

NASA/TM-20220014998



Comparison of Likelihood Methods for Generalized Linear Mixed Models with Application to Quiet Supersonic Flights 2018 Data

*Nathan B. Cruze, Kathryn M. Ballard, Aaron B. Vaughn, William J. Doebler,
Jonathan Rathsam, and Peter A. Parker
Langley Research Center, Hampton, Virginia*

November 2022

NASA STI Program . . . in Profile

Since its founding, NASA has been dedicated to the advancement of aeronautics and space science. The NASA scientific and technical information (STI) program plays a key part in helping NASA maintain this important role.

The NASA STI program operates under the auspices of the Agency Chief Information Officer. It collects, organizes, provides for archiving, and disseminates NASA's STI. The NASA STI program provides access to the NTRS Registered and its public interface, the NASA Technical Reports Server, thus providing one of the largest collections of aeronautical and space science STI in the world. Results are published in both non-NASA channels and by NASA in the NASA STI Report Series, which includes the following report types:

- **TECHNICAL PUBLICATION.** Reports of completed research or a major significant phase of research that present the results of NASA Programs and include extensive data or theoretical analysis. Includes compilations of significant scientific and technical data and information deemed to be of continuing reference value. NASA counter-part of peer-reviewed formal professional papers but has less stringent limitations on manuscript length and extent of graphic presentations.
- **TECHNICAL MEMORANDUM.** Scientific and technical findings that are preliminary or of specialized interest, e.g., quick release reports, working papers, and bibliographies that contain minimal annotation. Does not contain extensive analysis.
- **CONTRACTOR REPORT.** Scientific and technical findings by NASA-sponsored contractors and grantees.

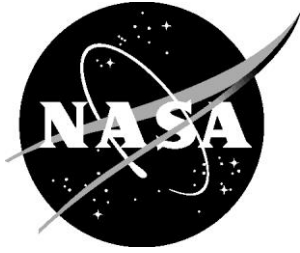
- **CONFERENCE PUBLICATION.** Collected papers from scientific and technical conferences, symposia, seminars, or other meetings sponsored or co-sponsored by NASA.
- **SPECIAL PUBLICATION.** Scientific, technical, or historical information from NASA programs, projects, and missions, often concerned with subjects having substantial public interest.
- **TECHNICAL TRANSLATION.** English-language translations of foreign scientific and technical material pertinent to NASA's mission.

Specialized services also include organizing and publishing research results, distributing specialized research announcements and feeds, providing information desk and personal search support, and enabling data exchange services.

For more information about the NASA STI program, see the following:

- Access the NASA STI program home page at <http://www.sti.nasa.gov>
- E-mail your question to help@sti.nasa.gov
- Phone the NASA STI Information Desk at 757-864-9658
- Write to:
NASA STI Information Desk
Mail Stop 148
NASA Langley Research Center
Hampton, VA 23681-2199

NASA/TM-20220014998



Comparison of Likelihood Methods for Generalized Linear Mixed Models with Application to Quiet Supersonic Flights 2018 Data

*Nathan B. Cruze, Kathryn M. Ballard, Aaron B. Vaughn, William J. Doebler, Jonathan Rathsam, Peter A. Parker
Langley Research Center, Hampton, Virginia*

National Aeronautics and
Space Administration

Langley Research Center
Hampton, Virginia 23681-2199

November 2022

The use of trademarks or names of manufacturers in this report is for accurate reporting and does not constitute an official endorsement, either expressed or implied, of such products or manufacturers by the National Aeronautics and Space Administration.

Available from:

NASA STI Program / Mail Stop 148
NASA Langley Research Center
Hampton, VA 23681-2199
Fax: 757-864-6500

This page intentionally left blank.

Abstract

Repeated measurement will be a feature of the survey data collected during the Quesst mission X-59 community response tests (CRT). Since each participant will report his or her categorical level of annoyance in response to multiple events, the responses from any single individual may be correlated with one another. Several models within the class of generalized linear mixed models (GLMM) are pertinent to the analysis of correlated categorical outcomes; the random intercept logistic regression model is one example. Both Bayesian and frequentist methods for fitting these models are available, with frequentist methods relying on some form of approximation (of either an integral or the integrand) that appears in the marginal likelihood function. Given several anticipated similarities of the X-59 CRT data to data collected during a past risk reduction, Quiet Supersonic Flights 2018 (QSF18), this short note is intended to create awareness. It documents an instance in which a reported population average dose-response relationship derived from QSF18 single event data was distorted by the integral approximation applied in likelihood-based methods. We review some of the available literature on the topic, compare the outputs of several different computational approaches implemented in available statistical software, and present simple corrective actions that may be useful during the Quesst mission.

Contents

1	Introduction	1
2	Preliminary Analysis of QSF18 Data	1
2.1	Random Intercept Logistic Regression	2
2.2	Population Average Dose-Response Curves	3
3	Available Methods and Revised Analysis of QSF18 Data	5
3.1	Methods for Estimating Parameters Provided in Common Software	5
3.1.1	Adaptive Gauss-Hermite Quadrature	5
3.1.2	Laplace Approximation	6
3.1.3	Pseudolikelihood Approximations	7
3.2	Experimenting with Multiple Choices of Software	8
3.2.1	Varying the Number of Quadrature Nodes in the <code>glmer</code> Function Call	8
3.2.2	Comparing Procedures Across Different Software	10
3.2.3	Improved Population Average Dose-Response Curves	11
4	Discussion and Conclusions	12
A	Bayesian Random Intercept Logistic Regression Model	14
B	Software Notes	14

List of Figures

1	Bar chart of categorical annoyance responses (left) and histogram of administered doses (right) collected during QSF18 study.	2
2	Notional conditional (subject-specific) dose-response curves.	3
3	Population average curves obtained by integration over the distribution of random effects. . .	4
4	Estimates of parameters of the marginal likelihood function obtained by varying the value of the <code>nAGQ</code> argument in the <code>glmer</code> function call.	9
5	Population average curves obtained after numerically integrating over the distribution of random intercepts. For clarity in plotting, corresponding confidence bands are omitted from the plot.	12

List of Tables

1	Methods of estimation provided by popular software. See notes in Appendix B for more detail.	8
2	Point estimates obtained from random intercept logistic regression by supplying select values of the <code>nAGQ</code> argument in the <code>glmer</code> function. These regressions were fit to data from the subset of 23 participants who indicated high annoyance at least once during the QSF18 study. . . .	10
3	Point estimates of parameters in the random intercept regression model obtained through different choices of software and method. See software notes in Appendix B for more detail. (*)Denotes the default setting in the <code>glmer</code> function call.	11

1 Introduction

The class of generalized linear mixed models (GLMM) extends linear mixed models to include correlated responses that are not normally distributed. In the context of past and forthcoming NASA community noise studies, a noise exposure level, i.e., a noise dose, will be used to predict perceptual response to the noise stimulus. The responses are an ordered categorical scale asking survey participants to rate their level of annoyance to noise exposure from low-amplitude sonic booms using one of the following options: *not at all annoyed*, *slightly annoyed*, *moderately annoyed*, *very annoyed*, and *extremely annoyed*. Given the regulatory emphasis surrounding community noise, these categories are often binned to create a binary outcome. The ‘very annoyed’ or ‘extremely annoyed’ responses are recoded as ‘highly annoyed’ responses, and the remaining categories are recoded as ‘not highly annoyed’. NASA has experience modeling both the polychotomous and binary outcomes in a Bayesian framework. The work of Lee et al. (2019) examined seven candidate Bayesian models, noting that the multilevel ordinal regression and multilevel logistic regression (specifically, a random intercept logistic regression) performed better than five other candidate models applied to Quiet Supersonic Flights 2018 (QSF18) data. Both of these candidate approaches are members of the class of GLMM.

For brevity, we focus only on a random intercept logistic regression model applied to the QSF18 single-event survey data captured near Galveston, Texas, in November 2018. Bayesian methods have been used extensively to fit the random intercept logistic regression to available QSF18 data (Lee et al., 2019, 2020; Vaughn et al., 2022; Doebler et al., 2022). More recent analysis of the same data has been performed using likelihood-based methods, in particular, using the solving routines available in the `lme4` package in the R statistical programming language (Bates et al., 2015; R Core Team, 2022). In the course of performing analysis on the entirety of the QSF18 single-event data, an unusual result was produced. Without obvious error or warning during processing, the population average dose-response curve obtained after marginalizing over the distribution of random effects in the frequentist random intercept logistic regression was markedly higher than its Bayesian counterpart suggested, and the resulting curve poorly fit the collected data. As the data collected during Quesst mission community response tests (CRT) to be conducted starting in 2024 may have similar features, the discrepancies between the two population average dose-response curves motivated our further investigation.

To that end, Section 2 describes the QSF18 single event data set and the initial analysis comparing population average models. A brief review of literature and software documentation on various likelihood-based methods is presented along with a revised analysis of the QSF18 data in Section 3. Discussion and conclusions that may also apply to Quesst mission CRT data are made in Section 4.

2 Preliminary Analysis of QSF18 Data

The QSF18 single-event data¹ collected in November 2018 over Galveston, Texas, consist of 4,998 dose-response pairs collected from 371 unique participants during nine test days. Categorical responses were captured and noise doses successfully estimated for 49 of 52 attempted low-amplitude sonic boom events over the study duration. While 55 participants responded just once to the single-event survey during the study, the majority of participants provided multiple responses. The median number of responses per participant was 9 responses with a maximum number of 44 responses provided by one individual in the study.

Captured categorical responses and administered doses are depicted in the left and right panels, respectively, of Figure 1. The single-event survey instrument allowed participants to indicate whether or not they heard the particular low-amplitude sonic boom event. For those that indicated that they heard the sonic boom, a follow-on question was asked: “How much did the sonic boom bother, disturb, or annoy you?” Participants indicated their annoyance on a five-point ordinal scale. A total of 2,804 responses (approximately 56% of all collected responses, shown in yellow) indicated that the low-amplitude sonic booms could not be heard, but as long as participants indicated their location on each administered questionnaire, they could still be assigned a dose. By assumption, those who indicated that they could not hear an event were assigned a ‘not at all annoyed’ response. On the high end of the scale, 34 responses were ‘very annoyed’ and 13 were ‘extremely annoyed’. After recoding to create the binary variable, only 47 (less than one percent

¹These data are publicly accessible in the supplemental files with Lee et al. (2020), <https://asa.scitation.org/doi/suppl/10.1121/10.0001021> [Last accessed: 10/04/2022]

of all responses) were ‘highly annoyed’ responses. Moreover, these responses came from a small group of 23 distinct participants. During the QSF18 study, noise exposure levels were measured in perceived-level decibels (PL dB). These doses ranged from 55 to 90 PL dB with a median dose of 72 PL dB. This range is roughly indicative of the range of anticipated doses to be administered in community overflights of the X-59 aircraft during the Quesst mission (Doebler & Loubeau, 2021). Additional details regarding QSF18 data and study procedures may be found in Lee et al. (2020) and Page et al. (2020).

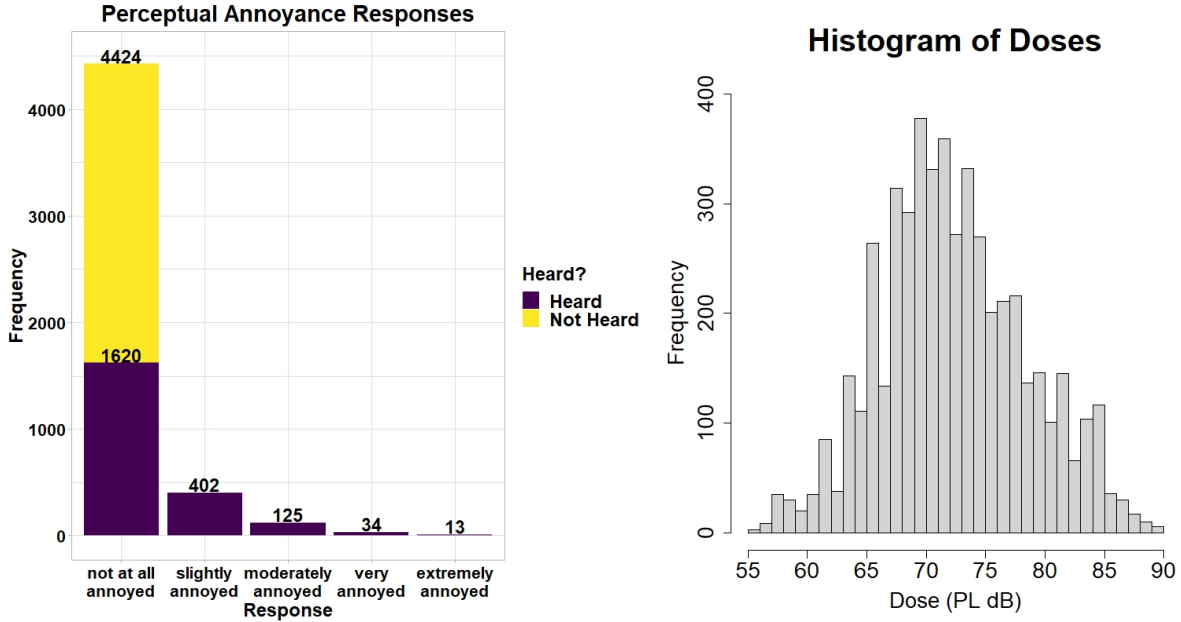


Figure 1: Bar chart of categorical annoyance responses (left) and histogram of administered doses (right) collected during QSF18 study.

2.1 Random Intercept Logistic Regression

The entire class of GLMM can be expressed compactly in matrix form as in Equation 1:

$$g(E[\mathbf{y}|\mathbf{u}]) = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}. \quad (1)$$

Upon a choice of transformation called a link function $g(\cdot)$, the mean value of a response \mathbf{y} , conditioned on one or more random effects \mathbf{u} , is linearly related to a set of fixed effects $\boldsymbol{\beta}$ and the cluster-specific random effects.

Defining the conditional mean response $E(y_{ij}|x_{ij}, u_i) \equiv p(x_{ij}, u_i)$ and choosing the logit^2 link function, the random intercept logistic regression model is specified as

$$\text{logit}(p(x_{ij}, u_i)) = \beta_0 + \beta_1 x_{ij} + u_i \text{ with } u_i \sim N(0, \sigma_u^2) \quad (2)$$

where $i \in \{1, 2, \dots, I\}$ is an index over the number of unique participants, $j \in \{1, 2, \dots, J_i\}$ indexes the (possibly different) number of responses provided by the i^{th} participant and x_{ij} denotes the noise exposure (dose) administered to the i^{th} participant during the j^{th} boom event. The scalar random effect u_i denotes a subject-specific intercept, assumed to be drawn from a normal distribution with mean zero and finite variance, σ_u^2 . The categorical response, y_{ij} , is a binary indicator variable, taking value 1 if the participant is highly annoyed and zero otherwise.

The expected value of a binary random variable is a well-defined probability or proportion. Whereas an unconstrained linear probability model may produce predicted probabilities outside the interval $(0, 1)$, the

²The logit function is the log of the odds ratio, $\text{logit}(p) \equiv \log\left(\frac{p}{1-p}\right) = k$, and taking its inverse $p = \text{logit}^{-1}(k) = \frac{1}{1+\exp(-k)}$

inverse logit maps all real values that may arise in the linear predictor space into this interval. For a specific participant, the probability that he or she is highly annoyed given dose and individual intercept is stated in Equation 3:

$$p(x_{ij}, u_i) = \text{logit}^{-1}(\beta_0 + \beta_1 x_{ij} + u_i) = \frac{1}{1 + \exp(-[\beta_0 + \beta_1 x_{ij} + u_i])}. \quad (3)$$

Equation 3 suggests a family of *conditional* or subject-specific dose-response curves as depicted in the plot of notional curves in Figure 2. In the context of community noise, this model assumes that the growth rate of annoyance is constant across all participants, but it allows for the possibility of different onset of annoyance for each participant.

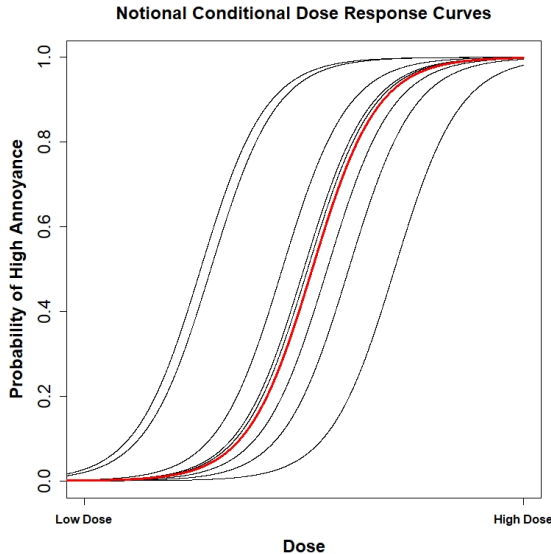


Figure 2: Notional conditional (subject-specific) dose-response curves.

2.2 Population Average Dose-Response Curves

Where Equation 3 depends on the subject-specific intercept, the ultimate data analysis goal of QSF18 (and of the Quesst mission) is to obtain a population-average (marginal) model—one that is free of additional conditions and simply relates the probability of high annoyance to the noise exposure level for a member of the population at large. Skrondal & Rabe-Hesketh (2009) and Pavlou et al. (2015) note that appropriate marginal predictions can be obtained by integration over the distribution of random effects. Thus, the strategy for obtaining a population average or marginal dose-response curve from a fitted random intercept model can be summarized as follows:

1. Obtain survey responses matched to doses. Since the QSF18 data were obtained from a panel sample, the responses provided by a participant may be correlated with his or her other responses, and the random intercept logistic regression or other GLMM may be an appropriate analysis model.
2. Fit the random intercept logistic regression model to obtain estimates of fixed effects $\hat{\beta}_0$, $\hat{\beta}_1$ and estimated variance component $\hat{\sigma}_u^2$. These parameter estimates are necessary plug-in values. (Estimates of subject-specific intercepts \hat{u}_i are also generated in this step, but they are nuisance parameters for the purposes of constructing the population average curve.)
3. Dropping subscripts, marginalize with respect to the distribution of random effects as in Pavlou et al. (2015):

$$\hat{p}(x) = \int_{-\infty}^{\infty} \frac{1}{1 + \exp(-[\hat{\beta}_0 + \hat{\beta}_1 x + u])} \phi(u|\hat{\sigma}_u^2) du \quad (4)$$

where $\phi(u|\hat{\sigma}_u^2)$ denotes a mean zero normal distribution with estimated variance $\hat{\sigma}_u^2$. The result is a function describing a probability of high annoyance (equivalently, the proportion of the population that is highly annoyed) as a function of dose alone and not depending on the intercept of any specific individual.

The points plotted in Figure 3 are observed proportions of highly annoyed responses within each 1 decibel-wide dose bin. Intuitively, the association between the probability of high annoyance and dose will be positive; the louder an event is, the more apt it is to be heard and possibly be disruptive. A single highly annoyed response out of just three total responses collected within the 56 dB bin causes the unusual outlier near the low end of the dose range. Similarly, a single highly annoyed response out of six collected in the 90 dB bin contributes to the next highest observed proportion of 0.167.

Two population average models have been plotted in Figure 3, both obtained by fitting separate random intercept logistic regression models and performing a numerical integration to marginalize with respect to the distribution of random effects as in Equation 4. The dashed line, corresponding to a Bayesian model with noninformative prior distributions as specified in Lee et al. (2020) and reproduced again in Vaughn et al. (2022) and Doebler et al. (2022), fits the preponderance of data points. (Details about this model are summarized in Appendix A.) By contrast, the population average model derived from a maximum likelihood approach produced using *default* settings in a call to the `glmer` function in the `lme4` library for R predicts probability of high annoyance that is uniformly higher across the entire dose range. Moreover, it does not fit the scatter of points well. The difference is concerning as it is understood that “we can perform Bayesian inference using noninformative or weakly informative priors and obtain results similar to classical estimates.” (Gelman et al., 2020, p. 16). Ultimately, the difference in population curves plotted in Figure 3 can be attributed to a limitation of the default settings in the `glmer` function to accurately approximate the (log)likelihood function for the random intercept logistic regression model fit to QSF18 data.

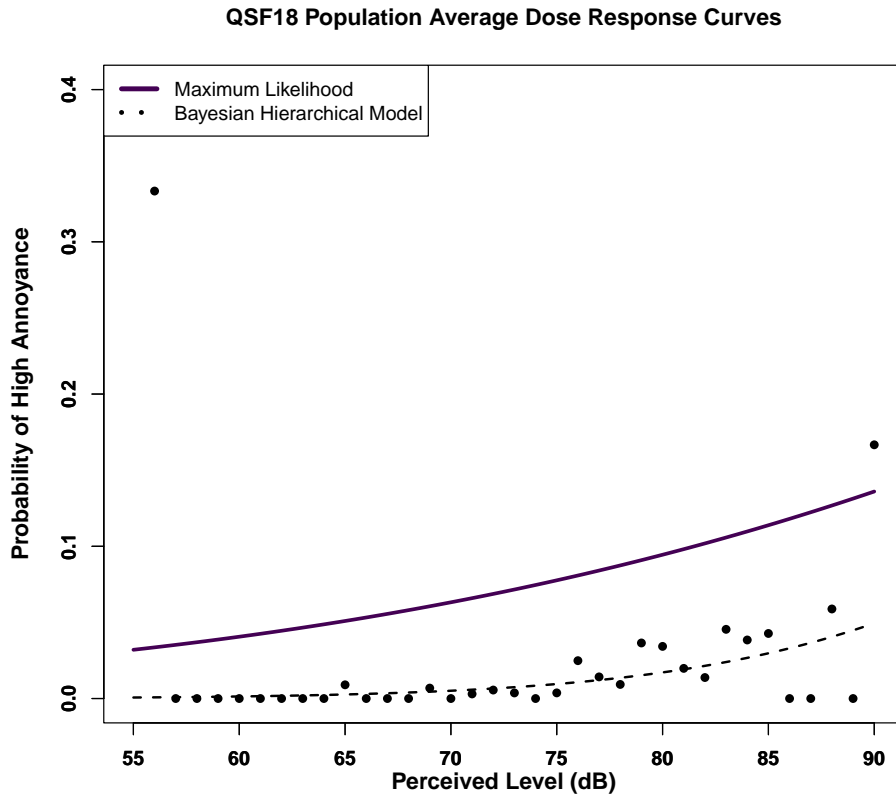


Figure 3: Population average curves obtained by integration over the distribution of random effects.

3 Available Methods and Revised Analysis of QSF18 Data

There is a considerable body of literature on maximum likelihood estimation and approximate likelihood inference for generalized linear mixed models. Several textbooks on longitudinal analysis include overviews of commonly used families of approximation techniques for likelihood-based methods; see, e.g., (Fitzmaurice et al., 2011, Ch.14, 15) and (Liu, 2016, Ch.4.2, 8.2, 8.3). A review article by Tuerlinckx et al. (2006) provides a thorough technical treatment of these families including penalized quasi-likelihood (PQL), marginal quasi-likelihood (MQL), Expectation-Maximization routines, Laplace approximations, Gaussian quadrature rules, and Markov Chain Monte Carlo (MCMC) techniques. Additionally, Kim et al. (2013), Breslow (2004), Bolker et al. (2009), Josephy et al. (2016), Ojo et al. (2017), Handayani et al. (2017), and Stroup & Claassen (2020) constitute a collection of software reviews, comparative studies, and review articles for users of GLMM offering guidance for selecting from the variety of different numerical procedures. Stroup & Claassen (2020) summarizes a few of the dilemmas faced by users of GLMM that spurred the many computational approaches to likelihood-based inference, including adaptive Gauss-Hermite quadrature and linearization strategies involving pseudolikelihood approaches (e.g., PQL and MQL).

Central to the discussion, integrals present in the marginal likelihood function of a GLMM typically cannot be evaluated analytically. Let \mathbf{y}_i denote a vector of all collected responses for the i^{th} participant, $L_i(\cdot)$ denote the contribution of the i^{th} participant to the marginal likelihood function, and \mathbf{y} and $L(\cdot)$ denote the complete-data counterparts, respectively. Then the marginal likelihood for the random intercept logistic regression is

$$\begin{aligned} L(\beta_0, \beta_1, \sigma_u^2 | \mathbf{y}) &= \prod_{i=1}^I L_i(\beta_0, \beta_1, \sigma_u^2 | \mathbf{y}_i) \\ &= \prod_{i=1}^I \int_{-\infty}^{\infty} \prod_{j=1}^{J_i} [p(x_{ij}, u_i)]^{y_{ij}} [1 - p(x_{ij}, u_i)]^{1-y_{ij}} \phi(u_i | \sigma_u^2) du_i \end{aligned} \quad (5)$$

where $p(x_{ij}, u_i)$ was defined in Equation 3, and $\phi(u_i | \sigma_u^2)$ again denotes the normal probability density function governing random intercepts. Given the collected data \mathbf{y} , maximum likelihood estimation seeks values of the common model parameters for which the marginal likelihood function obtains its global maximum, i.e., $\arg \max_{\beta_0, \beta_1, \sigma_u^2} L(\beta_0, \beta_1, \sigma_u^2 | \mathbf{y})$. In practice, computation is often performed on the natural logarithm of the

likelihood function. Since the integral in Equation 5 cannot be evaluated analytically, likelihood-based methods entail some form of approximation, and the optimizers $(\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}_u^2)$ of Equation 5 are obtained under the assumption that the approximation of the (log)likelihood function itself is sufficiently accurate. A variety of computing procedures have been developed, either to numerically evaluate the entire *integral*, or to approximate the *integrand* in the likelihood function in Equation 5 (Tuerlinckx et al., 2006).

3.1 Methods for Estimating Parameters Provided in Common Software

Rather than present in-depth descriptions of each of these methods, we give a more heuristic or procedural description of each, and we refer the reader to appropriate literature and software documentation for additional technical details. In particular, we focus on three broad classes of approaches to likelihood inference found in R, MATLAB, JMP, and SAS software, noting that apart from SAS, these separate choices of software or libraries don't necessarily provide all of these options. (Bates et al., 2015, 2022; R Core Team, 2022; MATLAB, 2021; JMP, 2022; Dong, 2020; SAS Institute Inc., 2020)

3.1.1 Adaptive Gauss-Hermite Quadrature

Adaptive Gauss-Hermite quadrature is an extension of traditional Gaussian quadrature that appeared in the context of nonlinear mixed effects models in Pinheiro & Bates (1995). In essence, it uses a recentering and scaling of traditional Gaussian quadrature to approximate integrals of the form

$$\int_{-\infty}^{\infty} \exp(-z^2) f(z) dz \approx \sum_{q=1}^Q \tilde{w}_q f(z_q) \quad (6)$$

where z_q indexes Q distinct quadrature nodes in the domain of the integral. The z_q are roots of Hermite polynomials $H_Q(z)$; the weights \tilde{w}_q have a particular functional form involving lower-order Hermite polynomials $H_{Q-1}(z_q)$ and the number of specified nodes Q (Liu, 2016, p. 262). Note that the term $\exp(-z^2)$ in the equation above is part of the kernel of a normal distribution, the distribution that governs random intercepts in the random intercepts logistic regression model. The contribution of the i^{th} participant to the marginal likelihood function is approximately

$$L_i(\beta_0, \beta_1, \sigma_u^2 | \mathbf{y}_i) \approx \sum_{q=1}^Q \tilde{w}_q \left\{ \prod_{j=1}^{J_i} [p(x_{ij}, z_q)]^{y_{ij}} [1 - p(x_{ij}, z_q)]^{1-y_{ij}} \right\} \quad (7)$$

where \tilde{w}_q and z_q are the appropriately scaled quadrature weights and nodes. In practice, maximum likelihood estimates are obtained from the log of the quadrature approximation.

We state the following as fact about adaptive Gauss-Hermite quadrature:

- Compared to traditional Gaussian quadrature, adaptive Gauss-Hermite quadrature assures that more quadrature points will lie in regions of interest (Liu, 2016, p. 262).
- In principle, the approximation of the integral can be improved with a higher number of quadrature points, and results estimated by other methods are often compared to [adaptive Gauss-Hermite quadrature] because the computation is very accurate when Q is large (Kim et al., 2013).
- The computational burden increases with the number of quadrature points and grows exponentially with the number of random effects (Fitzmaurice et al., 2011, p. 411). More complicated models with multiple random effects, especially with nesting or crossing of random effects, can quickly become computationally intractable.

As of this writing, SAS software provides an adaptive Gauss-Hermite quadrature option in the GLIMMIX procedure. The `glmex` function in the `lme4` library for R implements this technique *exclusively*. Neither JMP nor MATLAB software provide this functionality in the pertinent add-ins or toolboxes at this time.

3.1.2 Laplace Approximation

Some authors discuss Laplace approximation of the GLMM marginal likelihood function as a numerical integration approach (Liu, 2016; Stroup & Claassen, 2020), whereas others prefer to discuss it in the context of approximating the integrand and a motivation for linearization strategies (Tuerlinckx et al., 2006; Fitzmaurice et al., 2011). Discussions in both contexts are apt. Assuming a second order Taylor series approximation, the Laplace method is useful for approximating integrals of the form

$$\int_a^b \exp[Nf(z)] dz \approx \exp[Nf(z_0)] \int_a^b \exp\left[-\frac{1}{2}N|f''(z_0)|(z-z_0)^2\right] dz \quad (8)$$

where the function $f(z)$ is at least twice-differentiable and assumed to be unimodal with a maximum value obtained at some value z_0 (Liu, 2016, p. 259-260). N denotes a *large* number, and the limits of integration a and b may be infinite. It is further assumed that the matrix of second derivatives (Hessian matrix) of $f(z)$ evaluated at z_0 is positive definite. This combination of assumptions means that a second-order Taylor series expansion about the mode z_0 yields

$$f(z) \approx f(z_0) - \frac{1}{2}|f''(z_0)|(z-z_0)^2, \quad (9)$$

so that the integrand can be approximated with a more tractable form.

As it relates to GLMM and the random intercept logistic regression, the strategy is to express the marginal likelihood function in the form of the left hand side of Equation 8 by taking a natural logarithm and then exponentiating:

$$L_i(\beta_0, \beta_1, \sigma_u^2 | \mathbf{y}_i) = \int_{-\infty}^{\infty} \exp\left\{ \ln\left(\prod_{j=1}^{J_i} [p(x_{ij}, u_i)]^{y_{ij}} [1 - p(x_{ij}, u_i)]^{1-y_{ij}} \phi(u_i | \sigma_u^2) \right) \right\} du_i. \quad (10)$$

Laplace’s method can then be applied by taking a second order Taylor series expansion of the term in braces above about a mode \hat{u}_i , corresponding to an estimate of the random intercept for each of the i study subjects. In practice there is dependence of \hat{u}_i and the Hessian matrix evaluated at that point on the (unknown) estimates $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\sigma}_u^2$, and different approaches to this optimization exist (Tuerlinckx et al., 2006, p. 241-242). For example, MATLAB `fitglme` documentation³ identifies these different approaches as ‘Laplace’ and ‘ApproximateLaplace’; in the latter case, the fixed effects are said to be profiled out of the likelihood (Murphy & van der Vaart, 2000; Cole et al., 2013).

We state the following facts about Laplace approximation:

- Laplace approximation is *numerically* equivalent to Adaptive Gauss-Hermite quadrature with a single quadrature node (Tuerlinckx et al., 2006; Fitzmaurice et al., 2011; Liu, 2016; Stroup & Claassen, 2020; SAS Institute Inc., 2020).
- Kim et al. (2013) note that the Laplace method approximately integrates the objective function expanded about a mode \hat{u}_i and produces *asymptotically unbiased* estimates with less computational burden than adaptive Gauss-Hermite quadrature.
- The asymptotics in question are a function of cluster size (Tuerlinckx et al., 2006, p. 241). In the context of the QSF18 and Quesst mission studies, *cluster size refers to the number of responses provided by each participant* rather than the number of unique participants. Consequently, Laplace approximation may provide less accurate estimates for data featuring small cluster sizes (Kim et al., 2013).

Presently, MATLAB offers Laplace approximation in the `fitglme` function in the Statistics and Machine Learning Toolbox. The SAS GLIMMIX procedure offers Laplace approximation; the accompanying documentation emphasizes that the result is numerically but *not computationally* equivalent to adaptive Gauss-Hermite quadrature (SAS Institute Inc., 2020, p. 3817). Strictly speaking, R does not implement Laplace approximation, but the numerically equivalent results can be obtained by using a single quadrature node in the `glmer` function call. JMP software does not offer Laplace approximation in the Generalized Linear Mixed Model Add-In at this time.

3.1.3 Pseudolikelihood Approximations

The jargon ‘penalized quasi-likelihood’ (PQL) was developed in Breslow & Clayton (1993) to describe a linearization strategy for approximate inference in GLMM. The technique is also known as ‘pseudolikelihood’ approximation (Wolfinger & O’Connell, 1993; Stroup & Claassen, 2020). We opt to use the term ‘pseudolikelihood’ approximation as that term is used throughout SAS, JMP, and MATLAB software documentation. Pseudolikelihood methods approximate the integrand of the marginal likelihood function; the SAS GLIMMIX procedure documentation goes so far as to call it an approximation of the model (SAS Institute Inc., 2020, p 3805).

Pseudolikelihood estimation stems from a linear Taylor approximation. Re-expressing Equation 3, the mean of the response y_{ij} , conditional on the random intercept u_i , is

$$y_{ij} = \text{logit}^{-1}(\beta_0 + \beta_1 x_{ij} + u_i) + \epsilon_{ij} \quad (11)$$

where ϵ_{ij} is a mean-zero error term. Taking a linear Taylor approximation of Equation 11 expanded around current estimates of fixed effects $\hat{\beta}_0$, $\hat{\beta}_1$, and the mode \hat{u}_i , and then grouping like terms, a well-defined linear mixed model for a transformed response variable (the so-called pseudo-response) and transformed error term emerges:

$$y_{ij}^* \approx \beta_0 + \beta_1 x_{ij} + u_i + \epsilon_{ij}^*. \quad (12)$$

The pseudolikelihood estimation algorithm then makes repeated calls to a linear mixed model solver, cycling between a fitting step from which new estimates $\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{\sigma}_u^2$ are obtained, and an update step in which new \hat{u}_i and new pseudo-response variable y_{ij}^* can be calculated. The process iterates until a stated convergence tolerance is reached.

Further descriptions of the pseudolikelihood approach, including derivations for the general case of GLMM, the random intercept logistic regression, and other multilevel logistic regression models, can be

³<https://www.mathworks.com/help/stats/fitglme.html> [Last accessed 10/04/2022]

found in Handayani et al. (2017); pages 443-445, 466-470 of Fitzmaurice et al. (2011); and Tuerlinckx et al. (2006), pages 243-244, respectively. We make the following additional observations about pseudolikelihood methods:

- From the outset, pseudolikelihood approaches were devised as approximate computing methods for GLMM, and potential for biased estimates resulting from a local linear approximation have been studied in the literature, spurring other methodologies or bias corrections. See, e.g., McCulloch (1997), Breslow & Lin (1995), Breslow (2004).
- The accuracy of the approximation depends on the validity of assumed normality, and extreme departures such as binary response data, poses a particular problem (Tuerlinckx et al., 2006). A simulation study in Breslow (2004) showed that the estimate of the variance component was seriously underestimated by pseudolikelihood methods for clusters of size 2, i.e., matched pairs of binary responses per individual, but that the bias diminished as a function of cluster size.
- A variant of maximum likelihood, called restricted maximum likelihood (or residual maximum likelihood (REML)) was developed to overcome some of the known small sample biases associated with maximum likelihood estimates (Fitzmaurice et al., 2011, p.101-104). Some software packages, including MATLAB and SAS, offer both variants for pseudolikelihood approximation.

Presently, versions of pseudolikelihood estimation procedures are available in MATLAB, SAS, and in the Generalize Linear Mixed Model Add-in for JMP. This procedure is not provided in the `lme4` library for R, however, a function called `glmmPQL` is available in the `MASS` library for R (Ripley et al., 2022). We do not study the behavior of the `glmmPQL` function in this report.

3.2 Experimenting with Multiple Choices of Software

As Stroup & Claassen (2020) note, “it is clear there is no one-size-fits-all best method” for obtaining maximum likelihood estimates of parameters in a GLMM. Given the variety of computational strategies for producing maximum likelihood estimates, standard statistical software packages may implement more than one procedure or leave open the possibility that the same procedure can be tuned to offer better approximations of the likelihood function. Table 1 summarizes several of the aforementioned choices of statistical software, and the provided estimation methods; if the software provides more than one estimation procedure, its default procedure is listed first.

Table 1: Methods of estimation provided by popular software. See notes in Appendix B for more detail.

Software	Provided Estimation Procedures
R <code>lme4</code> package	Adaptive Gauss-Hermite Quadrature
MATLAB with Statistics Toolbox	Pseudolikelihood, Laplace Approximation
JMP with GLMM Add-in	Pseudolikelihood
SAS GLIMMIX Procedure	Pseudolikelihood, Laplace Approximation, Adaptive Gauss-Hermite Quadrature, and others

As the preliminary analysis of QSF18 data discussed in Section 2 was conducted using R, we begin by examining the impact of the number of quadrature nodes in the `glmer` function call on the magnitudes of estimates of common model parameters in the random intercept logistic regression. Then we compare the corresponding model outputs of different procedures in R, MATLAB, and JMP, which collectively provide the three previously described estimation procedures. Finally, we perform integration over the distribution of random effects using plug-in estimates we obtained to produce and compare corresponding population average models.

3.2.1 Varying the Number of Quadrature Nodes in the `glmer` Function Call

In particular, the `lme4` package in R exclusively implements adaptive Gauss-Hermite quadrature for the loglikelihood function of a GLMM in its `glmer` function. A specific argument `nAGQ` takes a non-negative

integer value, determining the number of quadrature nodes used in the adaptive Gauss-Hermite quadrature routine. *If the user does not explicitly specify a number of quadrature nodes in the `glmer` function call, it defaults to the value of `nAGQ=1`* (Bates et al., 2022, p.34). Recall that a single quadrature node is numerically equivalent to a Laplace approximation. As a matter of programming convention, the `nAGQ` argument can also take the value 0. This does not specify zero quadrature nodes. Rather, it specifies a single quadrature node (again, numerically equivalent to a form of Laplace approximation) and uses a faster but less exact form of parameter estimation for GLMMs by optimizing the random effects and the fixed-effects coefficients in the penalized iteratively reweighted least squares (PIRLS) step (Bates et al., 2022, p.34). The values of `nAGQ=0` and `nAGQ=1` denote to the two different approaches described for dealing with the dependence of random effects on unknown fixed effects and variance components (Tuerlinckx et al., 2006, p. 242).

Tuerlinckx et al. (2006) outlined a heuristic strategy for determining an appropriate number of nodes, suggesting a sequence of analyses with an increasing number of nodes, and noting the point at which the magnitudes of estimates do not change very much. This recommendation was demonstrated in Fitzmaurice et al. in the amenorrhea data case study, where the authors tabulated the sensitivity of the estimate of the variance component of a random intercept logistic regression model to the number of quadrature points (Fitzmaurice et al., 2011, Table 14.3, p.419). The `lme4` reference manual notes that models with “a single, scalar random-effects term could reasonably use up to 25 quadrature points per scalar integral.” (Bates et al., 2022, p.34) Following this reasoning, Figure 4 depicts the course of the trio of maximum likelihood estimates $\hat{\beta}_0$, $\hat{\beta}_1$, $\sqrt{\hat{\sigma}_u^2}$ as functions of the value of the `nAGQ` argument from 0 to 25, inclusive.

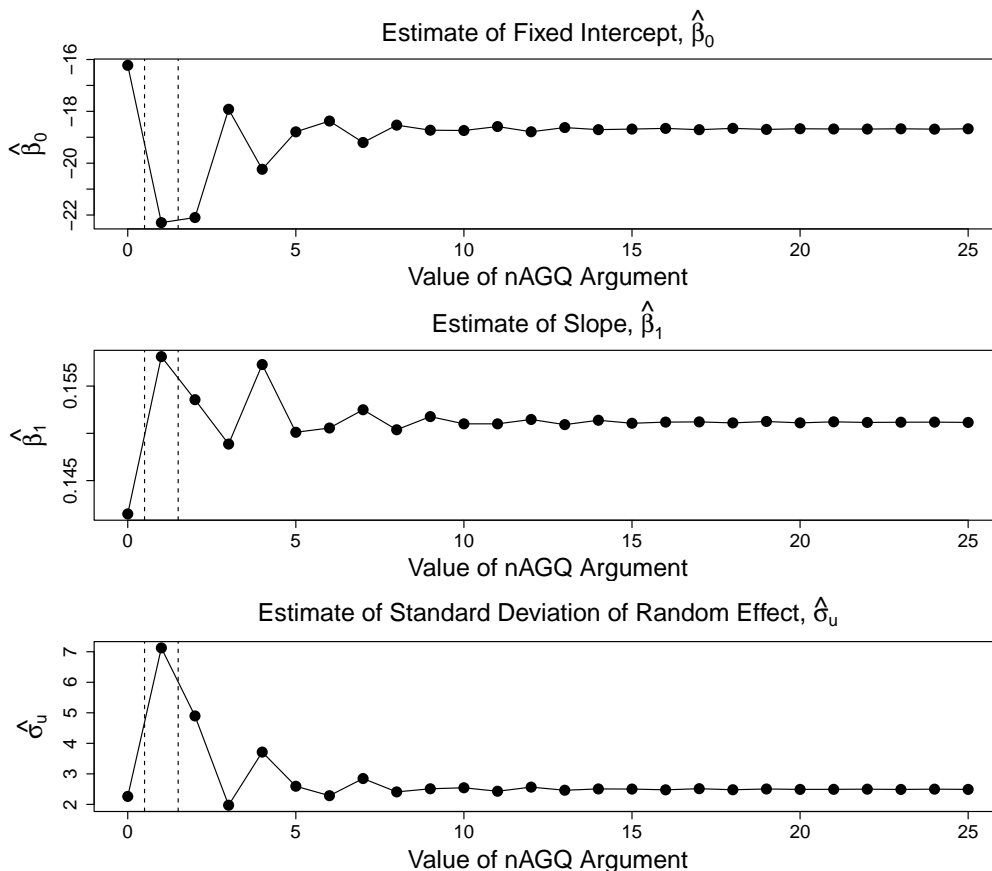


Figure 4: Estimates of parameters of the marginal likelihood function obtained by varying the value of the `nAGQ` argument in the `glmer` function call.

At the values of 0 and 1, both equivalent to forms of Laplace approximations, the estimates fluctuate greatly in magnitude. We note that the estimates obtained at `nAGQ=1` correspond to the default settings in `glmer`, and that they correspond to the plug-in values driving the unusual population average model plotted

earlier in Figure 3. As expected, an increase in the number of quadrature nodes improves the stability of estimates. We judge that 15 quadrature points may offer a sufficient approximation of the marginal likelihood function as little fluctuation is visible beyond that point.

The accuracy of Laplace approximation, and therefore, of the numerically-equivalent one-node quadrature, is known to improve with increase in cluster size (Kim et al., 2013; Liu, 2016, p. 260). As an additional experiment with quadrature, we take a subset of the QSF18 data corresponding only to those participants who were highly annoyed at least once during the study. This is a subset of 418 dose-response pairs provided by 23 unique participants over the study duration. One of the 23 participants responded just once, a second responded twice to the single event surveys, and the remaining 21 subjects each provided four or more responses with a median of 17 responses per individual. Refitting a random intercept logistic regression to this subset of data, Table 2 shows the change in parameter estimates for select values of `nAGQ`. Extra digits of precision are retained to show the minimal changes in point estimates even at coarser levels of approximation of the likelihood. Relative to the 418 collected responses, the 47 highly annoyed responses are now a more frequent occurrence (approximately 11% of cases) within the data set. Moreover, in defining this subset of data, the typical number of responses per participant (cluster size) increased from a median of 9 responses to a median of 17 responses. As the low- and high-number-of-node outcomes in Table 2 are similar in magnitude, this illustrates some of the asymptotic gains in accuracy associated with increased cluster size as well as improvement due to a (relatively) more frequent success outcome.

Table 2: Point estimates obtained from random intercept logistic regression by supplying select values of the `nAGQ` argument in the `glmer` function. These regressions were fit to data from the subset of 23 participants who indicated high annoyance at least once during the QSF18 study.

<code>nAGQ</code>	Intercept $\hat{\beta}_0$	Slope $\hat{\beta}_1$	Std. Dev. $\sqrt{\hat{\sigma}_u^2}$
0	-12.58976	0.1384273	0.2428118
1	-12.67904	0.1393362	0.2508190
2	-12.67893	0.1393376	0.2494038
4	-12.67903	0.1393415	0.2485943
15	-12.67905	0.1393418	0.2485820
25	-12.67905	0.1393418	0.2485820

3.2.2 Comparing Procedures Across Different Software

Kim et al. (2013) opined, “Investigators often use one of the three [likelihood-based] methods...because it is the default option of the package. However, the choice of method should be made according to the characteristics of data and the purpose of modeling.” Some of the aforementioned features of the QSF18 data include binary responses with rare success outcomes (i.e., high annoyance) and possibly small numbers of responses per participant over study duration. This combination of features is known to complicate the computation of maximum likelihood estimates of parameters in the random intercept logistic regression, especially for Laplace approximation and quadrature using small numbers of nodes.

Table 3 provides the point estimates for parameters of the random intercept logistic regression obtained under various settings of three popular likelihood-based methods: Laplace approximation, adaptive Gauss-Hermite quadrature, and a pseudolikelihood (linearization) approach. The point estimates obtained from the Bayesian hierarchical model are also presented for comparison. These tabulated triplets of estimates are the necessary plug-in estimates in the numerical integration step that produces the population average dose-response curve. Whereas the likelihood-based methods may entail some ‘loss’ of information due to the necessity of approximation and differences in approaches to it, Bayesian hierarchical models add information in the form of assumed prior distributions, and the obtained estimates may be sensitive to those choices of prior distribution. Thus, the goal in comparing these software outputs is not necessarily to insist on exact agreement across distinct methods, but rather, to help discern plausible results from implausible results.

Comparing estimates in the first and third rows and in the second and fourth rows of Table 3, the numerical equivalence of Laplace approximations and adaptive Gauss-Hermite quadrature with a single node is demonstrated. The default setting in the `glmer` function is a one-node quadrature, and in comparison with

the estimates obtained from higher numbers of quadrature nodes, the REML-like pseudolikelihood method, or the point estimates obtained from the Bayesian hierarchical model, these point estimates are not plausible. The single quadrature node offers only a coarse approximation to the marginal likelihood function of the logistic regression model. The estimates of the intercept and variance component ($\hat{\beta}_0$ and $\hat{\sigma}_u^2$) in particular seem to be affected, and consequently, the resulting population average dose-response curve that takes these values as given overestimates the probability of high annoyance across the entire range of noise exposure levels. This fully explains the discrepant results plotted previously in Figure 3.

Table 3: Point estimates of parameters in the random intercept regression model obtained through different choices of software and method. See software notes in Appendix B for more detail. (*)Denotes the default setting in the `glmer` function call.

Software	Method	Intercept $\hat{\beta}_0$	Slope $\hat{\beta}_1$	Std. Dev. $\sqrt{\hat{\sigma}_u^2}$
MATLAB <code>fitglme</code>	Approximate Laplace	-16.22	0.141	2.26
MATLAB <code>fitglme</code>	Laplace	-22.29	0.158	7.13
R <code>glmer</code>	Quadrature, 1 node (<code>nAGQ=0</code>)	-16.22	0.141	2.26
R <code>glmer</code>	Quadrature, 1 node (<code>nAGQ=1</code>)*	-22.29	0.158	7.13
R <code>glmer</code>	Quadrature, 4 nodes (<code>nAGQ=4</code>)	-20.24	0.157	3.13
R <code>glmer</code>	Quadrature, 15 nodes (<code>nAGQ=15</code>)	-18.68	0.151	2.50
JMP Pro	REML-like Pseudolikelihood	-18.91	0.153	2.73
R and JAGS	Bayesian Hierarchical Model	-19.00	0.153	2.61

3.2.3 Improved Population Average Dose-Response Curves

With regard to the QSF18 data, the strategy for obtaining a population average dose-response curve involves marginalizing with respect to the distribution of random intercepts, using the triplet of point estimates obtained from a selected method of fitting the random intercept logistic regression as plug-in estimates in Equation 4. Figure 5 is an update of Figure 3, including multiple population average curves based on the values of estimates obtained from select methods tabulated in Table 3. The range of the vertical axis has been reduced to add clarity to the plot, and consequently, outlier points are not displayed even though they remain in the data set.

The dashed population average line still corresponds to the Bayesian hierarchical model, which provides a useful reference. The darkest solid line, originally labeled ‘Maximum Likelihood’ in Figure 3, has been relabeled reflecting the knowledge that it was produced using the default settings in the `glmer` function call, corresponding to a one-point quadrature and numerically equivalent to a Laplace approximation of the marginal likelihood function. The overestimation of probability of high annoyance resulted because the one-point quadrature provides an approximation of the marginal likelihood function that is too coarse when applied to the totality of the QSF18 data. Consequently, the resulting plug-in point estimates of the fixed effects and variance component showed substantial biases, affecting the quality of the population average dose-response curve. After taking corrective actions and increasing the number of quadrature points, the resulting population average curves improve in terms of plausible fit to the cloud of data points, and the population average dose-response curve corresponding to a 15-node quadrature is nearly coincident with that of the Bayesian hierarchical model.

Additionally, a population average curve corresponding to use of the REML-like pseudolikelihood point estimates (obtained from JMP) in Equation 4 offers a similarly plausible fit to the plotted data. It is worth noting that the fixed effects estimates ($\hat{\beta}_0$, $\hat{\beta}_1$) obtained from the pseudolikelihood approach agree more closely with those obtained from the Bayesian hierarchical model than the corresponding estimates from the 15-point quadrature do, and yet the resulting population average curve for the 15-point quadrature is “closer” to the Bayesian counterpart. The estimated variance component ($\hat{\sigma}_u^2$) obtained from the pseudolikelihood approach is larger than the Bayesian counterpart, whereas the estimated variance component from the 15-node quadrature is smaller. This highlights the nonlinear interaction of plug-in estimates in the population average dose-response curve in Equation 4 and its sensitivity to the value of $\hat{\sigma}_u^2$ in particular.

QSF18 Population Average Dose Response Curves

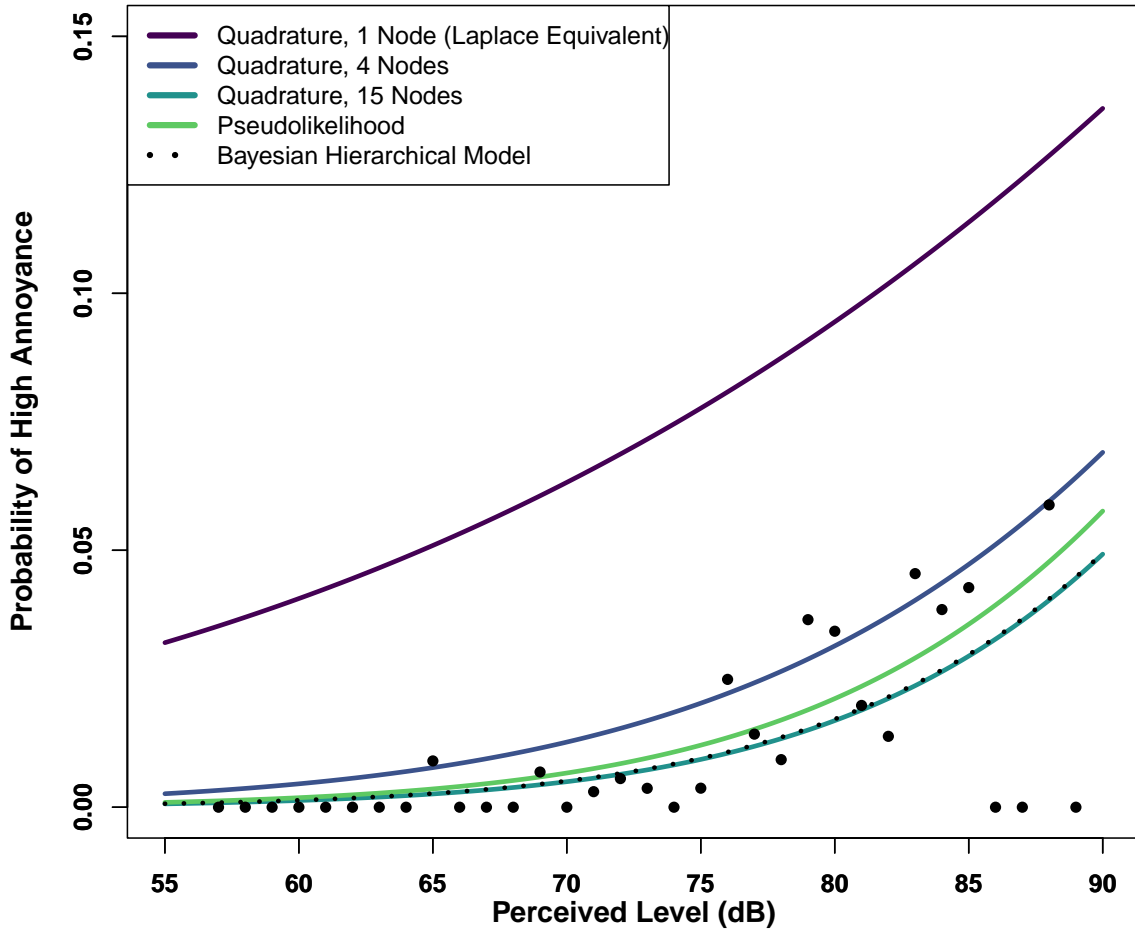


Figure 5: Population average curves obtained after numerically integrating over the distribution of random intercepts. For clarity in plotting, corresponding confidence bands are omitted from the plot.

4 Discussion and Conclusions

Large observed discrepancies in the plotted population average dose-response curves for QSF18 data can be explained by differences in approaches to approximating the marginal likelihood function of the random intercept logistic regression model. A multiplicity of likelihood-based methods, all involving some form of approximation, have been developed to fit many classes of generalized linear mixed models. Each method has strengths and weaknesses depending on the modeling context.

The QSF18 study served as a pilot study for the future Quesst mission CRT that will commence in 2024. As such, several features of the QSF18 data may be indicative of the objective measurements of noise exposure levels and the perceptual experiences of survey participants that will ultimately be collected during the Quesst mission. These data sources will be linked through a statistical model to produce a dose-response relationship. While the sample sizes at each CRT have yet to be determined, they are anticipated to be larger than the QSF18 study. At each CRT community, a longitudinal survey will be deployed, gathering repeated measurements from a recruited panel of participants. By its very design, the X-59 aircraft has been built to quiet the sonic booms produced during supersonic operation, and the resulting ‘sonic thumps’ may be less disruptive than traditional sonic booms, if they can be heard at all. Therefore, it is

expected that high annoyance will be a rare outcome in the collected survey data. Logistic regression models incorporating random effects will be likely analysis tools for exploratory analysis and dose-response modeling. The combination of rare, binary outcome and repeated measurements with a variable (and sometimes small) number of responses per participant in the QSF18 data manifested a few of the limitations of available likelihood-based methods for fitting the conditional and population average dose-response curves.

It is believed that data collected during each Quesst mission CRT will have similar features. Laplace approximation and the numerically equivalent one-node adaptive Gauss-Hermite quadrature have known limitations when the cluster size is small, but the accuracy may improve with an increased number of responses per participant. We note that each planned CRT is longer in duration with more opportunities for recruited participants to respond to surveys, and, in contrast to the QSF18 study, an incentive structure tied to the number of completed surveys will be offered to the recruited panel to encourage full participation. Nevertheless, the performance of Laplace approximation may very well depend on survey response rates, and these response rates may even differ across the five currently planned CRT sites. Pseudolikelihood approaches rely on local linear approximations to the logistic regression model and the resulting estimates may be subject to varying degrees of bias. Many practitioners point to adaptive Gauss-Hermite quadrature as a gold standard, provided a sufficient number of nodes are applied in the quadrature. In the context of the random intercept logistic regression, which has a single, scalar random effect term, the number of quadrature nodes is not so computationally prohibitive so that specifying a high number of nodes, say `nAGQ=25`, may be feasible even if data from multiple CRT sites are pooled together in one large dataset. If additional random effects terms are introduced during model development, e.g., random slopes and intercepts and/or random effects pertaining to location of each CRT site, then the computational burden grows exponentially, and smaller numbers of nodes may be a necessary compromise. We also note that as of this writing, the `glmer` function can only accommodate adaptive Gauss-Hermite quadrature with two or more quadrature nodes for models with a single, scalar random effect. Thus, the more complex models just described could either be fit using a single quadrature node (`nAGQ=0` or `1` in the `glmer` function call) provided it offered a sufficiently accurate approximation, or else, a different choice of software or library would be needed to facilitate the use of adaptive Gauss-Hermite quadrature.

A large number of candidate models with different combinations of covariates may be fit to the collected Quesst CRT data. Given their computational speed relative to Bayesian methods, likelihood-based methods will likely be employed during some of the early model-building activity. We recommend that simple assessments of the accuracy of the chosen likelihood-based estimation procedure, e.g., sensitivity of estimates to a varying number of quadrature points, be incorporated into preliminary analysis of Quesst mission CRT data and subsequent dose-response modeling.

Acknowledgments

This work was funded by the NASA Commercial Supersonic Technology Project. This research was performed by members of the Engineering Integration Branch, the Structural Acoustics Branch, and the Advanced Measurement and Data Systems Branch at NASA Langley Research Center in Hampton, Virginia. The authors thank Lisa Rippy, Chad Rice, and Amanda Cutright of the Engineering Integration Branch for additional editorial review and comments.

A Bayesian Random Intercept Logistic Regression Model

A Bayesian hierarchical model, equivalent to that of Lee et al. (2020), is specified as below:

$$y_{ij} \sim \text{Bernoulli}(p_{ij}) \quad (13)$$

$$p_{ij} = \text{logit}^{-1}(\beta_0 + \beta_1 x_{ij} + u_i) \quad (14)$$

$$u_i \sim N(0, \sigma_u^2) \quad (15)$$

$$\beta_0 \sim N(0, 10^2) \quad (16)$$

$$\beta_1 \sim N(0, 10^2) \quad (17)$$

$$\sigma_u^2 \sim IG(0.01, 0.01) \quad (18)$$

Whereas likelihood methods treat estimation as an optimization problem, Bayesian methods treat the problem as one of sampling from a posterior probability distribution. Equations 13-15 fully define the likelihood function. Noninformative prior distributions in Equations 16-18 complete the specification of the Bayesian model, namely mean-zero normal distributions with standard deviation of 10 for the fixed effects, and inverse gamma with shape and scale parameters both equal to 0.01 applied to the variance component. Given this choice of noninformative priors, it is expected that the results of the Bayesian hierarchical model should be comparable (but not necessarily identical) to likelihood-based results. The point estimates reported from the Bayesian model were posterior means obtained by (Doebler et al., 2022, Table 4), using Markov Chain Monte Carlo methods provided by R, JAGS, and related R packages (Plummer, 2003; Plummer et al., 2021) to fit the random intercept logistic regression.

B Software Notes

The following versions of software, packages, and add-ins were used in the initial and revised analysis:

- R v4.1.2
 - lme4 v1.1-27.1; `glmer` is a function in this R package that implements adaptive Gauss-Hermite quadrature methods for integral approximation in the loglikelihood function.
 - rjags v4-12
- JAGS v4.3.0
- JMP Pro v16.0.0 with the Generalized Linear Mixed Model Add-In v5. At the time of writing, JMP Pro only provides functionality for fitting generalized linear mixed models in a user community contributed add-in that provides REML-like pseudolikelihood methods for select classes of GLMM.
- MATLAB 2021a with Statistics and Machine Learning, and Optimization Toolboxes
 - `fitglm` is a function in the Statistics and Machine Learning toolbox that offers variants of Laplace approximation and pseudolikelihood methods.

References

- Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., Dai, B., Scheipl, F., Grothendieck, G., Green, P., Fox, J., Bauer, A., & Krivitsky, P. N. (2022). *Linear Mixed-Effects Models using 'Eigen' and S4*. <https://cran.r-project.org/web/packages/lme4/lme4.pdf>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J.-S. S. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology and Evolution*, 24(3), 127–135. <https://doi.org/10.1016/j.tree.2008.10.008>
- Breslow, N. (2004). Whither PQL? *Proceedings of the Second Seattle Symposium in Biostatistics: Analysis of Correlated Data*, 1–22. Springer New York. https://doi.org/10.1007/978-1-4419-9076-1_1
- Breslow, N. & Clayton, D. (1993). Approximate Inference in Generalized Linear Mixed Models. *Journal of the American Statistical Association*, 88(421), 9–25. <https://doi.org/10.1080/01621459.1993.10594284>
- Breslow, N. E. & Lin, X. (1995). Bias Correction in Generalised Linear Mixed Models with a Single Component of Dispersion. *Biometrika*, 82(1), 81–91. <https://doi.org/10.2307/2337629>
- Cole, S. R., Chu, H., & Greenland, S. (2013). Maximum Likelihood, Profile Likelihood, and Penalized Likelihood: A Primer. *American Journal of Epidemiology*, 179(2), 252–260. <https://doi.org/10.1093/aje/kwt245>
- Doebler, W. J. & Loubeau, A. (2021). The Noise Dose Range of the X-59 Estimated from Propagation Simulations. *Presentation at the Acoustical Society of America Meetings on Acoustics*. 3pCA1. Seattle, WA. <https://ntrs.nasa.gov/citations/20210024526>
- Doebler, W. J., Vaughn, A. B., , Ballard, K. M., & Rathsam, J. (2022). Simulation and Application of Bayesian Dose Uncertainty Modeling for Low-Boom Community Noise Surveys. *Proceedings of Meetings on Acoustics*, volume 45. <https://doi.org/10.1121/2.0001592>
- Dong, M. (2020). Generalized Linear Mixed Model Add-In v5 for JMP. SAS Institute Inc., Cary, NC. <https://community.jmp.com/t5/JMP-Add-Ins/Generalized-Linear-Mixed-Model-Add-in/ta-p/284627>
- Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2011). *Applied Longitudinal Analysis* (2nd ed.). John Wiley & Sons. <https://doi.org/10.1002/9781119513469>
- Gelman, A., Hill, J., & Vehtari, A. (2020). *Regression and Other Stories*. Analytical Methods for Social Research. Cambridge University Press. <https://doi.org/10.1017/9781139161879>
- Handayani, D., Notodiputro, K. A., Sadik, K., & Kurnia, A. (2017). A Comparative Study of Approximation Methods for Maximum Likelihood Estimation in Generalized Linear Mixed Models (GLMM). *AIP Conference Proceedings*, 1827(1), 020033. <https://doi.org/10.1063/1.4979449>
- JMP (2022). JMP Pro v16.0.0. SAS Institute Inc., Cary, NC, 1989–2022.
- Joseph, H., Loeys, T., & Rosseel, Y. (2016). A Review of R-packages for Random-Intercept Probit Regression in Small Clusters. *Frontiers in Applied Mathematics and Statistics*, 2. <https://doi.org/10.3389/fams.2016.00018>
- Kim, Y., Choi, Y.-K., & Emery, S. (2013). Logistic regression with multiple random effects: A simulation study of estimation methods and statistical packages. *The American Statistician*, 67(3), 171–182. <https://doi.org/10.1080/00031305.2013.817357>
- Lee, J., Rathsam, J., & Wilson, A. (2019). Statistical Modeling of Quiet Sonic Boom Community Response Survey Data. Technical report. NASA/TM-2019-220427. <https://ntrs.nasa.gov/api/citations/20190033466/downloads/20190033466.pdf>

- Lee, J., Rathsam, J., & Wilson, A. (2020). Bayesian statistical models for community annoyance survey data. *The Journal of the Acoustical Society of America*, 147(4), 2222–2234. <https://doi.org/10.1121/10.0001021>
- Liu, X. (2016). *Methods and Applications of Longitudinal Data Analysis*. Academic Press. <https://doi.org/10.1016/C2013-0-13082-6>
- MATLAB (2021). *9.10.0.1602886 (R2021a)*. Natick, Massachusetts: The MathWorks Inc.
- McCulloch, C. E. (1997). Maximum Likelihood Algorithms for Generalized Linear Mixed Models. *Journal of the American Statistical Association*, 92(437), 162–170. <https://doi.org/10.1080/01621459.1997.10473613>
- Murphy, S. & van der Vaart, A. (2000). On Profile Likelihood. *Journal of the American Statistical Association*, 95(450), 449–465. <http://www.jstor.org/stable/2669386>
- Ojo, O., Lougue, S., & Woldegerima, W. (2017). Bayesian generalized linear mixed modeling of tuberculosis using informative priors. *PLoS ONE*, 12(3). e0172580. <https://doi.org/10.1371/journal.pone.0172580>
- Page, J. A., Hodgdon, K. K., Hunte, R. P., Davis, D. E., Gaugler, T., Downs, R., Cowart, R. A., Maglieri, D. J., Hobbs, C., Baker, G., Collmar, M., Bradley, K. A., Sonak, B., Crom, D., & Cutler, C. (2020). Quiet Supersonic Flights 2018 (QSF18) Test: Galveston, Texas Risk Reduction for Future Community Testing with a Low-Boom Flight Demonstration Vehicle. Technical report. NASA/CR-2020-220589/Volume I. <https://ntrs.nasa.gov/citations/20200003223>
- Pavlou, M., Ambler, G., Seaman, S., & Omar, R. Z. (2015). A Note on Obtaining Correct Marginal Predictions from a Random Intercepts Model for Binary Outcomes. *BMC Medical Research Methodology*, 15(59). <https://doi.org/https://doi.org/10.1186/s12874-015-0046-6>
- Pinheiro, J. C. & Bates, D. M. (1995). Approximations to the Log-Likelihood Function in the Nonlinear Mixed-Effects Model. *Journal of Computational and Graphical Statistics*, 4(1), 12–35. <https://doi.org/10.1080/10618600.1995.10474663>
- Plummer, M. (2003). JAGS: A Program for Analysis of Bayesian Graphical Models using Gibbs Sampling. *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*, 1–10. <https://www.r-project.org/conferences/DSC-2003/Proceedings/Plummer.pdf>
- Plummer, M., Stukalov, A., & Denwood, M. (2021). *Rjags: Bayesian Graphical Models using MCMC*. <https://cran.r-project.org/web/packages/rjags/rjags.pdf>
- R Core Team (2022). *R: A Language and Environment for Statistical Computing*. <https://www.R-project.org/>
- Ripley, B., Bates, D. M., Hornik, K., Gebhardt, A., & Firth, D. (2022). *Package ‘MASS’: Support Functions and Datasets for Venables and Ripley’s MASS*. <https://cran.r-project.org/web/packages/lme4/lme4.pdf>
- SAS Institute Inc. (2020). *SAS/STAT 15.2 User’s GuideThe GLIMMIX Procedure*. Cary, NC: SAS Institute Inc. https://documentation.sas.com/api/docsets/statug/15.2/content/glimmix.pdf?locale=en#nameddest=statug_glimmix_details24
- Skrondal, A. & Rabe-Hesketh, S. (2009). Prediction in multilevel generalized linear models. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 172(3), 659–687. <https://doi.org/https://doi.org/10.1111/j.1467-985X.2009.00587.x>
- Stroup, W. & Claassen, E. (2020). Pseudo-Likelihood or Quadrature? What We Thought We Knew, What We Think We Know, and What We Are Still Trying to Figure Out. *Journal of Agricultural, Biological and Environmental Statistics*, 25(4), 639–656. <https://doi.org/10.1007/s13253-020-00402-6>

- Tuerlinckx, F., Rijmen, F., Verbeke, G., & De Boeck, P. (2006). Statistical Inference in Generalized Linear Mixed Models: A Review. *British Journal of Mathematical and Statistical Psychology*, 59(2), 225–255. <https://doi.org/10.1348/000711005X79857>
- Vaughn, A. B., Rathsam, J., Doeblner, W. J., & Ballard, K. M. (2022). Comparison of two statistical models for low boom dose-response relationships with correlated responses. *Proceedings of Meetings on Acoustics*, volume 45. <https://doi.org/10.1121/2.0001541>
- Wolfinger, R. & O'Connell, M. (1993). Generalized Linear Mixed Models: A Pseudo-Likelihood Approach. *Journal of Statistical Computation and Simulation*, 48(3-4), 233–243. <https://doi.org/10.1080/00949659308811554>