Food and Chemical Toxicology 151 (2021) 112125

Contents lists available at ScienceDirect

Food and Chemical Toxicology

journal homepage: www.elsevier.com/locate/foodchemtox

Bayesian hierarchical evaluation of dose-response for peanut allergy in clinical trial screening

Lynne T. Haber [a,*], John F. Reichard [a], Alice K. Henning [b], Peter Dawson [b], R. Sharon Chinthrajah [c], Sayantani B. Sindher [c], Andrew Long [c], Melissa J. Vincent [a,1], Kari C. Nadeau [c], Bruce C. Allen [d]

[a] University of Cincinnati, Department of Environmental and Public Health Sciences, 160 Panzeca Way, Cincinnati, OH, 45267, USA
[b] The Emmes Company, LLC, 401 North Washington St., Suite 700, Rockville, MD, 20850, USA
[c] Sean N. Parker Center for Allergy and Asthma Research at Stanford University and Division of Pulmonary, Allergy, & Critical Care Medicine, Department of Medicine, Stanford, CA, USA
[d] Independent Consultant, 101 Corbin Hill Circle, Chapel Hill, NC, 27514, USA

ARTICLE INFO

Handling Editor: Dr. Jose Luis Domingo

Keywords:
Peanut allergy
Risk assessment
Dose-response
Interindividual variability
Bayesian hierarchical modeling

ABSTRACT

Risk-based labeling based on the minimal eliciting doses (EDs) in sensitized populations is a potential replacement for precautionary allergen labeling of food allergens. We estimated the dose-response distribution for peanut allergen using data from double-blind placebo-controlled food challenges (DBPCFCs) conducted in the US at multiple sites, testing a population believed to be similar to the general U.S. food allergic population. Our final (placebo-adjusted) dataset included 548 challenges of 481 subjects. Bayesian hierarchical analysis facilitated model fitting, and accounted for variability associated with various levels of data organization. The data are best described using a complex hierarchical structure that accounts for inter-individual variability and variability across study locations or substudies. Bayesian model averaging could simultaneously consider the fit of multiple models, but the Weibull model dominated so strongly that model averaging was not needed. The ED01 and ED05 (and 95% credible intervals) are 0.052 (0.021, 0.13) and 0.49 (0.22, 0.97) mg peanut protein, respectively. Accounting for challenges with severe reactions at the LOAEL, by using the dose prior to the LOAEL as the new LOAEL, the ED01 drops to 0.029 (0.014, 0.074) mg peanut protein. Our results could aid in establishing improved food labeling guidelines in the management of food allergies.

1. Introduction

Food allergies constitute a significant public health issue and are an area where clear communication with the affected population is important. Potential exposure to food allergens in the United States (US) is currently communicated using precautionary allergen labels (PALs), such as "may contain allergen X" or "packaged in a facility that also processes allergen Y." Such labels do not include any information about the potential amount of the allergen in the food, and thus are not connected to any measure of risk (i.e., do not reflect whether exposure exceeds some potency-based dose metric). This means that such labels imply that the threshold for an allergic response is zero for all allergens for all individuals (since labeling based on the potential presence of any amount of allergen implies that any amount could be of concern). This labeling presents challenges to consumers, food manufacturers, and public health authorities, among others, in understanding and managing the potential risk of exposures in the food supply (DunnGalvin et al., 2015; NAS, 2017; Madsen et al., 2020).

Recent research, however, has indicated that minimal eliciting doses (EDs) can be identified for food allergens. The distribution of individual dose-response data can be used to develop a Reference Dose (RfD) (reviewed by Crevel et al., 2014; Madsen et al., 2020; NAS, 2017). Since the first use of dose distribution modeling for the response to allergenic foods by Bindslev-Jensen et al. (2002), a number of papers have been published on such modeling for a variety of populations and allergens (e.g., Blom et al., 2013; Blankestijn et al., 2017; Purington et al., 2018; Remington et al., 2020; Taylor et al., 2009, 2014).

The results of such modeling (Taylor et al., 2014) are incorporated into the Voluntary Incidental Trace Allergen Labelling (VITAL) program

* Corresponding author.
E-mail address: Lynne.Haber@uc.edu (L.T. Haber).
[1] Currently at Cardno ChemRisk

https://doi.org/10.1016/j.fct.2021.112125
Received 19 October 2020; Received in revised form 24 February 2021; Accepted 9 March 2021
Available online 12 March 2021

Abbreviations

BMA	Bayesian model averaging
CI	credible interval
CoFAR	Consortium for Food Allergy Research
DBPCFC	double-blind placebo-controlled food challenge
ED	eliciting dose
EPIT	epicutaneous immunotherapy
FDA	Food and Drug Administration
FEV1	forced expiratory volume in 1 s
GLRT	Generalized log-rank test
GMP	Good manufacturing practices
lb	lower bound
LOAEL	Lowest observed adverse effect level
LOO	leave one out
MCMC	Markov Chain Monte Carlo
NAS	National Academy of Sciences
NIH	National Institutes of Health
NOAEL	No observed adverse effect level
OFC	Oral food challenges
OIT	Oral immunotherapy
PAL	Precautionary allergen labels
RfD	Reference Dose
SPT	Skin prick test
ub	upper bound
US	United States
VITAL	Voluntary Incidental Trace Allergen Labelling

in Australia and New Zealand, which uses a risk-based approach for allergen labeling of food (Allergen Bureau, 2020). VITAL 2.0 established RfDs for 10 food allergens, and VITAL 3.0 expanded the list to 14 allergens (Allergen Bureau, 2019). VITAL RfDs have also been adopted by other countries, including Sweden (Sjogren Bolin, 2015) and Germany (Waiblinger and Schulze, 2018). In the US, the NAS recommended that a risk-based approach similar to VITAL be adopted to replace PALs systems (NAS, 2017). However, this recommendation has not yet been implemented, in part due to the limited modeling information from US populations, as well as incomplete details of the published modeling results (i.e., the data and methods used for modeling have not been shared in a manner that allows for independent verification of the results).

The purpose of the current paper is to estimate the dose-response distribution for peanut allergen, based on data from double-blind placebo-controlled food challenges (DBPCFCs) conducted during oral immunotherapy (OIT) or epicutaneous immunotherapy (EPIT) clinical trials in the US. Data were obtained from baseline DBPCFCs and DBPCFCs conducted on the placebo arm of the immunotherapy trials. The modeling approach accounted for variability between studies, geographic variability, and inter-individual variability. Intra-individual variability was addressed as a by-product of the probabilistic representation of the dose-response. We focus on the results with one allergen, peanut protein, so that closer consideration can be given to details of the modeling results and associated issues of interpretation. The study results provide insight into the dose-response for peanut allergen in the US population, and show the importance of accounting for multiple sources of variability. These results can be used to improve food labeling guidelines for peanut allergen, and thereby improve communication with allergic individuals.

2. Materials and methods

2.1. Data

2.1.1. Study design and population

Investigators conducting oral food challenges (OFC) with peanut protein as part of controlled clinical trials were identified with assistance from Marshall Plaut of the National Institutes of Health (NIH). All challenges were DBPCFCs. Three primary datasets were obtained from studies investigating potential peanut immunotherapies. One dataset was from a series of studies led by Kari Nadeau at Stanford University. The other two data sets were from the Consortium for Food Allergy Research (CoFAR), specifically CoFAR4 (Fleischer et al., 2013; Burks et al., 2015) and CoFAR 6 (Jones et al., 2017). Baseline challenge data, as well as data from post-therapy challenges of the placebo group, when applicable, were used in the peanut protein dose-response model (See Fig. 1.).

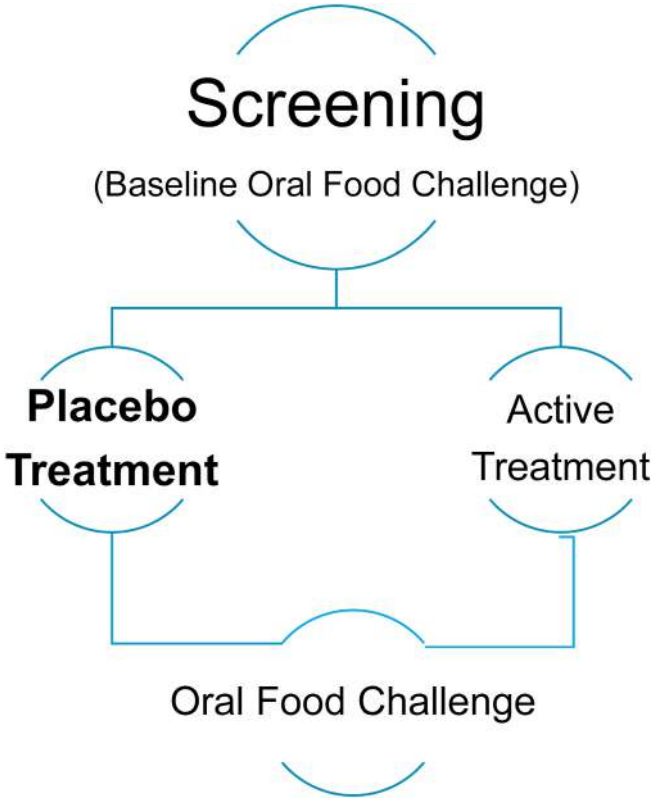

Fig. 1. General study design. All subjects who met the inclusion criteria underwent a DBPCFC at baseline. Some of the subjects (see Table 3 for details) in the placebo treatment arm underwent a second DBPCFC. Note that the placebo arm of the OIT refers to placebo therapeutic treatment, which is different from the placebo test by exposure to oat protein in the DBPCFC.

The test protein used by Stanford was FDA-standardized and validated GMP-grade peanut protein. Protein amounts were quantified based on protein gels and prepared in a GMP facility. For both CoFAR studies, commercially available peanut flour was used. Oat protein was used as a placebo for all studies, administered on an adjacent day (or potentially the same day, separated by at least 2 h, if there was no reaction in the first test), using the same protocol as for the peanut challenge. Peanut challenges were administered at 20-30-min intervals (CoFAR6) or 15 to 30-min intervals (CoFAR4, Stanford).

The Stanford data were compiled from a total of 16 different studies. Although some of these studies were multi-site studies, only data from

testing conducted at Stanford were included in the dataset. There was some minor variability in inclusion criteria across the studies; the universal inclusion criteria were: (1) skin prick test (SPT) average wheal >3 mm above negative control and history of clinical reaction to that food, or (2) peanut-specific IgE (sIgE) ≥4kU/L.

CoFAR4 and CoFAR6 were conducted at five different sites in the U. S. (New York, NY; Baltimore, MD; Little Rock, AR; Denver, CO; Durham, NC).[2] The inclusion criteria were: (1) convincing clinical history or physician's diagnosis of peanut allergy; (2) positive peanut SPT response (wheal diameter ≥3 mm, saline control corrected) and/or (3) sIgE ≥0.35 kU_A/L (CoFAR4), sIgE >0.35 kU_A/L (CoFAR6).

All studies required that subjects met the SPT and/or the sIgE criteria. Based on meeting these objective criteria, the subjects were considered "sensitized" and allergic. Therefore, the current analysis included *all* subjects who underwent baseline testing. That is, the analysis included some subjects who were excluded from the clinical trial due to their "passing" the baseline test by not reacting even at the highest dose tested; these subjects are in the set of right-censored observations.

Generally similar exclusion criteria were used for all three major datasets. The Stanford study excluded those individuals who had a prior history of an allergic reaction requiring intubation, or with associated hypotension. CoFAR4 and CoFAR6 excluded subjects with a history of intubation, or more broadly with a history of anaphylaxis to peanut, defined as involving hypoxia, hypotension, or neurologic compromise. All three studies also excluded subjects with asthma with FEV1 <80% of predicted value or clinical features of moderate or severe persistent asthma; and subjects with other significant nonallergic medical conditions. All three studies accepted patients who had experienced allergic reactions requiring epinephrine. Additional details on inclusion and exclusion criteria are provided in the Supplemental data.

2.1.2. Doses tested

Table 1 shows the doses tested in CoFAR4 and CoFAR6 at the initial baseline challenge, and in the repeat challenge in the immunotherapy placebo group. The majority of the doses tested in the Stanford studies are also shown in Table 1. However, there were some variations in the doses tested in the Stanford studies (and, to a lesser degree, in the CoFAR studies) beyond what is shown in Table 1, because the clinician had the discretion to repeat lower doses rather than escalating to a higher dose, if there was concern about a potential reaction, or to test higher doses if there was no reaction. The calculation of cumulative dose included all such repeated doses.

Unlike typical dose-response data, where separate groups are exposed to different doses, under the OFC study design, each person is tested with increasing doses, until there is one or more observable challenge-terminating reaction, or until the maximum protocol dose is reached. The dose prior to the challenge-terminating reaction is referred to as the individual no observed adverse effect level (NOAEL), and the dose at the terminating reaction is referred to as the individual lowest observed adverse effect level (LOAEL). Thus, the actual initial dose that would elicit a reaction is unknown but is bracketed by the NOAEL and LOAEL. Observations of this nature are referred to as censored. If the subject responded at the lowest dose tested, the NOAEL is 0, and the data are left-censored. If the highest dose tested elicited no response, the LOAEL is unbounded and the data are right-censored. All other cases are termed interval-censored. The statistical approaches applied account for all such censoring (left, right, and interval), as described below.

Table 1
Primary dosing schemes.

Study	Challenge	Doses (mg peanut protein)
CoFAR4	Baseline	Individual doses: 0.5, 2.5, 7.5, 25, 37.5, 50, 125, 250, 500 Possible cumulative doses: 0, 0.5, 3, 10.5, 35.5, 73, 123, 248, 498, 1000[a]
	Week 44 Repeat challenge	Individual doses: 0.5, 2.5, 7.5, 25, 37.5, 50, 125, 250, 500, 625, 875 Possible cumulative doses: 0, 0.5, 3, 10.5, 35.5, 73, 123, 248, 498, 1000[a], 1623, 2500[b]
CoFAR6[c]	Baseline	Individual doses: 1, 3, 10, 30, 100, 300, 600 Possible cumulative doses: 0, 1, 4, 14, 44, 144, 444, 1044
	Week 52 repeat challenge	Individual doses: 1, 3, 10, 30, 100, 300, 600, 1000, 3000 Possible cumulative doses: 0, 1, 4, 14, 44, 144, 444, 1044, 2044, 5044
Stanford	General design	Individual doses: 5, 20, 50, 100, 100, 100, 125, with possible additions of 400, 1000, 2500 Possible cumulative doses: 0, 5, 25, 75, 175, 275, 375, 500, and higher doses varying with the dosing pattern
	Alternative design	Individual doses: 0.1, 1.6, 6, 25, 50, 100, with possible additions of 400, 1000, 2500 Possible cumulative doses: 0.1, 1.7, 7.7, 33.1, 83.1, 183.1, and higher doses varying with the dosing pattern.

[a] Actually is 998 mg protein, but was sometimes rounded by the researchers to 1000 mg protein, and was treated as 1000 mg protein in the current analysis. There were also two subjects who received a cumulative dose of 5000 mg protein during the baseline challenge; they were not randomized into the study, but were included in the current analysis.

[b] Actually is 2498 mg protein, but was sometimes rounded by researchers to 2500 mg protein, and was treated as 2500 mg protein in the current analysis. There was also one subject who received a cumulative dose of 5000 mg protein.

[c] There was also one challenge with a cumulative dose of 2 mg protein and two challenges with a cumulative dose of 9 mg protein. In CoFAR6, one individual received a dose of 5044 mg protein.

2.1.3. Challenge termination criteria

For all three studies, the definition of a challenge-terminating reaction was based on defined criteria, with additional clinician discretion.[3] Challenges were stopped at the first sign of an objective symptom, or following significant or persistent subjective findings. This discretion is best illustrated with examples. If isolated vomiting were attributed to gagging, the challenge would not be terminated. Isolated transient mouth itch or isolated transient abdominal pain would not terminate dosing, but persistent mouth itch with mild abdominal pain (2 areas affected) could result in termination. Although abdominal pain is subjective, if it were severe and persistent with behavioral change such as crying or reduced activity, dosing could be terminated. Mild pruritus without objective rash would not terminate dosing but if combined with significant patient discomfort requiring treatment it would. Nausea alone is a subjective symptom that would not usually terminate a food challenge, but nausea with notable distress, for example, where the subject was clearly holding back from vomiting could be considered challenge-terminating, even without objective observation of vomit. Any reaction that required treatment, or where subject discomfort was significant enough that the subject refused the next dose, was also challenge-terminating.

[2] The specific sites were as follows: Department of Pediatrics, Mount Sinai School of Medicine, New York, New York; Department of Pediatrics, Johns Hopkins University School of Medicine, Baltimore, MD; Department of Pediatrics, University of Arkansas for Medical Sciences and Arkansas Children's Hospital, Little Rock Department of Pediatrics, Little Rock, Ark; National Jewish Health, Denver, Colorado; Department of Pediatrics, University of North Carolina, Chapel Hill.

[3] For CoFAR4, the stopping criterion was "first objective or significant/persistent subjective" responses. For CoFAR6, the criterion was "symptoms indicate a positive reaction."

2.1.4. Data preparation for modeling

All analyses were based on cumulative dose (peanut or placebo) administered over the timeframe of the challenge. Two different approaches were used to account for responses to the placebo challenge (oat protein). In the first approach, challenges in which there was a challenge-terminating reaction to placebo were removed regardless of the dose at which the response occurred or the severity of the response. In the second approach, the dose and severity of the response (based on the type of symptom(s) and severity of symptom(s), as characterized by the clinicians) were compared for the peanut challenge and the placebo challenge, in a manner similar to that of Westerhout et al. (2019). The goal of this approach was to remove only the challenges where the response could not be distinguished from the placebo response. Thus, the challenge was removed if the placebo response occurred at a dose lower than the dose at which the response to peanut was observed. In addition, if the response to placebo and peanut occurred at the same dose, then the challenge was removed only if the response to placebo was of the same or greater severity as the response to peanut.

In order to visualize intra-individual variability, the first and second challenge NOAELs were compared from individuals with multiple challenges. First and second challenge LOAELs were similarly compared graphically. Spearman's ranked correlation test was used to evaluate the association between paired NOAELs and between paired LOAELs.

While the placebo analysis and adjustments used the clinician's severity ratings for comparisons *within* a study (as described above), the severity rating approach of Zhu et al. (2015) was used for analyses that integrated across studies, specifically the association between LOAEL and severity (Fig. 6), and the "back one down" analysis described later in this paragraph. For these latter analyses, the Zhu et al. (2015) approach meant that the severity of each challenge-limiting response was defined based on both the clinical severity of individual reactions, as determined by the monitoring clinician, and the number of affected organ systems or regions of the respiratory tract. The Zhu approach categorizes allergic responses as mild, medium, or severe based on the nature of the reaction and the number of affected organ systems (among skin, gastrointestinal, upper respiratory, and lower respiratory). Due to concerns about potential for severe reactions at the challenge-terminating dose, a supplemental analysis was conducted based on the dose prior to the challenge-terminating dose for those challenges with severe reactions based on the Zhu rating system. Specifically, the subjects with a severe reaction at the LOAEL were identified; then this "back one down" analysis substituted the original NOAEL for the LOAEL, and identified the dose below the original NOAEL as the new NOAEL. The only exceptions were for three challenges where there was a severe reaction at the lowest discrete dose, and so it was not possible to identify a lower NOAEL, since the NOAEL was already left-censored. These challenges were retained for the "back one down" analysis using the original NOAEL and LOAEL. The other alternative would have been to remove these three observations for the "back one down" analysis, but they provided useful information regarding the potential for severe reactions at the lowest dose.

The data used for modeling are archived at https://osf.io/2vab4/; additional details of the data processing and cleaning steps are provided in the supplemental materials.

2.2. Modeling approach

2.2.1. Models

Based on the general design of OFCs, as described above, the data were considered to be analogous to the results of a failure analysis, where a sequence of stresses, or just time itself, is thought to incrementally increase the probability of a response. The models utilized for the analyses are based on well-known failure models (the inverses of survival models) or closely related functions (Wheeler et al., 2020). The five fully parametric models fit to the data were the Generalized Pareto, Log-Laplace, Logistic, Lognormal, and Weibull. See Table 2 for the

Table 2
Model equations with parameter priors.

Hierarchical Structure	Parameters (priors)	Transformed parameters	Model equation for Individual i
Generalized Pareto			
standard	l ~ C(0,1) b ~ G(2,1)	λ(i) = exp(-l)	$F_i(d) = 1 - (1 + \frac{d}{\lambda(i)})^{-b}$
studies	l(s) ~ C(0,1), s = 1,2,3 b ~ G(2,1)	λ(i) = exp(-l(s(i)))	
centers	l ~ C(0,5) sigmac ~ C^+(0,5) lc(c) ~ N(l, sigmac) b ~ G(2,1)	λ(i) = exp(-lc(c(i)))	
individuals-within-centers	l ~ C(0,5) sigmac ~ C^+(0,5) lc(c) ~ N(l, sigmac) sigmai ~ C^+(0,5) li(i) ~ N(lc(c(i)),sigmai) b ~ N(0,1)	λ(i) = exp(-li(i))	
Log-Laplace			
standard	l ~ C(0,1) b ~ N(0,1)	λ(i) = l β = exp(b)	$F_i(d) = 0.5exp(\frac{ln(d) - \lambda(i)}{\beta})$ *for* $ln(d) \leq \lambda(i)$ $F_i(d) = 1 - 0.5exp(\frac{\lambda(i) - ln(d)}{\beta})$ *for* $ln(d) > \lambda(i)$
studies	l(s) ~ C(0,1), s = 1,2,3 b ~ N(0,1)	λ(i) = l(s(i)) β = exp(b)	
centers	l ~ C(0,5) sigmac ~ C^+(0,5) lc(c) ~ N(l, sigmac) b ~ N(0,1)	λ(i) = lc(c(i)) β = exp(b)	
individuals-within-centers	l ~ C(0,5) sigmac ~ C^+(0,5) lc(c) ~ N(l, sigmac) sigmai ~ C^+(0,5) li(i) ~ N(lc(c(i)),sigmai) b ~ N(0,1)	λ(i) = li(i) β = exp(b)	
Logistic			
standard	l ~ C(0,1) b ~ N(0,1)	λ(i) = exp(l) β = exp(b)	$F_i(d) = 1 - \frac{1}{1 + \lambda(i) \cdot d^{\beta}}$
studies	l(s) ~ C(0,1), s = 1,2,3 b ~ N(0,1)	λ(i) = exp(l(s(i))) β = exp(b)	
centers	l ~ C(0,5) sigmac ~ C^+(0,5) lc(c) ~ N(l, sigmac) b ~ N(0,1)	λ(i) = exp(lc(c(i))) β = exp(b)	
individuals-within-centers	l ~ C(0,5) sigmac ~ C^+(0,5) lc(c) ~ N(l, sigmac) sigmai ~ C^+(0,5) li(i) ~ N(lc(c(i)),sigmai) b ~ N(0,1)	λ(i) = exp(li(i)) β = exp(b)	
Lognormal			
standard	l ~ C(0,1) b ~ G(1,1)	λ(i) = l	$F_i(d) = \Phi(\frac{ln(d) - \lambda(i)}{b})$
studies	l(s) ~ C(0,1), s = 1,2,3 b ~ G(1,1)	λ(i) = l(s(i))	
centers		λ(i) = lc(c(i))	

(*continued on next page*)

Table 2 (*continued*)

Hierarchical Structure	Parameters (priors)	Transformed parameters	Model equation for Individual i
	l ~ C(0,5) sigmac ~ C^+(0,5) lc(c) ~ N(l, sigmac) b ~ G(1,1)		
individuals-within-centers	l ~ C(0,5) sigmac ~ C^+(0,5) lc(c) ~ N(l, sigmac) sigmai ~ C^+(0,5) li(i) ~ N(lc(c(i)),sigmai) b ~ G(1,1)	λ(i) = li(i)	
Weibull			
standard	l ~ C(0,1) b ~ N(-0.66,1)	λ(i) = exp(-l) β = exp(b)	$F_i(d) = 1 - exp(-\lambda(i)\cdot d^{\beta})$
studies	l(s) ~ C(0,1), s = 1,2,3 b ~ N(-0.66,1)	λ(i) = exp(-l(s(i))) β = exp(b)	
centers	l ~ C(0,5) sigmac ~ C^+(0,5) lc(c) ~ N(l, sigmac) b ~ N(-0.66,1)	λ(i) = exp(-lc(c(i))) β = exp(b)	
individuals-within-centers	l ~ C(0,5) sigmac ~ C^+(0,5) lc(c) ~ N(l, sigmac) sigmai ~ C^+(0,5) li(i) ~ N(lc(c(i)),sigmai) b ~ N(-0.66,1)	λ(i) = exp(-li(i)) β = exp(b)	

s(i) = study for individual i; c(i) = center for individual i.
C(x,y) = Cauchy distribution centered at x with scale y.
C^+(x,y) = Half-Cauchy distribution centered at x with scale y (restricted to positive values).
G(α,β) = Gamma distribution with parameters α and β.
N(m,s) = Normal distribution with mean m and standard deviation s.

equations for each model. Each of these models has two parameters, which we will refer to as a location parameter (l, in Table 2) and a scale parameter (b, in Table 2).

2.2.2. Hierarchical structure

The data are grouped in terms of "studies" (i.e., CoFAR4, CoFAR6, and Stanford), "centers" within studies (the five study locations for CoFAR, and the16[4] studies within the Stanford dataset), and individuals within centers. The four hierarchical structures considered here were implemented in a stepwise manner, increasing complexity at each step.

We started with the simplest structure, referred to here as the "standard" structure, in which every observation is assumed to have been generated from the same underlying survival distribution (where the distribution is defined by one of the five models introduced above). This is not a hierarchical model *per se*.

Next, we considered a "study-level" structure in which observations within a study were generated from the same distribution, but the location parameter in each model under consideration varied from study to study. Strictly speaking, this was not a hierarchical model; the study-specific parameter estimates are independent of one another (as in a fixed-effects modeling approach).

The next step up included center-specific location parameters: every observation within a given center was assumed to be generated from the same distribution, but the location parameter varied from center to center. In this case, the center-specific parameter values were assumed to vary according to a normal distribution with a mean and standard deviation that were themselves estimated (so-called hyper-parameter). This is a fully hierarchical structure, akin to a random effects approach.

Finally, we considered a structure that hypothesized that individuals varied from one another with respect to their survival distributions, even within a given center. This "individuals-within-centers" structure had different location parameters for every individual. Those individual parameters were assumed to be normally distributed around center-specific means and, as in the center-level hierarchical structure, the center-specific means were assumed to vary around a "grand" mean. This structure therefore represents two levels of hierarchy.

Table 2 displays the inter-relationships among the parameters included in each of these hierarchical structures. Note that, regardless of structure, multiple observations from the same subject, if available, were always assumed to have been generated from the same survival distribution.

2.2.3. Bayesian implementation

We have adopted a Bayesian approach to model fitting and estimation for these analyses. In general, a Bayesian approach proceeds via the following steps. (1) Candidate data-generation processes are identified. In this analysis, we are considering 20 such processes: the five survival models coupled with the four hierarchical structures. (2) The parameters of those processes are assigned prior distributions. The prior distributions represent one's understanding of, or belief about, the likelihood of various values for those parameters, prior to the analysis of the data under consideration. In this way, one can incorporate knowledge of the system and understanding of the biological/toxicological situation. For the current modeling, fairly diffuse priors were used, to allow the data to drive the estimation more than any particular prior assumptions. (3) Finally, the likelihood of the data set under consideration is integrated with the first two steps. The parameter values that jointly tend to yield higher likelihoods for both the observations and with respect to their priors are identified. Most typically, the output consists of a distribution for the parameters, the posterior distribution. This posterior distribution is used to characterize the uncertainty in the parameter estimates and in the estimates derived from them. In this analysis, the posterior distributions were used to estimate the values of eliciting dose (ED), and the associated credible intervals (CIs, the Bayesian analog of confidence intervals) as a measure of uncertainty.

Table 2 displays the prior distributions used in this analysis, one set of priors for each of the 20 processes. For convenience (i.e., so as to make the specification of the priors independent of the actual range of dose values used in any particular analysis), those priors are expressed relative to doses that have a maximum value of 1; i.e., they are for "scaled doses." In order to use those priors here, all doses in the data set under consideration were divided by the maximum observed cumulative dose (5044 mg) when input into the analysis. ED estimates derived in the course of the estimation were subsequently multiplied by that maximum value to yield values on the original scale.

The priors listed in Table 2 are largely taken from Wheeler et al. (2020). There are two exceptions. Following Gelman et al. (2006), we have used half-Cauchy distributions for the standard deviation hyper-parameters (sigmai and sigmac) in all models. Additionally, to improve performance of the Weibull model with respect to convergence for the individuals-within-centers structure, the following changes were made. The scale parameters for all Cauchy (or half-Cauchy) distributions were increased to 5; the mean of the Normal distribution for model parameter b was changed from 0 to −0.66 and its standard deviation was increased to 1. For consistency, the corresponding changes were made to all the

[4] The final dataset for modeling was based on 15 studies ("centers") from Stanford, because all observations from one study were excluded as part of the placebo adjustment described in 2.1.4.

hierarchical structures for the Weibull model.

RStan (versions 2.19.2 and 2.21.2) was used to implement a Markov Chain Monte Carlo analysis to update those parameters and provide posterior distributions for them and for model predictions of interest (e.g., estimates of eliciting dose, ED, expressed as cumulative dose of peanut protein). Results are summarized here primarily in terms of the means and the 95% CIs for those predictions.

Model averaging was conducted largely according to Wheeler et al. (2020). The methods used are based on Bayesian model averaging methods (Fragoso et al., 2018). Whenever possible, the methods used are based on weights computed using the R package "LOO." That package computes approximate leave-one-out predictive probability estimates for each model. Two methods, "stacking" and "BMA+", were considered for use to compute model weights and to obtain model-averaged estimates (Vehtari et al., 2017; Yao et al., 2018).

In the case of the individuals-within-centers hierarchy, the LOO computations were problematic for estimating model weights, as further addressed in Allen et al. (submitted). In that case, alternative metrics were considered for model weighting. The choice of an alternative weighting was based on comparisons of the weighting results from the previous, simpler hierarchical structures. The results of those comparisons suggested that averaging based on mean log-likelihood across the samples from the MCMC simulation was a good stand-in for LOO-based averaging. Consequently, model predictions for the individuals-within-center structure were averaged based on the model-specific values of those mean log-likelihoods when applied to the individuals-within-centers structure. Additional supporting material is provided in Supplemental Table S-1.

An additional complication was noted in appropriately accounting for the variability in the location (l) parameter when calculating the ED estimates. For the analyses using a hierarchical structure (i.e., when center or individuals-within-centers levels were in the hierarchy), the "l" parameter varied by individual[5]; this parameter was distributed around the "pooled" "l" parameter representing the mean of the distribution. However, the ED estimates should be computed using the individual-specific "l" parameters, rather than using the pooled parameter value. This is because the probability of response is not a simple linear function of the "l" parameter, and so computing the probability of response using the average of the "l" parameter estimates is not the same as computing the probabilities of response from the individual-specific "l" parameter estimates and then averaging those probabilities.

Therefore, ED estimates were computed as follows. We pre-selected a vector of cumulative dose values that spanned the range of cumulative doses that were administered across all of the studies under consideration. For each of those cumulative dose values, we computed probabilities of response. For the standard and study-level structures, those probability calculations were based on the iteration-specific values of the model parameters, l or l(s), respectively, and b. For the center and individuals-within-center hierarchical structures, 500 randomly selected λ(i) parameter values were obtained according to the normal distributions shown in Table 2, again using the iteration-specific values of the model parameter l in conjunction with the iteration-specific values of sigmac and sigmai. Those λ(i) values were used with the iteration-specific b parameter value to compute, and average, the probabilities of response. Those averages represent the means of the probabilities over a random set of individuals in the general population and thus characterize the expected response at each of those doses. Specific ED estimates (i.e., ED01, ED05, and ED10) were obtained by linear interpolation using the doses with computed probabilities bracketing the desired probabilities (e.g., using the doses that yield probabilities just below and just above 0.01 to compute the ED01). As all such calculations are MCMC-iteration specific, the summary statistics reported here (mean and 95% credible intervals, as noted above) for the ED values were also obtained directly from the MCMC samples. Sensitivity analysis confirmed that averaging 500 random values was sufficient to yield estimates stable to at least 2 decimal places.

The model code is archived at https://osf.io/2vab4/.

3. Results

3.1. Data description and preparation

The total combined data set includes challenge data provided by CoFAR4, CoFAR6, and Stanford. The initial data set included 565 challenges of 498 subjects. The Stanford dataset is more than twice as large as the sum of the other two datasets, and includes a wider age range of test subjects, as well as a generally wider dose range. In total, the overall range of cumulative doses is 0.1–3625 mg peanut protein (with three subjects tested at doses up to 5000 mg protein and one tested up to 5044 mg protein).

Two approaches were used to account for responses to the placebo challenge (see 2.1.4), in addition to an analysis where all challenges with a response to placebo protein were included. When challenges in which there was a challenge-terminating reaction at any placebo dose were removed, the resulting dataset includes a total of 538 challenges of 473 subjects. When the approach of Westerhout et al. (2019) was used to remove only the challenges where the response cannot be distinguished from the placebo response, the resulting "placebo-adjusted" dataset includes 548 challenges from 481 subjects, including 67 repeat challenges[6] (summarized in Table 3).

A sensitivity analysis finds no meaningful difference in the ED values calculated when all placebo protein responders are included, or using either approach to account for challenges in which there was a reaction to the placebo protein (see Supplemental Table S-2). Since the placebo-adjusted approach represents the most refined approach toward considering placebo responses, the placebo-adjusted dataset was used for the remainder of the analyses presented here.

For the subjects who underwent two rounds of testing, Fig. 2 compares the NOAEL (dose prior to the challenge-terminating reaction) in the first and second challenges. This approach allows a direct comparison of multiple responses to testing within individuals and therefore illustrates intra-individual variability. Left-censored data points are not

Table 3
Summary of study data (placebo-adjusted).

	CoFAR4	CoFAR6	Stanford
Number of challenges	65	106	377
Number of repeat tests	19	22	26
Dose range (mg peanut protein cum doses, baseline/repeat test)	0.5–1000[a]/ 0.5–2500[b]	1 - 1044/ 1 - 5044	0.1–3625
Challenges with No Allergic Reaction	6	10	43
Challenges with Allergic Reaction	59	96	334
% of Challenges with Positive Reactions	90.8	90.6	88.6
Patient Demographics:			
Age range	12–37	4–20	0–54
Pre-pubescent[c] males	10	46	182
Post-pubescent males	20	7	34
Pre-pubescent[c] females	0	24	106
Post-pubescent females	16	7	27

[a] There were two additional challenges up to 5000 mg peanut protein.
[b] There was one additional challenge up to 5000 mg peanut protein.
[c] Pre-pubescent: <15 for males, <13 for females.

[5] In the case of the center-level hierarchy, the individual-specific "l" values are the same for individuals within the same center (see Table 2).

[6] One individual underwent three separate peanut challenges.

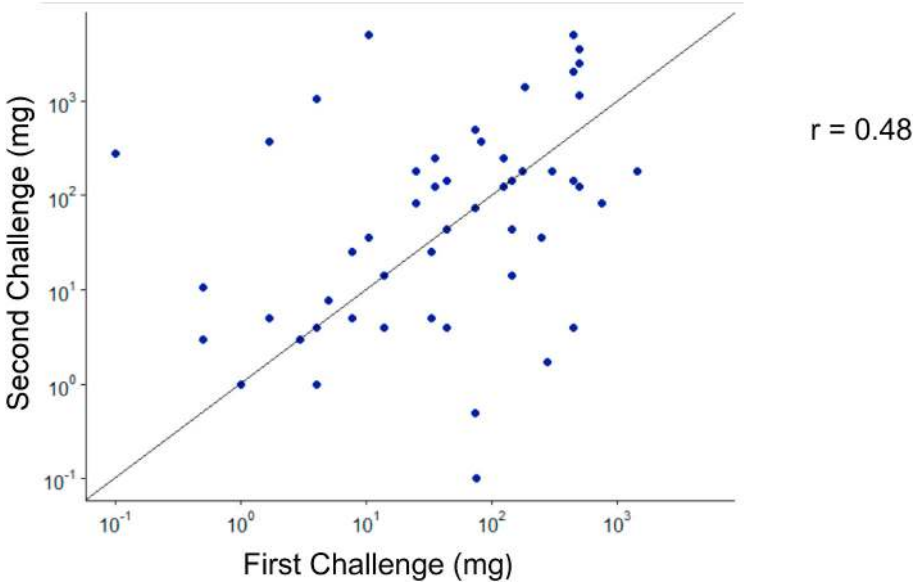

Fig. 2. Comparison of the NOAELs in the first and second challenges for subjects that were tested twice. Note that left-censored challenge results are not included in this display of the data, because there is no NOAEL for such results, although they were included in the dose-response analysis. Each point represents one individual in two challenges. The unity line is shown as a solid line. r = 0.48.

shown in this figure, because in these cases there is no NOAEL. As shown, there is a moderate correlation between the NOAEL in the first and second challenge (r = 0.48). However, there are some substantial outliers, by more than an order of magnitude. Similar numbers of subjects fall above and below the line of unity, indicating that there is no clear trend for the NOAEL in the first challenge to be higher or lower than the NOAEL in the second challenge. Similar results are obtained in comparing the LOAEL at the first and second challenge (r = 0.4, excluding the right-censored data; see Supplemental Figure S-1).

3.2. Bayesian hierarchical modeling results

A Bayesian hierarchical analysis was used to evaluate the dose-response. In and of itself, the Bayesian approach facilitates model fitting and parameter estimation through the specification of priors for potentially numerous parameters, i.e., when individual-level variability is incorporated. Moreover the priors specified can reflect and account for biological intuition, for example with respect to the expected change in response at lower doses, or the dose levels likely to yield any given probability of reaction. The hierarchical analysis allows for the modeling to consider the variability associated with various nested levels of the data organization. Finally, Bayesian model averaging (BMA) was applied as a way to take model uncertainty into account more appropriately than if one were to choose an individual model or models. The advantage of BMA is that poorly-fitting models and models that are less biologically plausible receive low weights. For the less complex hierarchical structures (i.e., those that did not include individual-level variability), we applied the two BMA approaches described in the Methods section (2.2.3). Those results are discussed further in Allen et al. (submitted). Because the focus here is on the best-choice individuals-within-centers structure, and, as noted in 2.2.3, LOO-based averaging was not considered appropriate for that structure, weights based on mean data log-likelihoods in the posterior sample were considered. On that basis, the Weibull model dominates so strongly (had a weight of essentially 1 while the other models had weights nearly equal to zero), that no averaging of model predictions was necessary.

The hierarchical analysis compared the model fit to the data at each successive level of the hierarchy, starting with the "standard" model (i.e., the model for which it is assumed that each individual tested has the same underlying probability of reaction at any given dose level; see 2.2.2). Elaborations of that "standard" model consisted of models for which the "location" parameter was allowed to vary in relation to some aspect of the challenges, as described in the rest of this paragraph. For example, the first level in the hierarchical structure assessed whether adding variability related solely to the three sources of the data (CoFAR4, CoFAR6, and Stanford) improves the model fits; it does not significantly do so. However, accounting for variability across "centers" (level 2) (i.e., the five study locations in CoFAR, and the 15 studies within the Stanford dataset, after making the placebo adjustment) does improve fits significantly (see Supplemental Table S-1). Furthermore, adding individual-level variability (level 3) "on top of" that center-level variability (the hierarchical structure referred to herein as "individuals-within-center") improves model performance even more (see Supplemental Table S-1). The individuals-within-center hierarchical structure is the primary focus of, and source of estimates presented in, this report.

Figs. 3–5 present the model-predicted probabilities of response, averaged over all tested individuals, for the individuals-within-center hierarchy. It is difficult, however, to compare these results with the observations (data) graphically. Because of the censored nature of the data (generally the ED for any individual challenge is only known to fall between certain bounds, the NOAEL and the LOAEL), the actual response at any given dose is not known. Therefore, Figs. 3–5 show the cumulative distributions of the NOAELs and of the LOAELs, as a way of bounding the actual (unknown) response.

As shown (Fig. 3), the best estimates for the predicted probabilities from the five individual models all fall between the NOAEL and LOAEL bounds. In fact, the model-predicted rates of elicitation as a function of dose (what we refer to as the "dose-response" relationship) are generally very similar across models, although the Weibull model differs a bit from the other models at the tails of the distribution. In particular, the Weibull model predicts slightly greater response at dose levels below about 2 mg peanut protein. Fig. 4 shows the results for the Weibull model alone, with its credible intervals also shown. The Weibull model is emphasized because, when considering model averaging, the weight calculations for the individuals-within-center level of hierarchy put essentially all the weight on that model, for the current dataset.

Fig. 5 shows the same observations and Weibull model predictions, but limited to the low-dose region and plotted on the natural scale for the explanatory variable (cumulative mg peanut protein), rather than the log scale used for the other figures. This change to the x axis was

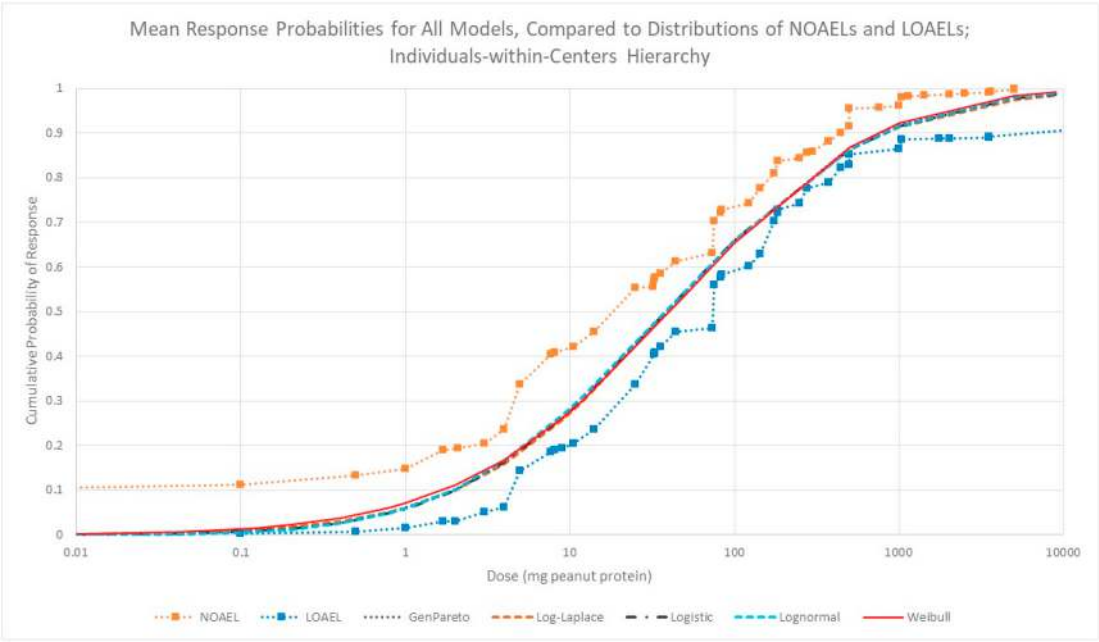

Fig. 3. Mean response probabilities for all models, compared to distributions of NOAELs and LOAELs; individuals-within-centers hierarchy. Results are shown for each model fit to the data. Note that the NOAEL and LOAEL data points represent the cumulative probability distribution at the corresponding dose. Note that the LOAEL corresponding to the NOAEL in a given challenge would be at a different dose (by definition), and the LOAELs corresponding to a given NOAEL may vary, depending on the exact dosing scheme (see Table 1). This means that the cumulative probability of the NOAELs at a given dose is not directly related to the cumulative probability for the LOAELs at the same dose.

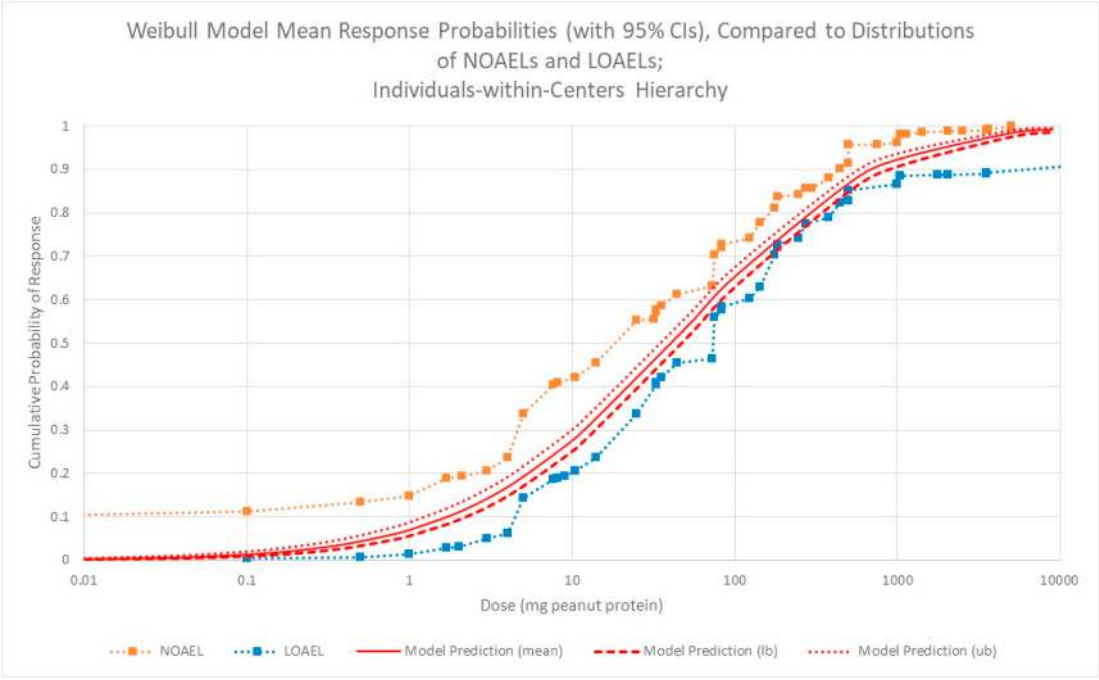

Fig. 4. Weibull model mean response probabilities (with 95% credible intervals), compared to distributions of NOAELs and LOAELs; individuals-within centers hierarchy. lb = lower bound; ub = upper bound.

made to emphasize that the shape of the curves (and observations) is actually moderately steep at low doses, contrary to the impression conveyed when the response probabilities are displayed relative to the log-transformed doses. The Weibull model predictions of elicitation rates are well within the cumulative distribution curves for the NOAELs and the LOAELs.

In an attempt to provide a refined graphical representation of the data for comparison with the model estimates, Fig. 5 also plots the cumulative distribution of the arithmetic average of the NOAEL and LOAEL for each challenge. This curve provides a rough approximation to the cumulative distribution of EDs, as opposed to bounds on the EDs. We say "rough approximation" because the actual location of the ED in each challenge is uncertain, as discussed above. As shown in Fig. 5, the average values are fairly consistent with the model predictions, although

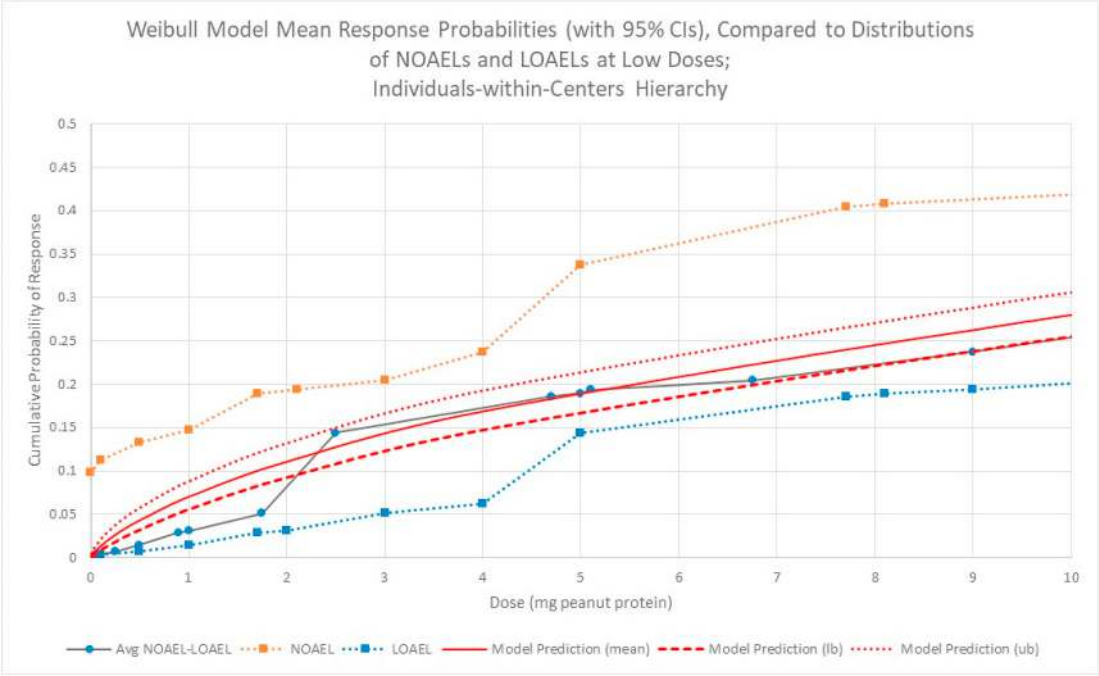

Fig. 5. Weibull model mean response probabilities (with 95% credible intervals), compared to distributions of NOAELs and LOAELs, low-dose region; individuals-within-centers hierarchy. lb = lower bound; ub = upper bound. Note that, unlike Figs. 3 and 4, the x axis is on a linear scale. The average values are computed as the average for each challenge, not as the average of the two cumulative distribution curves. Because NOAEL and LOAEL values were not "drawn" from the same set of possible values for each challenge (e.g., the dose intervals differed across studies), the cumulative distribution of the average does not fall exactly in the middle of the NOAEL and LOAEL curves.

the model predictions are slightly greater than the estimated averages at doses below about 2 mg. Aside from the fact that the curve position is uncertain, it should be noted that it is based on relatively few observations. With a total of 548 challenges, an ED05 corresponds to only about 27 people, and an ED01 to about 5 people. This is why the modeling is important, since predictions in the lower tail are subject to relatively large uncertainty when based on observations alone; small changes in the number of positive responses, or slightly different doses tested, could result in moving the observed averages that are below the lines to being within the credible limits.

The numerical estimates and credible intervals of the ED01, ED05 and ED10 for each individual model are shown in Table 4. Numerical estimates and credible estimates for additional ED values (analogous to the work of Houben et al., 2020) are in Supplemental Table S-3.

3.3. Considering severity

In considering what doses might be appropriate as exposure limits, it is important to consider both the probability of a response at any given dose, and the severity of the responses. In order to evaluate severity in a systematic manner, the integrated severity grading system developed by Zhu et al. (2015) was used to categorize the challenge-limiting responses, based on the severity ratings of the monitoring clinician, and on the number of affected organ systems or regions of the respiratory tract (see Section 2.1.4). Using this approach, the median doses for mild, moderate and severe responses for the combined data set are 25 mg, 44 mg, and 133 mg peanut protein, respectively (Fig. 6). The median doses for the mild and moderate categories are remarkably close to those reported by Zhu et al. (2015) of 25 and 50 mg, respectively. The median for the severe category differs by somewhat more, with Zhu et al. (2015) reporting a median of 250 mg.

As noted, the terminating dose (LOAEL) was based on a clinician evaluation of whether the reaction was sufficiently severe to stop the dose escalation. Almost a quarter of the challenges (121 out of 548 total challenges) are rated severe, although many of those (83 challenges) are considered severe based on the number of affected systems, rather than the severity of a single effect. Of the challenges that were rated as severe based on a single effect, almost all were based on wheezing (34 of 38 challenges) or airway obstruction (3 challenges); one severe challenge had mild tachycardia. Due to concerns about the potential for severe reactions, the analysis was repeated based on the dose prior to the challenge-terminating dose for the challenges with severe reaction ratings (the "back one down" dose). In other words, if there was a severe reaction at the LOAEL, the NOAEL for the main analysis became a LOAEL, and the dose below the original NOAEL became the new NOAEL.[7] As expected, the calculated ED values decrease for this analysis. For the model-of-choice (Weibull failure model run on the individuals-within-centers hierarchy), the EDx values change as indicated in Table 5.

3.4. Sensitivity analyses

Sensitivity analyses were also conducted to determine whether the EDx values varied by sex or age. There is no substantial[8] difference when the data are stratified by sex (Supplemental Table S-3). There is also no substantial difference when the data are stratified by age, either considering 18 and older as the breakpoint, or using sex-specific breakpoints (13 and older for females, 15 and older for males) (Supplemental Table S-4). The relatively small number of adults in the sample population (49 subjects, 55 challenges) is reflected in the wide confidence limits for the adult population ED values (Table S-3).

[7] For three challenges, there was a severe reaction at the lowest dose, and so it was not possible to identify a lower NOAEL. These challenges were retained for the "back one down" analysis using the original NOAEL and LOAEL.

[8] Although visual examination of the data suggests some apparent differences, the CIs overlap and the mean estimate for any one category is included within the CIs for the other categories.

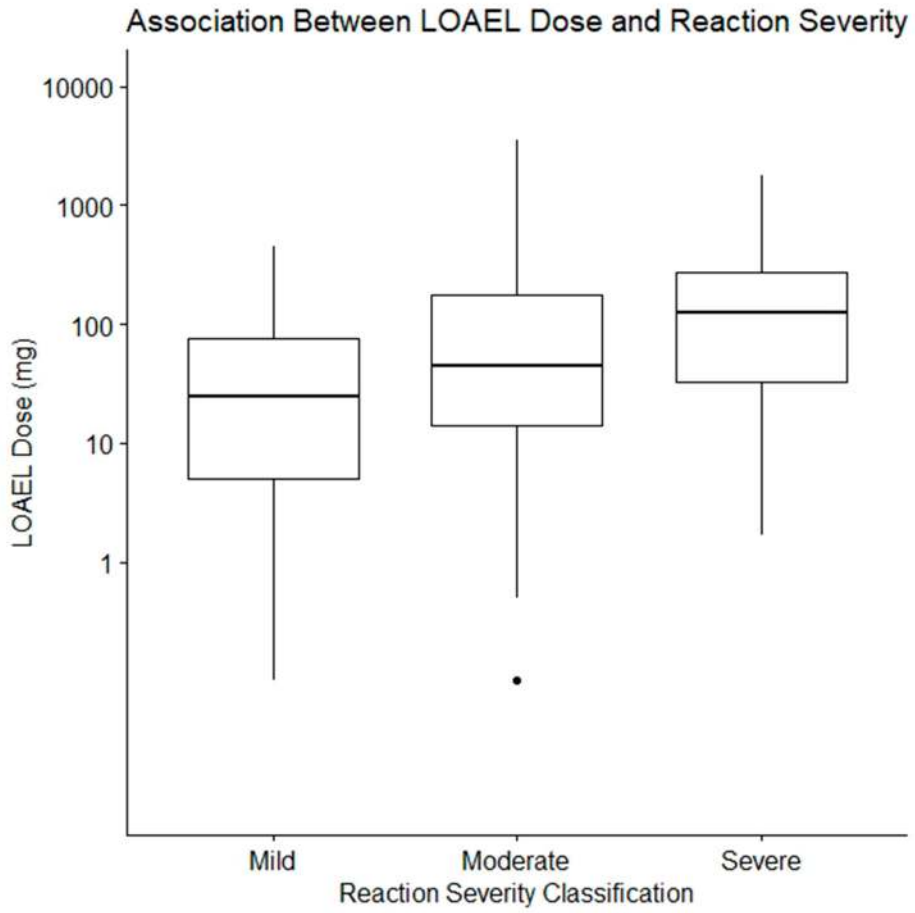

Fig. 6. Box and whiskers plot of the association between LOAEL (mg peanut protein) and reaction severity.

Table 4
ED estimates by model (means and 95% credible intervals).

Model Wts:	Failure Model				
	Generalized Pareto	Log-Laplace	Logistic	Lognormal	Weibull
	0	0	0	0	1
ED01	0.090 (0.040, 0.19)	0.075 (0.031, 0.18)	0.10 (0.05, 0.23)	0.12 (0.05, 0.27)	**0.052 (0.021, 0.13)**
ED05	0.62 (0.30, 1.19)	0.62 (0.30, 1.21)	0.66 (0.32, 1.27)	0.67 (0.33, 1.27)	**0.49 (0.22, 0.97)**
ED10	1.63 (0.80, 2.98)	1.66 (0.82, 3.05)	1.68 (0.84, 3.04)	1.65 (0.83, 2.95)	**1.44 (0.69, 2.70)**

Individuals-within-centers hierarchy. Note: Weibull model gets all the weight; its ED estimates are bolded to emphasize that.

Table 5
Sensitivity to severity of response. ED estimates for Weibull model (individuals-within-centers hierarchy) by treatment of severe responses (means and 95% credible intervals).

	Treatment of Severe Responses[a]	
	Standard	Back-one-Down
ED01	**0.052 (0.021, 0.13)**	0.029 (0.014, 0.074)
ED05	**0.49 (0.22, 0.96)**	0.29 (0.085, 0.54)
ED10	**1.43 (0.69, 2.69)**	0.88 (0.41, 1.75)

[a] Standard treatment leaves NOAEL and LOAEL as recorded by clinicians (basis for primary analyses). Back-one-down moves NOAELs and LOAELs to preceding doses if the response at the LOAEL was severe. Individuals-within-centers hierarchy.

4. Discussion

We conducted dose-response modeling on a dataset of DBPCFCs conducted in the US, including a total of 548 peanut challenges and 67 repeat tests, after adjusting for placebo responses. Testing was conducted using standardized approaches, and the dataset used came from four locations within the US, and so reflected some regional variability. Multiple sources of variability were considered in the analysis, including variability between studies (not significant), between sites or "centers" within a study (significant), and between individuals (significant). The impact of intra-individual variability was not a separate consideration in the modeling, but was reflected in the probabilistic nature of the dose-response analysis; even though each individual has an individual-specific dose-response curve, each challenge for the same individual is conditionally independent and could result in responses occurring at different doses, including falling between a different NOAEL-LOAEL

pair, as dictated by the probabilities defined by that dose-response curve. Additional strengths of the study included testing doses as low as 0.1 mg peanut protein, and testing a broad age range of subjects (0–54 in the Stanford dataset).

Like all studies of this sort, our study had several limitations. The number of challenges in adults was relatively small, limiting our ability to evaluate age-specific differences. Similarly, as discussed further below, there are only a few challenges with responses in the low-dose region. In addition, although the clinicians strove to be inclusive (e.g., by holding office hours on weekends, decreasing the need to take time off to come to the clinic), the studies were limited to those who came to allergy clinics. In general, the sampled population appears to be comparable to the general population with food allergies. We harmonized the categorization of symptoms of food challenges across the studies to optimize comparisons for our analyses.

A Bayesian hierarchical approach was used for the analysis. The Bayesian approach both facilitated the estimation of a large number of parameters, and allowed the use of parameter priors consistent with biological intuition. The hierarchical modeling was important for capturing key systematic sources of variability. For example, one important advance of the modeling conducted by Remington et al. (2020) was consideration of variability among studies. This was important because Taylor et al. (2010) observed a substantially different dose-response in the high-dose region for a diagnostic challenge trial in Nancy, France, compared with data gleaned from publications. We did not find similar significant differences among our data sources (CoFAR4, CoFAR6 and the Stanford data), but did find that model fit was significantly improved by considering the variability among centers and among subjects. Our approach also included Bayesian model averaging, to avoid having the results dependent on model choice, although all of the model weight ended up being on the Weibull model, and so formal model averaging was not conducted. This finding that all of the model weight was on the Weibull model is specific to the dataset analyzed here, and there is no expectation that this result should generally apply to modeling of the response to other allergens, or even other data sets of the response to peanut allergen (e.g., Taylor et al., 2014; Purington et al., 2018; Remington et al., 2020; Houben et al., 2020). We do note, however, that the Weibull model was chosen as the best-fitting model (based on the Akaike Information Criterion, AIC) by Purington and colleagues, in a deterministic evaluation of much of the Stanford data, and that the Weibull model received almost all of the weight in the Bayesian analysis of peanut response by Remington and colleagues. We also note in passing here that the modification of the prior for the Weibull model location parameter appears not to be the reason that model gets the preponderance of the weight. As discussed in detail in Allen et al. (submitted), little sensitivity to prior definition was evidenced and the data set is large enough to dominate over the prior specifications that were investigated (including the original one from Wheeler et al., 2020).

Consistent with results from other studies (e.g., Taylor et al., 2010), we observed a wide range of eliciting doses, reflecting the wide inter-individual variability in sensitivity. Indeed, even though all subjects were confirmed as sensitized to peanut protein, some subjects did not react even at the highest dose tested (i.e., there were 59 observations that were right-censored and included in the modeling, but had no LOAEL). Conversely, 54 (nearly 10%) of the challenges included in the modeling were left-censored (i.e., resulted in reactions even at the lowest dose tested, with no NOAEL).

An important contribution to the uncertainty in the ED values in Table 5 is the coarseness of the "grid" of tested doses in the low-dose range. Of the 54 left-censored challenges, 45 (8% of all challenges) had a LOAEL of 5 mg (which is substantially above the ED10); three of these 45 responses were rated "severe". The remaining nine challenges (nearly 2%) had LOAELs between 0.1 and 2 mg (with the variation in this lowest dose reflecting variability in study design). There were an additional eight challenges that did not elicit a response at 0.1 mg, but elicited responses at a cumulative dose of 1.7 mg. Overall, a total of 17 challenges (about 3% of all challenges, including both left-censored and interval-censored challenges) had a LOAEL less than 2 mg.

Despite these uncertainties, there was a generally good fit of all the models to the data. The Weibull model, which received all the weight in the model-averaging, appears to overpredict slightly the response below about 2 mg (Fig. 5). However, there are important considerations in interpreting the visual presentation of the data. Each challenge has its own NOAEL and LOAEL, and the actual threshold for that challenge is at an unknown point between the NOAEL and LOAEL. Fig. 5 approximates that threshold by taking the average of the NOAEL and LOAEL *for that challenge*. This average is different from the average cumulative frequencies of the NOAELs and LOAELs at a given dose (since the latter cumulative frequencies accumulate values that come from different challenges). This is why the data points for the average of the NOAEL and LOAEL do not fall directly in the middle between the NOAEL and LOAEL cumulative response curves. These considerations illustrate the importance of evaluations based on the mathematical fit to the data, and not solely visual fit. This is the first publication that we are aware of that provides detailed fit information for modeling of peanut allergen dose-response.

4.1. Comparison to other studies

The ED estimates returned by our analyses are somewhat lower than other recent modeling results. There are several potential reasons for this difference, including differences in the study populations, differences in the analytical approach, and differences in the actual dose-response data. We calculate an ED01 (95% credible interval) of 0.052 mg cumulative dose of peanut protein (0.021, 0.13) for the primary analysis and 0.029 mg (0.014, 0.074) for the back-one-down analysis that takes into account severe responses at the LOAEL. Remington et al. (2020) conducted an analogous analysis using a hierarchical Bayesian analysis with Bayesian model averaging, and calculated an ED01 of 0.7 (0.5, 1.3) based on cumulative dose and 0.2 (0.1, 0.4) based on discrete dose. Although our credible interval does not overlap with the ED01 credible interval for cumulative dose from the Remington analysis, it does overlap with the credible interval for their analysis based on discrete dose. VITAL 3.0 reported an ED01 and ED05 of 0.2 mg and 2.1 mg, based on thresholds for 1306 individuals (Allergen Bureau, 2019). The RfD derived in that analysis based on the ED01 was the same as that reported by VITAL 2.0 (Allen et al., 2014; Taylor et al., 2014), which was also based on an ED01, but considered multiple models and results in adults vs. children. In previous work by one of the collaborating groups on the current paper (KN) with a related dataset, an ED05 (and 95% confidence interval) of 0.49 mg (0.24, 0.73) was calculated based on cumulative dose (Purington et al., 2018); this value was comparable to our calculated ED05 of 0.49 mg (0.22, 0.96). This similarity between our results and those of Purington et al. (2018) suggests that the difference between our work and that of the Taylor group may be related to differences in the underlying dose-response data, rather than the mathematical analysis approach. However, we also note that calculating the ED estimates using the individual-specific "l" parameters (see Section 2.2.3), substantially reduced our calculated ED estimates compared to preliminary analyses using the pooled parameter value. If Remington et al. (2020) used pooled "l" parameter values, this could explain at least some of the difference.

There are a number of aspects to consider in comparing our results with those of previous evaluations. It is possible that the difference reflects differences between the US population and the European and Australian populations that form the basis for much of the VITAL database. However, it is noted that early introduction of peanut has been found to desensitize children, rather than sensitizing children (reviewed by Kusari et al., 2018; Agyemang and Sicherer, 2019), and peanut consumption is common in the US. Allen et al. (2014) reported that populations with peanut allergy from the United Kingdom, France, The Netherlands, and the United States were significantly different based on

the generalized log-rank test (GLRT). They did not find "substantial" differences in the ED05 estimates for the latter three countries (which fell between 2.0 and 4.0 mg of peanut protein), but it appears that the number of tested subjects from the US was relatively small. We note that part of the reason that our results are so similar to those of Purington et al. (2018) is because the dataset used by those authors comprises a large portion of our current dataset. However, approximately 30% of our dataset was not included in the Purington et al. (2018) analysis, and we did not find any significant impact of the source of data (CoFAR4, CoFAR6, or Stanford), indicating that the overall dose-response from the CoFAR studies is comparable to that from the Stanford dataset.

It is also possible that there are other ways that the populations we studied differ from those in the VITAL analyses, based on study design. Although the dose-response was based on testing at allergy clinics, we consider the test population similar to that of the general U.S. food allergic population, although this similarity has not been demonstrated. It is important to note that many individuals wanted to participate in the screening for clinical trials and this was based on the subjectivity of the parent, not a referral by a physician, and so the test subjects should not be assumed to be a sensitive subset of the allergic population. Houben et al. (2020) calculated that 500 mg of peanut protein (the highest dose tested for some of our data sets) is approximately the ED65 for the general population studies that they modeled. *If* our study had only included people who reacted at 500 mg and below, then we would have included only 65% of the allergic population, based on the Houben study. However, that was not the case. We included everyone in our analysis who met the criteria based on SPT and IgE, even if they did not react at the high dose in the baseline food challenge, and so this would not have been a reason that we captured a sensitive portion of the population. Like other DBPCFC studies used for dose-response modeling, the studies analyzed here excluded people with a history of severe anaphylaxis (involving hypoxia, hypotension or neurological compromise). However, note that this criterion is based on the severity of *response*, which is, at most, only partially related to the dose-response, since the magnitude of the dose that precipitated the severe anaphylaxis is not known. Furthermore, since this is a standard exclusion, it would not lead to differences from other modeling studies.

With regard to the analytic approach, our comparison focuses on a comparison with the Bayesian analysis of Remington et al. (2020). However, a direct comparison is not possible. Their data set included the "study" level of hierarchy, but it appears that there were not additional substudies ("centers") in their study design. In our analysis, "study" was not significant, but variability among "centers" alone was significant in our analysis. In addition, individual-level variability was not part of the Remington study analysis. The analyses upon which the current estimates are based built on the code supplied by a coauthor of that study (M. Wheeler), but that code was extended so as to include modeling of the variation from one individual to another. We believe consideration of inter-individual variability in the modeling is important for the derivation of health-protective standards or labeling, since we expect each individual to have their own dose-response curve, with potentially different curve shapes and potencies. However, in preliminary modeling analyses, inclusion of the individuals within centers level of the hierarchy did not necessarily result in lower ED values compared to the centers-alone level of hierarchy for individual models. Because the individuals-within-centers was preferred over the centers-level hierarchy, the centers-level results were not further pursued or carried to the level of evaluating model weighting, and so those model results are not presented here.

There are also some differences between our work and previous studies in the responses modeled. We modeled cumulative dose based on the NOAEL and LOAEL for the clinician-determined challenge-terminating reaction, where these reactions were either objective symptoms or *significant or persistent* subjective findings. As described in detail by Westerhout et al. (2019), even objective symptoms may be intermittent or transient. While Westerhout and colleagues developed a systematic approach toward identifying NOAELs and LOAELs for such datasets, the observation of an intermittent response suggests that different results could be obtained on repeat testing. The primary purpose of the data used for our analyses was in support of the development of OIT or EPIT, rather than being primarily for dose-response analyses. Thus, although consistent stopping criteria were used, clinician judgement was also involved. Our analyses find significant center-to-center variability. To the extent that this variability reflects, at least in part, the effect of different practices/judgments of clinicians across those centers, there may be some suggestion that those practices/judgments affect determinations of eliciting doses. The preferred hierarchy for the current analysis was individuals-within-centers, meaning that any such center-to-center differences have been captured in the modeling.

In summary, it appears that the differences between our results and those of Remington et al. (2020) are related to differences in the underlying dose-response data, although methodological differences are also possible. The difference in the data is less likely to be due to differences in the sampled population, but may relate to clinician judgement, or to specifics of the study design, including the doses tested or dose spacing. Without specific details on the underlying dose-response data analyzed by Remington et al. (2020), further evaluation of the differences is not possible. It is also unclear based on the available documentation whether Remington et al. (2020) used only the center of the distribution of what for them would have been study-specific location parameters, or whether they considered the full variation of that parameter among individuals, as we have done here.

4.2. Repeat challenges

With regard to repeat challenges, although we find that results of the repeat tests are moderately correlated with the initial testing, several of the repeat test NOAELs differ by an order of magnitude or more from the original test, and several differ by several orders of magnitude (Fig. 2). We find no trend for lower NOAELs or LOAELs in the initial or repeat testing, consistent with the results of Purington et al. (2018).[9] The Purington study also found a significant increase in severity on the repeat trial, a relationship that we did not evaluate. The high intra-individual variability is also consistent with previous reports, which noted that allergen reactivity might be affected by circadian, menstrual, and other biological cycles; activity (including exercise); infections; alcohol status; medications; sleep deprivation and many other factors (Crevel et al., 2014; NAS, 2017; Dua et al., 2019). In one of the few studies quantitating such differences, Dua et al. (2019) noted that exercise and sleep deprivation both lowered the mean threshold for response. Opinions in the literature differ as to whether such response modifiers should be accounted for in the RfD (Dua et al., 2019) or instead in clinician advice to individual patients (Crevel et al., 2014).

4.3. Use of cumulative dose

The analyses in this report were all conducted based on cumulative dose. This was based on the observation that cumulative dose is more stable than discrete dose, and more biologically meaningful. Taylor et al. (2014) calculated RfDs for a number of different food allergens and noted that ED values based on discrete dose are generally somewhat lower than those based on cumulative dose, but overall the differences were small. Small differences between ED values based on cumulative and discrete doses were also noted by Westerhout et al. (2019), although Remington et al. (2020) noted somewhat larger differences. Biologically, higher ED values might be calculated based on cumulative dose if

[9] The conclusions of Purington et al. (2018) are based on evaluation of a total of a total of 21 positive repeat challenges, including 16 to peanut. Our dataset included 67 repeat challenges, including both the Stanford dataset (overlapping with the Purington data) and repeat challenges from CoFAR4 and CoFAR6.

the full reaction is manifested after the time to the next dose increment. Blumchen et al. (2014) found a mean latency to reaction of almost an hour, but our investigators have found that >80% of reactions occur within 15–20 min of dosing. NAS (2017) suggested that the dose escalation approach could result in some desensitization to the allergen, and thus higher thresholds based on cumulative than discrete doses, but Hourihane et al. (2017) observed fewer subjects than predicted reacting after a single dose at the ED05 for peanut protein, indicating that the ramp-up design may not result in higher ED values.

4.4. Considering severity of response

An important consideration in setting RfDs is the severity of response. Considering this issue from a risk perspective, there are several overlapping issues that need to be teased apart. The first is the relationship between a prior history of severe responses and the observed dose-response. Results have been mixed regarding whether there is a relationship between this prior history and the reactive dose (Wensing et al., 2002; Hourihane et al., 2005; Taylor et al., 2010; Eller et al., 2012). However, as noted, this consideration of history does not explicitly address the dose that caused the severe reaction. Without information on the size of that dose, it is not surprising that there are inconsistencies in the literature regarding the correlation with the reactive dose in controlled studies.

A second issue is whether higher doses, *on average*, result in more severe responses, where again, results have been reported to be mixed (reviewed by Dubois et al., 2018). Evaluation of this issue is complicated by the typical design of DBPCFCs, which usually terminate dosing when objective or significant subjective responses occur. Thus, even though it would be expected that *an individual's* response would get more severe with increasing dose, such higher-dose data (for responses above the eliciting dose) are often not available. Given the wide variability in individual EDs, it is not surprising that severe reactions can occur at any dose, but this does not address the question of whether severity *on average* increases with dose. In a rare study that continued dosing above the usual cutoffs, Wainstein et al. (2010) continued dosing at higher levels beyond those causing mild reactions, and observed anaphylaxis in a number of subjects. The nature of the response at the eliciting dose depends on both the shape of an individual's dose-response-severity curve, and the spacing between doses. Because log-dose spacing is common, this means that the increment between successive doses is typically larger at higher doses, which could also be a reason for more severe responses at higher doses. That is, this larger increment means that there is a larger dose interval between the higher doses, and so there is a larger dose range where the response can transition from no response to mild response to a severe response. This means that observing a severe response at the LOAEL at these higher doses does not necessarily mean that a subject's first response as dose increases would be severe if there were additional intermediate doses.

Similarly to the results of Zhu et al. (2015), we find that the median LOAELs increase with the severity of response, although there is substantial overlap across categories, consistent with the issues noted related to study design. This is also consistent with the results of Pettersson et al. (2018), who developed a multiple linear regression model that predicted severity in a DBPCFC based on lower ED, along with age, SPT ratio, sigE, and a shorter reaction time in the DBPCFC. Pettersson et al. (2018) found that their model explained 23.5% of the severity, and ED only contributed 4.4% to the explained variance, perhaps due to the inclusion of biological predictors of response (SPT, sigE) and the design issues noted above. Sicherer et al. (2000) reported a weak correlation to no correlation between reactive dose and severity of reaction, and Wensing et al. (2002) reported an inverse correlation. Ballmer-Weber et al. (2015) found that subjective responses typically occurred at lower doses than objective responses, but did not further evaluate dose-response by severity.

A third issue related to severity is the severity of response for those sensitive individuals with low eliciting doses. This relates to the concern about whether the most sensitive allergic individuals are protected at an RfD, and if there were a reaction at the RfD, whether it would be severe. In order to evaluate the protectiveness of RfDs determined in DBPCFCs, Hourihane et al. (2017) treated 381 children at an allergy clinic with a single dose of the previously-calculated ED05 of 1.5 mg peanut protein (Allen et al., 2014; Taylor et al., 2014). Subjects in the single dose study were included regardless of the previous reaction severity, unlike some of the studies used in calculating the ED05, which have excluded people who had experienced severe anaphylaxis. Only 2.1% of the subjects (CI 0.6–3.4%) had an objective reaction (compared with a prediction of 5%), and all reactions were classified as mild. The study authors interpreted the results as indicating that any reactions at the ED01 of 0.2 mg (i.e., the RfD) would be mild. In our study, almost a quarter of the subjects had a severe reaction at the LOAEL, but these reactions were generally at relatively high doses compared to our ED10 (i.e., >5 mg). Of seven subjects in our study with severe reactions at ≤5 mg, four had NOAELs of 0.1–1 mg. The other three subjects had severe reactions at the lowest dose tested, but that dose was 5 mg for all three subjects. To address the concern about a potential for a severe response at the challenge-terminating dose, we also modeled the data for the subjects with a severe response by shifting down one dose group for these subjects, so that the prior NOAEL became a LOAEL. This approach decreases the ED01 to 0.029 mg.

5. Conclusion

Our analysis was based on three datasets that were collected as part of the placebo arms of OIT/EPIT trials, using standardized approaches, at different geographical locations in the US. Our modeling accounted for variability across studies, centers, and individuals, as well as model uncertainty. Our ED01 value of 0.052 mg peanut protein calculated in the primary analysis is somewhat lower than those that were the basis for the VITAL RfD for peanut protein, most likely reflecting differences in the underlying dose-response data or in details of the modeling approach. We have noted a number of uncertainties in the model results. Many of these uncertainties are likely to apply to other allergen dose-response modeling, but are more apparent here because of the level of detail provided for the input data and modeling methods. The underlying data that we modeled and the model code are also publicly available, in order to allow others to reproduce our results. Our analysis with a US population enrolled in OIT/EPIT trials provides an improved approach and benchmark for the development of labeling guidelines in the US. We believe that this additional level of detail and transparency, as well as accounting for multiple sources of variability will help to provide additional confidence in the health-protectiveness of the modeling results, and result in improved management of food allergy.

CRediT authorship contribution statement

Lynne T. Haber: Conceptualization, Funding acquisition, Methodology, Project administration, Supervision, Writing – original draft, Writing – review & editing. **John F. Reichard:** Conceptualization, Data curation, Formal analysis, Methodology, Validation, Visualization, Writing – review & editing. **Alice K. Henning:** Conceptualization, Data curation, Writing – review & editing. **Peter Dawson:** Data curation, Writing – review & editing. **R. Sharon Chinthrajah:** Investigation. **Sayantani B. Sindher:** Investigation. **Andrew Long:** Investigation. **Melissa J. Vincent:** Conceptualization, Data curation, Writing – review & editing. **Kari C. Nadeau:** Conceptualization, Investigation, Writing – review & editing. **Bruce C. Allen:** Conceptualization, Formal analysis, Methodology, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: The authors declare the following financial interests/personal relationships which may be considered as potential competing interests. Funding for the contributions of LH, JR, AH, PD, MV, and BA was provided by the Institute for the Advancement of Food and Nutrition Sciences (IAFNS) through an ILSI North America Food and Chemical Safety Committee grant. In 2021 ILSI North America tranformed into IAFNS. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing policies or endorsement of the sponsor or the authors' employers. The sponsor was not involved in the study design, collection, analysis, interpretation, writing, or decision to submit and where to submit for publication. The author team provided progress briefings to the sponsoring organization on the approach and interim results, but neither the final results nor a draft of the manuscript was shared with the sponsors prior to submittal. The sponsors were allowed to ask clarifying questions at the briefings, but not to comment on the methods or results. JR, AH, PD, and BA report no other known competing interests, and AL reports no known competing interests.

LH reports recent projects with Specialised Nutrition Europe (SNE) (another organization with food-related interests) and the Flavor and Extract Manufacturers Association (FEMA), and is a subcontractor on a separate IAFNS project related to food risk assessment.

SC reports grants from NIAID, CoFAR, Aimmune, DBV Technologies, Astellas, and Regeneron; is an advisory member for Alladapt, Genentech, Novartis, and receives personal fees from Before Brands.

SS receives grant support from NIH. She is involved in clinical trials with Regeneron, Aimmune Therapeutics, DBV Technologies, Adare Pharmaceuticals, Sanofi, and Novartis.

KN reports grants from National Institute of Allergy and Infectious Diseases (NIAID), National Heart, Lung, and Blood Institute (NHLBI), and National Institute of Environmental Health Sciences (NIEHS); Food Allergy Research & Education (FARE), Director of World Allergy Organization (WAO) Center of Excellence at Stanford; Advisor at Cour Pharma; Co-founder of Before Brands, Alladapt, Latitude, and IgGenix; National Scientific Committee member at Immune Tolerance Network (ITN) and National Institutes of Health (NIH) clinical research centers; DSMB member for NHLBI, US patents for basophil testing, multifood immunotherapy and prevention, monoclonal antibody from plasmoblasts, and device for diagnostics.

Acknowledgements

This work was supported by the Institute for the Advancement of Food and Nutrition Sciences (IAFNS), USA, through Food and Chemical Safety Committee contract NA-UCINCINNATI-20171018. IAFNS is a nonprofit science organization that pools funding from industry collaborators and advances science through the in-kind and financial contributions from public and private sector participants. IAFNS had no role in the design, analysis, interpretation, or presentation of the data and results.

We are grateful to Matt Wheeler for sharing his model code and for pointing us to the preprint of the paper on arXiv. We are grateful to Stefano Luccioli and colleagues at the US FDA for the opportunity to discuss aspects of the study design and preliminary results with them. Andy Maier also provided useful ideas during early phases of this project. Scott Sicherer provided useful insights regarding clinician determination of the stopping dose. Brian Palmer, Brendan Hanson, and Trevor Hay provided useful technical assistance. Marepalli B. Rao shared his expertise on Bayesian modeling. We are also grateful for the additional insights provoked by questions from the peer reviewers.

We acknowledge the receipt of deidentified data provided through collaboration with the Consortium for Food Allergy Research (CoFAR), involving 2 NIAID-sponsored clinical trials supported by the National Institutes of Health (NIH)/National Institute of Allergy and Infectious Diseases (NIAID) grants U19AI066738 and U01AI066560. The CoFAR studies were also supported by NIH/National Center for Advancing Translational Sciences (NCATS) grant nos. UL1 TR0001082 (Colorado), UL1 TR000067 (Mount Sinai), UL1 TR000039 (Arkansas), UL1 TR000083 (North Carolina), and UL1 TR000424 (Johns Hopkins) from the National Center for Research Resources (NCRR), a component of the NIH and partial support for one of the trials was obtained from DBV Technologies (Montrouge, France) provided to the Consortium of Food Allergy Research (CoFAR). We acknowledge all those research subjects and personnel contributing to the CoFAR studies, including the investigators: Stacie M Jones, David M Fleischer, Scott H Sicherer, A Wesley Burks, Donald Y M Leung, Robert W Lindblad, Peter Dawson, Alice K Henning, Andrew H Liu, Donald Stablein, Lloyd Mayer, Brian P Vickery, Amy M Scurlock, M Cecilia Berin, David Chiang, Brian P Vickery, Robbie D Pesek, Christine B Cho, Wendy F Davidson, Marshall Plaut, Robert A Wood and Hugh A Sampson.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.fct.2021.112125.

References

Agyemang, A., Sicherer, S., 2019. The importance of early peanut ingestion in the prevention of peanut allergy. Expet Rev. Clin. Immunol. 15 (5), 487–495.

Allen, K.J., Remington, B.C., Baumert, J.L., Crevel, R.W., Houben, G.F., Brooke-Taylor, S., Kruizinga, A.G., Taylor, S.L., 2014. Allergen reference doses for precautionary labeling (VITAL 2.0): clinical implications. J. Allergy Clin. Immunol. 133 (1), 156–164.

Allergen Bureau, 2020. The VITAL® program. http://allergenbureau.net/vital/. (Accessed 25 August 2020).

Allergen Bureau, 2019. Summary of the 2019 VITAL scientific expert panel recommendations. http://allergenbureau.net/wp-content/uploads/2019/09/VSEP-2019-Summary-Recommendations_FINAL_Sept2019.pdf. (Accessed 21 August 2020).

Ballmer-Weber, B.K., Fernandez-Rivas, M., Beyer, K., Defernez, M., Sperrin, M., Mackie, A.R., Salt, L.J., Hourihane, J.O.B., Asero, R., Belohlavkova, S., Kowalski, M., 2015. How much is too much? Threshold dose distributions for 5 food allergens. J. Allergy Clin. Immunol. 135 (4), 964–971.

Blumchen, K., Beder, A., Beschorner, J., Ahrens, F., Gruebl, A., Hamelmann, E., Hansen, G., Heinzmann, A., Nemat, K., Niggemann, B., Wahn, U., 2014. Modified oral food challenge used with sensitization biomarkers provides more real-life clinical thresholds for peanut allergy. J. Allergy Clin. Immunol. 134 (2), 390–398.

Bindslev-Jensen, C., Briggs, D., Osterballe, M., 2002. Can we determine a threshold level for allergenic foods by statistical analysis of published data in the literature? Allergy 57 (8), 741–746.

Blankestijn, M.A., Remington, B.C., Houben, G.F., Baumert, J.L., Knulst, A.C., Blom, W. M., Klemans, R.J., Taylor, S.L., 2017. Threshold dose distribution in walnut allergy. J. Allergy Clin. Immunol.: In Pract. 5 (2), 376–380.

Blom, W.M., Vlieg-Boerstra, B.J., Kruizinga, A.G., van der Heide, S., Houben, G.F., Dubois, A.E., 2013. Threshold dose distributions for 5 major allergenic foods in children. J. Allergy Clin. Immunol. 131 (1), 172–179.

Burks, A.W., Wood, R.A., Jones, S.M., Sicherer, S.H., Fleischer, D.M., Scurlock, A.M., Vickery, B.P., Liu, A.H., Henning, A.K., Lindblad, R., Dawson, P., 2015. Sublingual immunotherapy for peanut allergy: long-term follow-up of a randomized multicenter trial. J. Allergy Clin. Immunol. 135 (5), 1240–1248.

Crevel, R.W., Baumert, J.L., Baka, A., Houben, G.F., Knulst, A.C., Kruizinga, A.G., Luccioli, S., Taylor, S.L., Madsen, C.B., 2014. Development and evolution of risk assessment for food allergens. Food Chem. Toxicol. 67, 262–276.

Dua, S., Ruiz-Garcia, M., Bond, S., Durham, S.R., Kimber, I., Mills, C., Roberts, G., Skypala, I., Wason, J., Ewan, P., Boyle, R., 2019. The effect of sleep deprivation and exercise on reaction threshold in peanut-allergic adults: a randomized controlled study. J. Allergy Clin. Immunol. 144 (6), 1584–1594.

Dubois, A., Turner, P.J., Hourihane, J., Ballmer-Weber, B., Beyer, K., Chan, C.H., Gowland, M.H., O'Hagan, S., Regent, L., Remington, B., Schnadt, S., Stroheker, T., Crevel, R., 2018. How does dose impact on the severity of food-induced allergic reactions, and can this improve risk assessment for allergenic foods?: report from an ILSI Europe Food Allergy Task Force Expert Group and Workshop. Allergy 73 (7), 1383–1392. https://doi.org/10.1111/all.13405.

DunnGalvin, A., Chan, C.H., Crevel, R., Grimshaw, K., Poms, R., Schnadt, S., Taylor, S.L., Turner, P., Allen, K.J., Austin, M., Baka, A., 2015. Precautionary allergen labelling: perspectives from key stakeholder groups. Allergy 70 (9), 1039–1051.

Eller, E., Hansen, T.K., Bindslev-Jensen, C., 2012. Clinical thresholds to egg, hazelnut, milk and peanut: results from a single-center study using standardized challenges. Ann. Allergy Asthma Immunol. 108 (5), 332–336.

Fleischer, D.M., Burks, A.W., Vickery, B.P., Scurlock, A.M., Wood, R.A., Jones, S.M., Sicherer, S.H., Liu, A.H., Stablein, D., Henning, A.K., Mayer, L., 2013. Sublingual immunotherapy for peanut allergy: a randomized, double-blind, placebo-controlled multicenter trial. J. Allergy Clin. Immunol. 131 (1), 119–127.

Fragoso, T.M., Bertoli, W., Louzada, F., 2018. Bayesian model averaging: a systematic review and conceptual classification. Int. Stat. Rev. 86 (1), 1–28.

Gelman, A., 2006. Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). Bayesian analysis 1 (3), 515–534.

Houben, G.F., Baumert, J.L., Blom, W.M., Kruizinga, A.G., Meima, M.Y., Remington, B. C., Wheeler, M.W., Westerhout, J., Taylor, S.L., 2020. Full range of population Eliciting Dose values for 14 priority allergenic foods and recommendations for use in risk characterization. Food Chem. Toxicol. 146, 111831. https://doi.org/10.1016/j.fct.2020.111831.

Hourihane, J.O.B., Allen, K.J., Shreffler, W.G., Dunngalvin, G., Nordlee, J.A., Zurzolo, G. A., Dunngalvin, A., Gurrin, L.C., Baumert, J.L., Taylor, S.L., 2017. Peanut Allergen Threshold Study (PATS): novel single-dose oral food challenge study to validate eliciting doses in children with peanut allergy. J. Allergy Clin. Immunol. 139 (5), 1583–1590.

Hourihane, J.O.B., Grimshaw, K.E.C., Lewis, S.A., Briggs, R.A., Trewin, J.B., King, R.M., Kilburn, S.A., Warner, J.O., 2005. Does severity of low-dose, double-blind, placebo-controlled food challenges reflect severity of allergic reactions to peanut in the community? Clin. Exp. Allergy 35 (9), 1227–1233.

Jones, S.M., Sicherer, S.H., Burks, A.W., Leung, D.Y., Lindblad, R.W., Dawson, P., Henning, A.K., Berin, M.C., Chiang, D., Vickery, B.P., Pesek, R.D., 2017. Epicutaneous immunotherapy for the treatment of peanut allergy in children and young adults. J. Allergy Clin. Immunol. 139 (4), 1242–1252.

Kusari, A., Han, A., Eichenfield, L., 2018. Recent advances in understanding and preventing peanut and tree nut hypersensitivity. F1000Research 7, 1716. https://doi.org/10.12688/f1000research.14450.1. F1000 Faculty Rev.

Madsen, C.B., van den Dungen, M.W., Cochrane, S., Houben, G.F., Knibb, R.C., Knulst, A. C., Ronsmans, S., Yarham, R.A., Schnadt, S., Turner, P.J., Baumert, J., 2020. Can we define a level of protection for allergic consumers that everyone can accept? Regul. Toxicol. Pharmacol. 117, 104751. https://doi.org/10.1016/j.yrtph.2020.104751.

NAS (National Academies of Sciences, Engineering, and Medicine), 2017. Finding a Path to Safety in Food Allergy: Assessment of the Global Burden, Causes, Prevention, Management, and Public Policy. The National Academies Press, Washington, DC. https://doi.org/10.17226/23658.

Pettersson, M.E., Koppelman, G.H., Flokstra-de Blok, B., Kollen, B.J., Dubois, A., 2018. Prediction of the severity of allergic reactions to foods. Allergy 73, 1532–1540. https://doi.org/10.1111/all.13423.

Purington, N., Chinthrajah, R.S., Long, A., Sindher, S., Andorf, S., O'Laughlin, K., Woch, M.A., Scheiber, A., Assa'Ad, A., Pongracic, J., Spergel, J.M., 2018. Eliciting dose and safety outcomes from a large dataset of standardized multiple food challenges. Front. Immunol. 9, 2057.

Remington, B.C., Westerhout, J., Meima, M.Y., Blom, W.M., Kruizinga, A.G., Wheeler, M. W., Taylor, S.L., Houben, G.F., Baumert, J.L., 2020. Updated population minimal eliciting dose distributions for use in risk assessment of 14 priority food allergens. Food Chem. Toxicol. 111259.

Sicherer, S.H., Morrow, E.H., Sampson, H.A., 2000. Dose-response in double-blind, placebo-controlled oral food challenges in children with atopic dermatitis. J. Allergy Clin. Immunol. 105 (3), 582–586.

Sjogren Bolin, Y., 2015. Undeclared milk , peanut , hazelnut or egg - guide on how to assess the risk of allergic reactions in the population. Livsmedelsverkets rapport 17/2015. Livsmedelsverket Natl. Food Agency. https://www.livsmedelsverket.se/globalassets/publikationsdatabas/rapporter/2015/rapport-17-riskvarderingsguide-allergener.pdf. (Accessed 25 August 2020).

Taylor, S.L., Baumert, J.L., Kruizinga, A.G., Remington, B.C., Crevel, R.W., Brooke-Taylor, S., Allen, K.J., of Australia, T.A.B., Houben, G., 2014. Establishment of reference doses for residues of allergenic foods: report of the VITAL expert panel. Food Chem. Toxicol. 63, 9–17.

Taylor, S.L., Moneret-Vautrin, D.A., Crevel, R.W., Sheffield, D., Morisset, M., Dumont, P., Remington, B.C., Baumert, J.L., 2010. Threshold dose for peanut: risk characterization based upon diagnostic oral challenge of a series of 286 peanut-allergic individuals. Food Chem. Toxicol. 48 (3), 814–819.

Taylor, S.L., Crevel, R.W., Sheffield, D., Kabourek, J., Baumert, J., 2009. Threshold dose for peanut: risk characterization based upon published results from challenges of peanut-allergic individuals. Food Chem. Toxicol. 47 (6), 1198–1204.

Vehtari, A., Gelman, A., Gabry, J., 2017. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Stat. Comput. 27 (5), 1413–1432.

Waiblinger, H.U., Schulze, G., 2018. Action levels for food allergens: an approach for official food control in Germany. J. AOAC Int. 101 (1), 17–22.

Wainstein, B.K., Studdert, J., Ziegler, M., Ziegler, J.B., 2010. Prediction of anaphylaxis during peanut food challenge: usefulness of the peanut skin prick test (SPT) and specific IgE level. Pediatr. Allergy Immunol. 21, 603–611.

Wensing, M., Penninks, A.H., Hefle, S.L., Koppelman, S.J., Bruijnzeel-Koomen, C.A., Knulst, A.C., 2002. The distribution of individual threshold doses eliciting allergic reactions in a population with peanut allergy. J. Allergy Clin. Immunol. 110 (6), 915–920.

Westerhout, J., Baumert, J.L., Blom, W.M., Allen, K.J., Ballmer-Weber, B., Crevel, R.W., Dubois, A.E., Fernández-Rivas, M., Greenhawt, M.J., Hourihane, J.O.B., Koplin, J.J., 2019. Deriving individual threshold doses from clinical food challenge data for population risk assessment of food allergens. J. Allergy Clin. Immunol. 144 (5), 1290–1309.

Wheeler, M.W., Westerhout, J., Baumert, J.L., Remington, B.C., 2020. Bayesian Stacked parametric survival with frailty components and interval-censored failure times. Risk Anal. risa 13585. https://doi.org/10.1111/risa.13585 (epub ahead of print).

Yao, Y., Vehtari, A., Simpson, D., Gelman, A., 2018. Using stacking to average Bayesian predictive distributions (with discussion). Bayesian Analysis 13 (3), 917–1007.

Zhu, J., Pouillot, R., Kwegyir-Afful, E.K., Luccioli, S., Gendel, S.M., 2015. A retrospective analysis of allergic reaction severities and minimal eliciting doses for peanut, milk, egg, and soy oral food challenges. Food Chem. Toxicol. 80, 92–100.