - S(PL_p), PL_p + S(PL_p)]$. These single-event doses are then combined into cumulative Σ doses using Equation 2 and cumulative responses can then be generated.

4.2. Response Generation

Cumulative response values are randomly drawn from a Bernoulli distribution based on a logistic function (see Equation 3) with user-defined slope and intercept (β_0 and β_1) parameters. The response generation process is implemented by generating a uniform random number ranging from 0 to 1 and comparing it with the probability of high annoyance for the defined β parameters at a given cumulative dose value. If the random number is greater than the probability of high annoyance, then the response is assigned a 0 for “not highly annoyed,” and vice versa with 1 for “highly annoyed.” The cumulative dose value at which the response is generated in this simulation is the PLDNL ($\Sigma = 0.5$). The response generation process is repeated with all participant doses in each simulated dataset.

Eleven different pairs of β parameters were used to generate responses to allow for the observance of Σ analysis result trends as a function of %HA range. The specific parameter values are provided in Table 2. Sets of β parameters were selected such that the 1%HA point is at 35 PLDNL, as observed for QSF18 cumulative dose-response data [18], and that the %HA at 45 PLDNL varied from 5% to 99% as depicted in Figure 2(a). Based on QSF18 results, 5%HA might be unattainable for the given cumulative dose range; however, if slightly annoyed responses were considered instead, 30% may be a realistic range. Figure 2(b) compares the percent of HA responses for the subsequently described four dose designs with design A having slightly fewer HA responses and design C having relatively more highly annoyed responses. Across 30 realizations for each case, the total percent of HA responses were all within 0.5% with a standard deviation of 0.1 to 0.2.

Table 2: Response generation parameters and quantities of interest.

%HA Range	5%	10%	20%	30%	40%	50%	60%	70%	80%	90%	99%
β_0	-10.4	-13.0	-15.8	-17.7	-19.3	-20.7	-22.1	-23.6	-25.5	-28.4	-36.8
β_1	0.165	0.240	0.321	0.375	0.419	0.460	0.500	0.544	0.598	0.679	0.919
1%HA (dB)	35.0	35.0	35.0	35.0	35.0	35.0	35.0	35.0	35.0	35.0	35.0
50%HA (dB)	62.8	54.2	49.3	47.3	46.0	45.0	44.2	43.4	42.7	41.8	40.0

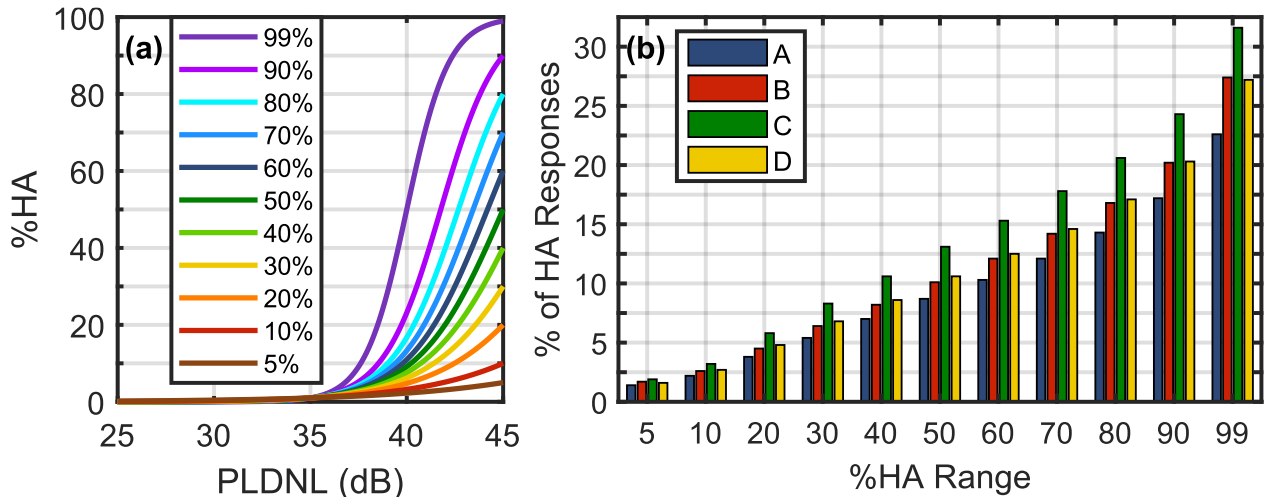


Figure 2: (a) Dose-response curves and (b) percent highly annoyed (%HA) for each of the 11 response generation cases.

4.3. Dose Designs

Current dose schedule planning for X-59 community tests strives to achieve a uniform distribution across the attainable range of single-event and cumulative doses [25]. The uniform distribution target is motivated by dose uncertainty work by Horonjeff [26], which indicated a uniform distribution of delivered doses had the lowest error of estimated dose response curves compared to three other distributions. There may be further considerations in determining an ideal dose design and target distribution, such as the specific models and analyses intended for the collected X-59 data.

An overview of the four dose designs examined in the present analysis is given in Table 3. The flight days and sorties are operational constraints while the number of single events, single-event and cumulative targets, and weighting methods are important factors of a dose design. The targets are how many doses the noise exposure scheduling software attempts to fill per bin, while the weighting method determines the tradeoff between filling bins and striving for a uniform distribution. Design A is considered the baseline or current dose design for X-59 tests. Design B is similar to A but is a different random generation with slightly more single events and sorties. In order to achieve a more uniform single-event dose distribution than produced by designs A or B, design factors are varied for designs C and D. Design C tends toward maximizing the number of single events, and design D weights for a uniform single-event distribution rather than attempting to fill single-event target bins.

Table 3: Design factors for the four dose designs.

	A	B	C	D
Number of Single Events	86	90	118	102
Flight Days	26	26	26	26
Sorties	50	53	66	58
Single-event Targets	3	3	9	NA
Cumulative Targets	1	1	2	2
Weighting Method	Fill bins	Fill bins	Fill bins	Uniform

The four dose designs utilized the anticipated dose range of the X-59. This dose range is based on the expected inner carpet mean (ICM) PL of the X-59 [27]. For each dose design, binned doses span this anticipated ICM PL dose range as shown in Figure 3(a) and 3(b). Figure 3(c) and 3(d) show the single-event and cumulative doses with the same number of total counts after applying the uniform spread in the simulation due to geographical variation of participants. The resulting cumulative dose distributions are relatively uniform across the range for both the target and smeared distributions with the notable exception of A being less uniform. Further perturbations are expected in real data due to various other sources of variability such as turbulence.

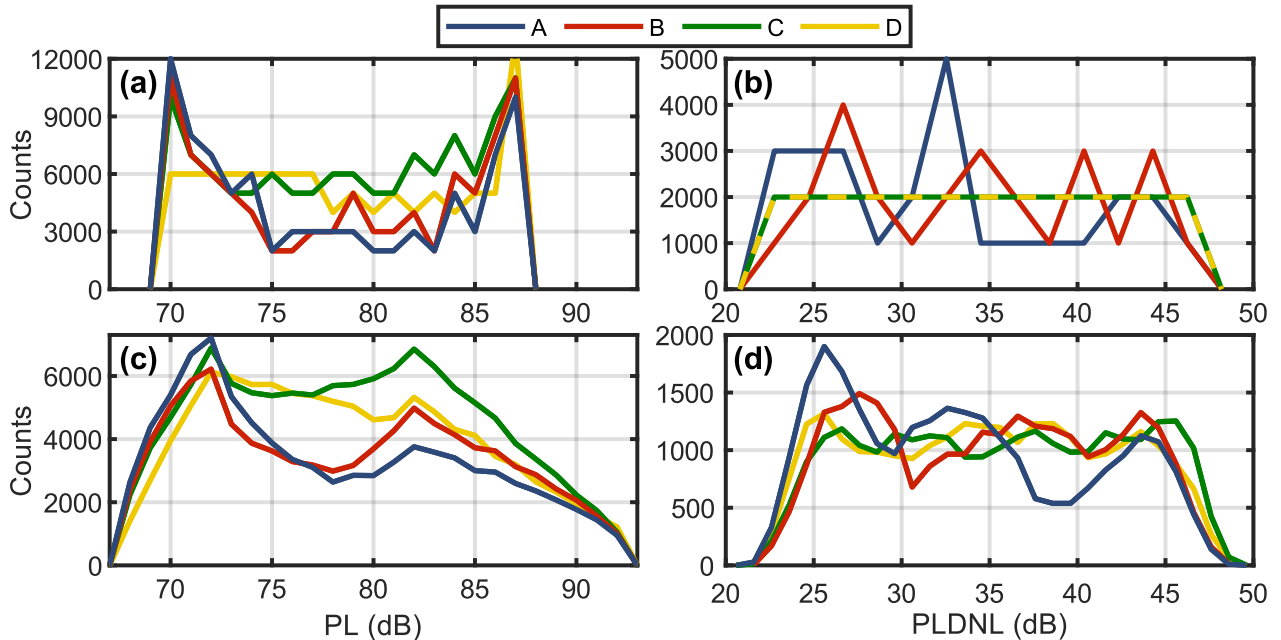


Figure 3: Planned dose design histograms for (a) single-event and (b) cumulative dose and their respective distributions due to geographical variation in (c) and (d) for 1000 simulated participants. Lines trace the midpoint of 1-dB wide dose bins, except (b) which has 13-evenly spaced bins.

The resulting four dose designs are shown in Figure 4 in terms of number of events (vertical axis), maximum single event (horizontal axis), and actual (dots) and target (diamond) cumulative

dose in PLDNL (color). Though this depiction provides a visualization of the varying dose combinations being analyzed, it does not inherently inform whether the data are ideal for Δ analysis. Having multiple ways of achieving the same PLDNL (see Figure 1) might reveal useful designs for Δ analysis. In Figure 4, there are multiple cases with the same PLDNL value; however, the trend is dominated by the loudest event, as noted by the horizontal color variation, which is expected considering the relatively few events and the logarithmic relationship between single-event and cumulative dose given in Equation 1. Other trends are apparent, such as more five and six event days for C and most days at two or six events for D. For all four designs, notice that there are no data in the upper left corner of the plots (six events at a low maximum single-event dose), a potentially important omission given the prospective future operational scenario with numerous low-level flight events.

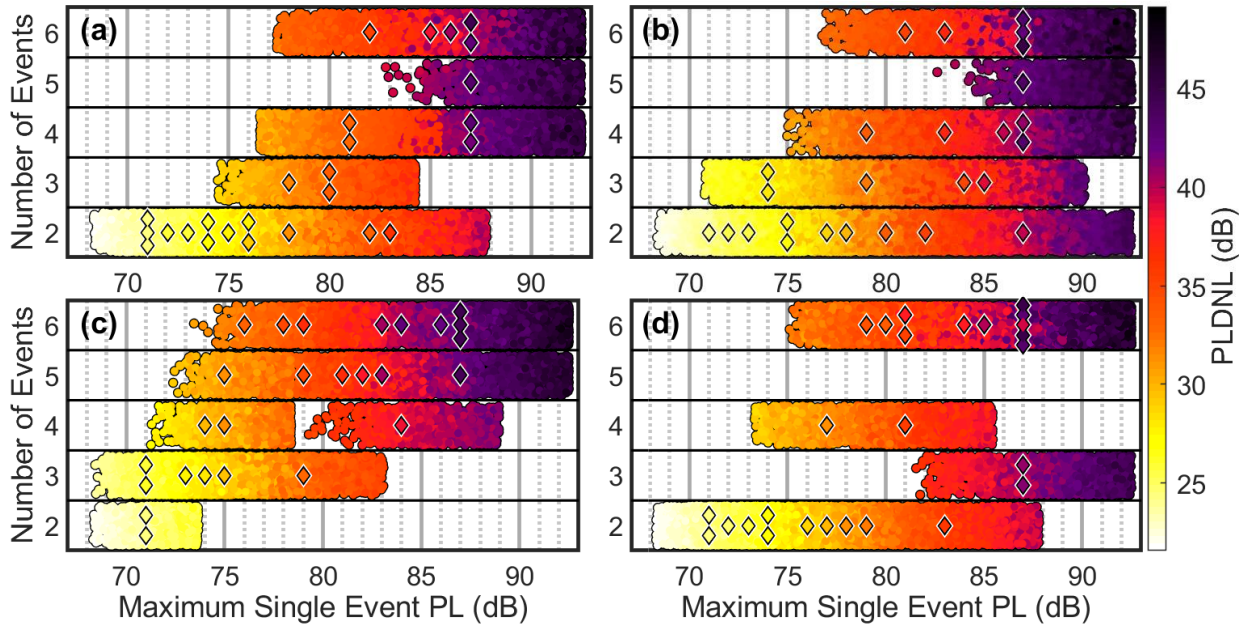


Figure 4: Simulated dose design as a function of maximum single event PL, number of events, and PLDNL for designs (a) A, (b) B, (c) C, and (d) D. Jitter is added to the number of events values and diamonds denote the planned dose design.

5. RESULTS

Key results of this analysis are the 95% CI widths of the Δ posterior distribution for various simulations. The choice of $\Delta = 0.5$ in the response generation means that the Δ posterior distributions should peak at 0.5 and the CI is equidistant from the $\Delta = 0$ and 1 bounds. While the Δ peak is not necessarily expected to be 0.5 for real-world data, previous simulations showed that the CI width is not impacted significantly by the placement of the Δ peak [11], and therefore, information regarding the general CI width trends should be transferrable to future collected data. Narrower CI yields a more precise Δ estimate and suggest a more informative dose-response dataset. For a given response case, comparing CI width across dose designs indicates which designs are more desirable.

An example of Δ posterior distributions at the 10%HA response case for the four dose designs is depicted in Figure 5. Setting the same limits on the ordinate allows for easy comparison, as the more peaked distributions have narrower CI. Each distribution appears normal, and for this case, dose designs B and D yield a more precise estimate of Δ . Recall that all four designs have the same number of cumulative dose-response pairs and similarly uniform PLDNL distributions (see Figure 3(d)). Despite design C having the most HA responses (see Figure 2(b)), it has the broadest distribution. And even though designs A and B are both based on the baseline schedule, they perform quite differently. These results demonstrate that dose-design components beyond the distribution and number of HA responses can have significant impact on the precision of estimating Δ . While these results are from a single response case, the same trends persist for the other response cases.

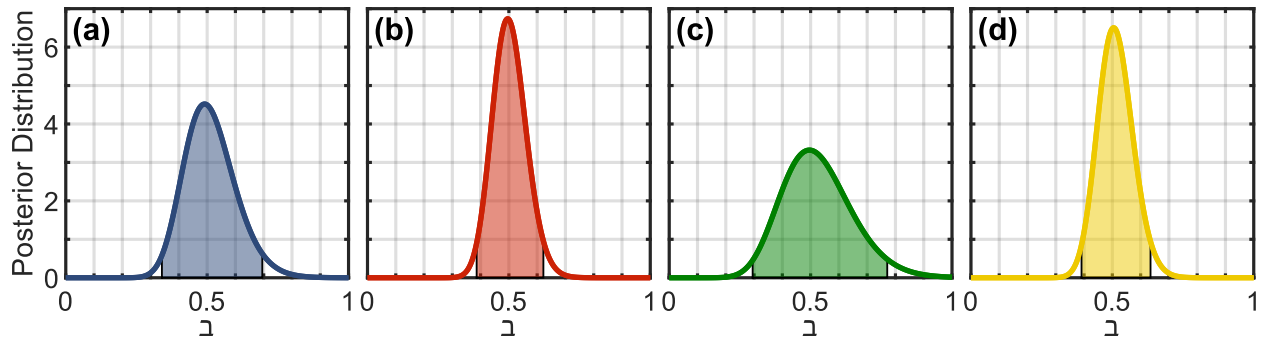


Figure 5: Posterior distributions of α at the 10%HA response case for (a) A, (b) B, (c) C, and (d) D dose designs.

A summary of all 30 replicates at each of the 11 response cases and four dose designs are captured in Figure 6. Each box and whisker plot denotes the 0th, 25th, 50th, 75th, and 100th percentiles across the 30 replicates. Figure 6(a) depicts the α posterior distribution 95% CI widths whereas Figure 6(b) notes the peak α values. Across all four dose designs, the CI widths become broader as %HA range decreases. Similarly, the peak α value more widely varies at lower %HA range cases. However, as noted in Figure 5, designs B and D consistently have narrower CI widths, regardless of the response case and despite having a smaller percent of HA responses than C (see Figure 2(b)), indicating that percent or number of HA responses is not the only factor impacting efficiency. As the interval estimate width tends to be related to the square root of the sample size asymptotically, these more efficient designs (B and D) differ from the other designs in CI width by a factor of roughly 2, which corresponds to an increase of the effective sample size by a factor of 4. Alternatively, using the more efficient designs, one might be able to conduct a test with a fourth as many responses in order to gain the same precision on α . While dose-response data collected in future tests are likely less idealized and confounded with other issues, such as dose uncertainty, these results further demonstrate the impact of dose design on potential α analysis results.

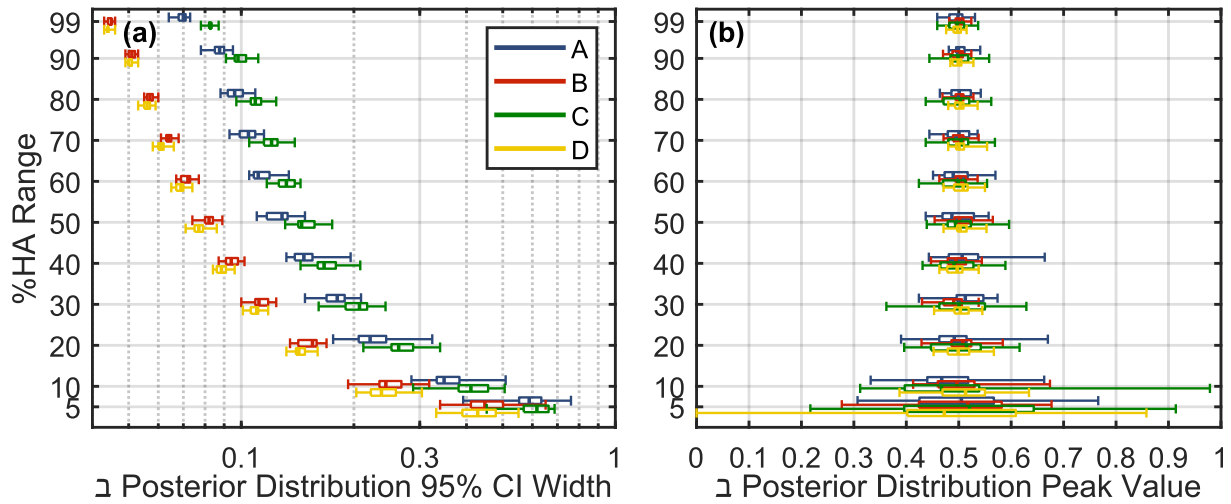


Figure 6: α posterior distributions (a) 95% credible interval width and (b) peak value for the four dose designs at the 11 response cases.

All four designs appear consistent, as they are all asymptotically leading toward the correct α parameter, but two are considerably more efficient in getting there. Interestingly, the two preferred designs are not superlative in their percent or number of HA responses, indicating that there is a structural aspect to the dose design that is contributing positively to resolving α . One consideration may be that given relatively fewer HA responses (see Figure 2(b)), there may be greater importance in the extremes of the potential predictor variables (i.e., maximum single event, DNL, and number of events). For example, for design D in Figure 4(d), there are five combinations of six single events

with a maximum single event at 87 dB. These five combinations have differing PLDNL values, which contrasts the example in Figure 1 with the same PLDNL values, but this may suggest that achieving a maximum potential predictor value in differing ways may help yield more informative Δ results. Some design factors that might not necessarily facilitate better Δ results are maximizing the number of single events (e.g., design C) and targeting a uniform cumulative dose distribution. Future work that systematically tests differing dose designs may lead to the ability to characterize the utility of a given dose design for Δ analysis. If a simple measure were developed for checking the goodness of a design for this analysis, it could provide a single-ended procedure for use in dose design efforts without simulation.

5. CONCLUDING REMARKS

This paper explores preliminary cumulative dose designs for upcoming X-59 community tests using an analysis technique that resolves the best predictor of cumulative annoyance. Dose-response data are simulated based on four dose designs and responses that demonstrate a range from idealized to more realistic scenarios. While all four dose designs have the same amount of data and similarly uniform distributions, results suggest that two designs are preferable in terms of producing more precise estimations of Δ , which indicates the ability of the analysis to resolve the underlying mechanisms driving annoyance.

The dose design for X-59 community tests is not finalized nor is the choice limited to the four presently considered designs. Additional investigation into the design elements that yield more precise estimates of Δ can help inform the formation of an optimal dose design for the Δ analysis. Potentially, a single parameter could be found that describes the goodness of dose design for Δ analysis. A design built for the Δ analysis could increase the understanding of the underlying noise mechanism that drives annoyance to multiple quieter supersonic overflights, which in turn would 1) update the cumulative noise metric to be more reflective of people's annoyance, 2) inform more effective noise mitigation techniques, and 3) support the implementation of regulations and flight operations to minimize potential community annoyance.

ACKNOWLEDGMENTS

This work is funded by the NASA Commercial Supersonic Technology Project. The authors are grateful for the following reviewers of this manuscript: K. Ballard, J. Rathsam, G. Shah, N. Cruze, and R. Cabell. In addition, we acknowledge J. Page, S. Lympany, and B. Harker of Blue Ridge Research and Consulting for their work on the preliminary dose designs.

REFERENCES

1. D.J. Maglieri, P.J. Bobbitt, K.J. Plotkin, K.P. Shepherd, P.G. Coen, and D.M. Richwine. Sonic boom: Six decades of research. Technical Report No. NASA/SP-2014-622. <https://ntrs.nasa.gov/citations/20150006843>. Last accessed 2023-05-03.
2. <https://nasa.gov/quesst>. Last accessed 2023-05-03.
3. J. Rathsam, P.G. Coen, A. Loubeau, L. Ozoroski, and G. Shah. Scope and goals of NASA's Quesst Community Test Campaign with the X-59 aircraft. *14th ICBEN Congress on Noise as a Public Health Problem*, June 2023.
4. J.A. Page et al. Waveforms and sonic boom perception (WSPR): Low-boom community response program pilot test design, execution, and analysis. Technical Report No. NASA/CR-2014-218180, 2014. <https://ntrs.nasa.gov/citations/20140002785>. Last accessed 2023-05-03.
5. J.A. Page et al. Quiet supersonic flights 2018 (QSF18) test: Galveston, Texas risk reduction for future community testing with a low-boom flight demonstration vehicle. Technical Report No. NASA/CR-2020-220589/Volume I, 2020. <https://ntrs.nasa.gov/citations/20200003223>. Last accessed 2023-05-03.

6. T.J. Schultz. Synthesis of social surveys on noise annoyance. *Journal of the Acoustical Society of America*, 64(2): 377–405, 1978.
7. G.J.J. Ruijgrok. *Elements of Aviation Acoustics*. Delft University Press, Delft, Netherlands, 1993.
8. S. Fidell and V. Mestre. *A Guide to US Aircraft Noise Regulatory Policy*. Springer International Publishing, pages 7–74 and 99–111, 2020.
9. J.M. Fields. The effect of numbers of noise events on people’s reactions to noise: An analysis of existing survey data. *Journal of the Acoustical Society of America*, 75(2): 447–467, 1983.
10. A.W. Christian. Psychoacoustics applied to aerial mobility vehicles [Conference presentation]. *Advances in Noise Control Engineering – A TQA/NAE Workshop*, 2021. <https://doi.org/10.13140/RG.2.2.14134.42566>. Last accessed 2023-05-03.
11. A.B. Vaughn and A.W. Christian. Distinguishing response behaviors within cumulative noise metrics for Quiet Supersonic Flights 2018 data [Conference presentation]. Acoustics Technical Working Group Meeting, 2023. <https://ntrs.nasa.gov/citations/20230002570>. Last accessed 2023-05-03.
12. S. Fidell, R. Horonjeff, B. Tabachnick, and S. Clark. Independent analyses of Galveston QSF18 social survey,” Technical Report No. NASA/CR-20205005471, 2020. <https://ntrs.nasa.gov/citations/20205005471>. Last accessed 2023-05-03.
13. S. S. Stevens. Perceived level of noise by Mark VII and decibels (E). *Journal of the Acoustical Society of America*, 51(2B): 575–601, 1972.
14. J. Rachami and J. Page. Sonic boom modeling of advanced supersonic business jets in NextGen. *48th AIAA Aerospace Sciences Meeting Including the New Horizons Forum and Aerospace Exposition*, 1385, 2010.
15. S. Fidell. Relationships among near-real time and end-of-day judgments of the annoyance of sonic booms. *Proceedings of Meetings on Acoustics*, 19(1): 040047, 2013.
16. R.F. Bass. Real Analysis for Graduate Students. Createspace Ind Pub, pages 131–145, 2013.
17. K.M. Ballard, A.B. Vaughn, W.J. Doebler, and J. Rathsam. Analysis of carryover effects in a low boom laboratory listener study. *Proceedings of Meetings on Acoustics*, 45(1): 040002, 2022.
18. A.B. Vaughn, J. Rathsam, W.J. Doebler, and K.M. Ballard. Comparison of two statistical models for low boom dose-response relationships with correlated responses. *Proceedings of Meetings on Acoustics*, 45(1): 040001, 2022.
19. J. Lee, J. Rathsam, and A. Wilson. Bayesian statistical models for community annoyance survey data. *Journal of the Acoustical Society of America*, 147(4): 2222–2234, 2020.
20. A.W.F. Edwards. Likelihood. Cambridge University Press, London, UK, pages 8–42, 1984.
21. J.O. Berger, B. Liseo, and R.L. Wolpert. Integrated likelihood methods for eliminating nuisance parameters. *Statistical Science*, 14(1): 1–22, 1999.
22. J.K. Kruschke. Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan, 2nd ed. Elsevier Science, London, UK, 2014.
23. K.R. Shah and B.I.K.A.S. Sinha. *Theory of optimal designs*, Vol. 54. Springer Science & Business Media, New York, USA, 2012.
24. H.M.E Miedema, H. Vos, and R.G. de Jong. Community reaction to aircraft noise: Time-of-day penalty and tradeoff between levels of overflights. *Journal of the Acoustical Society of America*, 107(6): 3245–3253, 2000.
25. J. Page, S. Lympany, and B. Harker. DRD NE-01 Noise Exposure Planning Report [Preliminary]. HMMH Report No. 312430.001.001, 2021.
26. R.D. Horonjeff. An examination of dose uncertainty and dose distribution effects on community noise attitudinal survey outcomes. *Journal of the Acoustical Society of America*, 150(3): 1691–1701, 2021.
27. W.J. Doebler and A. Loubeau. Updated noise dose range of NASA’s X-59 aircraft estimated from propagation simulations [Conference presentation]. *Journal of the Acoustical Society of America*, 152(4): A86, 2022. [Slides] <https://ntrs.nasa.gov/citations/20220017665>. Last accessed 2023-05-03.