UNITED STATES ENVIRONMENTAL PROTECTION AGENCY
WASHINGTON, D.C. 20460

OFFICE OF CHEMICAL SAFETY
AND POLLUTION PREVENTION

June 6, 2022

MEMORANDUM

Subject: 1,2-Benzisothiazolin-e-one (BIT) – Repeat Insult Patch Test in Humans – Non-Guideline

PC Code: 098901	DP Barcode: N/A
Decision No.:	Registration No.:
Petition No.:	Regulatory Action: Data Evaluation Record (DER)
Risk Assess Type: Single chemical	Case No.: 3026
TXR No.:	CAS No.: 2634-33-5
MRID No.: 51171302	40 CFR: N/A

From: Jorge G. Muñiz Ortiz, Ph.D., DABT, Toxicologist
Risk Assessment Branch 1
Tim McMahon, Ph.D., Toxicologist
Risk Assessment Branch 2
Antimicrobials Division (7510M)

Thru: Elizabeth Donovan, Branch Chief
Judy Facey, Ph.D., Risk Assessment Process Leader
Risk Assessment Branch 2
Antimicrobials Division (7510M)

To: Stephen Savage, Chemical Review Manager
Reevaluation Branch
Antimicrobials Division (7510M)

Agency Conclusion: The Agency has reviewed the data submission conducted on humans with BIT. This study is classified as ACCEPTABLE/NONGUIDELINE for regulatory purposes.

CITATION:

Ladics, G.S. (2020) BIT: Repeated insult patch test. Report No. NB-200034-1730. Laboratory Final Report No. 90RC-181. Unpublished. MRID 51171302.

ATTACHMENT

Data Evaluation Record (DER)
Repeat Insult Patch Testing
Non-Guideline
For
1,2-Benzisothiazolin-e-one

PC Code: 098901

EPA Reviewer: Jorge G. Muñiz Ortiz, Ph.D., DABT **Signature**: _____
RAB1, Antimicrobials Division (7510M) **Date**: _____
EPA Secondary Reviewer: Tim McMahon, Ph.D. **Signature**: _____
RAB2, Antimicrobials Division (7510M) **Date:** _____

DATA EVALUATION RECORD

STUDY TYPE: Skin Sensitization non-guideline (Repeat Insult Patch Test) - Human

PC CODE: 098901 **DP BARCODE: N/A**

TEST MATERIAL (PURITY): 1,2-Benzisothiazolin-e-one (19% active ingredient)

SYNONYMS: BIT; PROXEL GXL

CITATIONS: Ladics, G.S. (2020) BIT: Repeated insult patch test. Report No. NB-200034-1730. Laboratory Final Report No. 90RC-181. Unpublished. MRID 51171302.

LABORATORIES: Hill Top Research, Inc., P.O. Box 429501, Cincinnati, OH 45242.

EXECUTIVE SUMMARY:

1,2-Benzisothiazolin-e-one (BIT) was examined for the potential to induce contact sensitization in 111 human subjects (121 subjects entered the study, however 10 withdrew from the study for various reasons). The study was also conducted to observe potential irritation effects from exposures to BIT. Prior to the study, a brief medical history and written informed consent were obtained from each subject. For the main study, one group of subjects (Subjects 1-60) received an induction and elicitation concentration of 500 ppm (27.8 µg/cm^2) Proxel GXL (19% BIT in dipropylene glycol and water) diluted in Rhoplex AC-64 vehicle. A second group (Subjects 61-121) received an induction and elicitation concentration of 1000 ppm (55.6 µg/cm^2) BIT diluted in Rhoplex AC-64 vehicle. A third group (all subjects 1-121) received Rhoplex AC-64 vehicle alone. The subjects from all three groups were exposed to the respective doses during the induction and elicitation phases. All test substances were applied at a volume of 0.2 ml with a pipette directly to the sites on the lateral surface of the right and left upper arm lasting 24 h per application, allowed to air dry for 10 min and covered with a semi-occluded patch. The induction period lasted approximately 3 weeks. Induction applications were done on the same site (documented as the "O" site). If reactions became too strong, a first or second adjacent site (documented as "M" and "M-1" sites, respectively) had to be employed to complete induction. A naïve alternate site (documented as the "A" site) and the original induction site ("O") were used for the challenge applications. A rest period of approximately 2 weeks followed the induction period before the first challenge application. If rechallenge was performed, a single application

of the test materials was made to naïve sites on the lower back. The patch application site was marked with a line of gentian violet with a cotton tip applicator across the top adhesive part of the patch.

Reactions were scored Monday, Wednesday and Friday, 48 or 72 hours after each induction application (24 or 48 h after removal of patches) and 48 h and 96 h after challenge application (24 and 72 h after patch removal).

Responses to the skin from each patch application were examined and graded under a 100-watt incandescent blue bulb. The same scorer evaluated the reaction sites on any group of subjects and was blinded as to treatment assignments and any previous scores. There was no spot checking by an independent/alternative scorer as a quality assurance/quality control measure. Reactions to the test materials were documented on case report forms by the numerical and letter grade scoring system defined below.

Results of the induction phase showed that one subject, no. 46, exhibited a response of Grade 1. However, the response was observed with both the test substance at 27.8 $\mu g/cm^2$ and with the vehicle control, Rhoplex AC-64. The response was observed from the second application to the ninth and final application. Subjects 66 and 96, both exposed to 55.6 $\mu g/cm^2$ showed a response only during the challenge phase. Subject 46 showed a Grade 1 reaction in the original and adjacent sites at 48 and 96 h after exposure to both the test substance and the vehicle control. Subject 66 showed a Grade 1 reaction in the original site at 48 h after exposure, Grade 2 with spreading at the original and adjacent sites at 96 h after exposure to the test substance. Subject 66 also showed a Grade 1 and Grade 2 with spreading response in the original and adjacent sites at 96 h post exposure, respectively. Subject 96 only showed a Grade 1 response to the challenge in the original site 48 h post exposure. Subject 66 was rechallenged, however, no response to the test substance or vehicle control were observed.

Subjects 12, 35, 67, 85, and 115 showed Grade 1 reactions to the vehicle Rhoplex AC-64 sporadically during the induction phase. Subject 12 showed a Grade 2 reaction to the vehicle in the fifth application. None of these subjects showed a reaction during the challenge phase to the vehicle or BIT at any of the concentrations tested.

Subjects 12 and 20 showed Grade 1 reactions to BIT at 27.8 $\mu g/cm^2$ sporadically during the induction phase, however, did not show any reaction during the challenge phase.

Subjects 63, 67, 75, 80, 103, and 109 showed Grade 1 reactions to BIT at 55.6 $\mu g/cm^2$ sporadically during the induction phase. None of the subjects showed any reaction during the challenge phase to the same dose of BIT.

Based on the study results and statistical analysis, BIT is not a dermal sensitizer at the doses used in this study. The No Observed Adverse Effect Level (NOAEL) for this study is 55.6 $\mu g/cm^2$; there was no LOAEL established in this study.

This study is classified as **acceptable/non-guideline.** It was submitted by the registrant for fulfillment of a guideline and provides information on elicitation thresholds to BIT in humans.

This study is considered adequate for quantitative use in risk assessment in a weight-of-evidence determination for the dermal sensitization elicitation threshold to BIT.

COMPLIANCE: The analysis of this unpublished study was not conducted in compliance with the U.S. EPA Good Laboratory Standards 40 CFR Part 160.

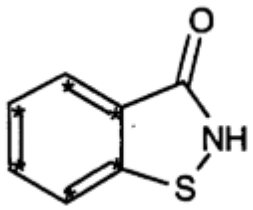
CONFIDENTIALITY: No claim of confidentiality is made for any information contained in this study on the basis of falling within the scope of FIFRA sec. 10(d)(1)(A), (B), or (C).

I. MATERIALS AND METHODS

A. MATERIALS:

1. Test Materials: 1,2-Benzisothiazolin-3-one
Description: Brown, slightly thickened liquid
Lot/Batch #:
Purity: 19% active ingredient in dipropylene glycol and distilled water

CAS # of TGAI: 2634-33-5
Structure:



2. Sample Preparation, Vehicle and/or Positive Control: All samples were made fresh daily prior to each induction and/or challenge application. Proxel GXL 19% was diluted 50% in distilled water. To prepare the working solutions (1000 ppm and 500 ppm) it was diluted further in Rhoplex AC-64. The solutions were used within 1-3 hours of preparation. A positive control was not used in this study. There was no solubility and stability analysis of the test substance or dilutions with the vehicle.

B. STUDY DESIGN and METHODS:

The objective of the present study was to examine whether BIT (Proxel GXL) has the potential to cause allergic contact dermatitis. To this end, human subjects were recruited for repeated insult patch testing at two concentrations to determine the presence of irritation and dermal sensitivity at 500 ppm (27.8 µg/cm^2) and 1000 ppm (55.6 µg/cm^2). The subjects from each respective group were exposed to the test substance at these doses during the induction and elicitation phases.

Study Participants

A total of 121 subjects were originally recruited for this study, however, 10 had to be removed for various reasons (work schedule changes, missed applications or scoring, loss of job, or traveling). All the subjects were 18 years old or older. Before the start of the testing a brief medical history and written informed consent were obtained from each subject. Subjects were rejected from the study for any of the following reasons:

- History of poor health or insulin-dependent diabetes

- History of bilateral mastectomy or any mastectomy within the past year
- History of active skin cancer
- Current skin disease, which may contraindicate participation, including psoriasis or active eczema, even if currently controlled through medication
- History of participation in a Draize type patch test within the past three months
- Current use of anti-inflammatory steroids or antihistamine medications
- Severe asthma
- Use of topical drugs at site of patching

Any subject absent more than once during the induction phase or at any time during the challenge week was withdrawn from the study.

Institutional Review Board (IRB) Approval and Informed Consent

Each subject was provided with an informed consent form for them to sign acknowledging participation in the study. The form detailed the procedure to which they were going to be submitted to. The study does not mention an IRB approval procedure.

Induction Testing

The following methodology on patch testing is reproduced from the study report:

“Each sample in the test kit will be applied directly to the test sites on the skin of the upper arm for contact periods lasting 24 hours per application. The samples will be air dried until it forms a "film" on the skin. The test sites will then be covered with a semi-occluded patch. Induction applications will be made three times per week (on Monday, Wednesday, and Friday) for three successive weeks. The nine applications made during these three weeks will be termed Induction Application Nos. 1 through 9, respectively. During the fourth week (on Monday), any subject who is absent for one of the regularly scheduled induction applications will receive a make-up induction application. All induction applications for individual test samples will be made to the same site (the site receiving the original sample at Induction Application No. 1) unless reactions become so strong as to make this inadvisable. In these cases, subsequent applications of the offending test material will be made to an adjacent area, and a second change of site will be made if a second strong reaction occurs. If a third strong reaction to the test preparation is manifested, patches of this test material will be discontinued until after the two-week rest period has been completed. The use of a first and second adjacent site will be identified on case report forms as M and M-1 sites, respectively, to indicate movement of test sites from the original (O) application site. A 10 to 15-day rest period will follow the final induction application” (page 4 of Appendix I of study report).

Challenge Testing

"Following the rest period, on Monday of the sixth week, a challenge application of the sample in the test kit will be made to each panelist. During the challenge application the samples will remain in contact with the skin for a period of 24 hours. Challenge will consist of application to a naive site located adjacent to the original (0) application site. Simultaneous application to a pre-exposed site (i.e., the original site used for Induction Application No. 1) will be made concurrently with the challenge at a naïve site (page 4 of Appendix I of study report).

"Observations at a naive site will provide a basis for an interpretation of contact sensitization. Data obtained from a pre-exposed site (during the challenge period) will be reported and used to support the conclusions drawn from observations at the naive site, but positive reactions at a pre-exposed site will not be interpreted as significant evidence of contact sensitization unless confirmed by observations at a naïve site (page 4 of Appendix I of study report)."

Rechallenge Testing

Rechallenge testing was performed by applying the sample or sample components to the naïve site to confirm reactions indicative of contact sensitization.

Evaluations

"Reactions will be scored Monday, Wednesday and Friday, 48 or 72 hours after each induction application (24 or 48 hours after removal of patches) 48 hours and 96 hours after challenge application (24 and 72 hours after patch removal) (page 5 of Appendix I of study report).

"Skin responses to each patch application will be examined and graded under light supplied by a 100-watt incandescent blue bulb. On any group of subjects, the same scorer will carry out all evaluations of the reaction sites. The scorer will be blinded as to treatment assignments and any previous scores. Reactions to the test materials are to be documented on case report forms by the numerical and letter grade scoring system defined in this protocol (page 4 of Appendix I of study report)."

Reactions to test materials were scored according to the following scale:

0 - No evidence of any reaction
1- Reaction readily visible but mild unless letter grade appended (see grade E and F below). Mild reactions include weak but definite erythema, and weak superficial skin responses such as glazing, cracking, or peeling.
2 - Definite papular response (append E or F or S if appropriate).
3 - Definite edema (append E or F or S if appropriate).
4 - Definite edema and papules (append E or F or S if appropriate).
5 - Vesicular/bullous eruption.
E - If the letter E is appended to the numerical grade, it documents the presence of strong erythema at the patch site.
F - If the letter F is appended to the numerical grade, it documents the presence of strong effects on superficial layers of the skin. These include fissures, a film of dried serous exudate, small petechial

erosions and/or scabs.
S - If the letter S is appended to the numerical grade, it documents the presence of a reaction spreading beyond test site.

Patch applications were terminated or moved to naïve adjacent sites if any of the grades were 2, 3, 4, 5 or E, F and S in any combination with a numerical grade including 1E or 1F. A Grade 1 with ** appended, indicated the presence of glazing, cracking, or peeling but no erythema.

Symbols for tabulating data (in addition to scoring grades):

O – Original application site.
M – Adjacent site for application after first strong reaction during induction.
M-1 – Second adjacent site for application after second strong reaction during induction.
A – Naïve adjacent site used during challenge application.
NP – Number of panelists not included in score totals: (See L, *, below).
MU – Make-up session for panelists with earlier absence(s).
( ) – Scores in ( ) are residual or flaring reactions on sites of earlier applications, recorded only when of special interest. (Not included in score totals).

Symbols used to document deviation from experimental plan:

R – Applied to adjacent site because of adhesive reaction or for other reasons irrelevant to test material reactions.
X – Patch omitted due to previous strong test reaction(s).
XR – Patch omitted for reasons unrelated to the test.
L – Test patch lost soon after application. (Not included in score totals.)
- – Panelist absent
* – Residual reaction to earlier application after absence. (Not included in score totals.)

Statistical Analysis

The authors of the study did not perform a statistical analysis of the data.

Statistical analysis of the data in the study was performed by Dr. Jonathan Cohen from ICF (Memorandum, J. Cohen, ICF, January 28, 2022). Five statistical analyses were performed to potentially identify a dose at which elicitation would be statistically significantly different from controls. The first test was to summarize the percentages of positive results for each combination of test type and dose. The second test was a McNemar test for each dose level to determine if the probabilities of a positive result are the same for induction and challenge tests among subjects who participated in both phases. The third statistical analysis was a Fisher's exact test for each test type that compared the positivity rates for the high (55.6 $\mu g/cm^2$) and low (27.8 $\mu g/cm^2$) dose levels. The fourth statistical test was a Fisher's exact test for each test type as the third test but including the vehicle only group. The final and fifth statistical analysis performed was fitting a set of non-linear mixed statistical models for each test type to predict the probability of a positive result as a function of the dose and a random effect for each subject. A logistic model is used to model the log-odds of a positive result as a linear or quadratic function of the dose plus the

normally distributed random effect.

II. RESULTS

Induction and Challenge Test Results

Table 1 shows all the subjects that reacted from 9 applications of the vehicle Rhoplex AC-64 or BIT at 27.8 µg/cm^2 or 55.6 µg/cm^2 to test for induction and 4 applications to test for elicitation or challenge.

Subjects 46, 66, and 96 exhibited irritation scores of Grade 1 or higher during the challenge. During the induction, Subject 46, who was exposed to 27.8 µg/cm^2 BIT (Proxel GXL), showed irritation scores of Grade 1 from the second application throughout the 9th and final application. The same results were observed with the Rhoplex AC-64 vehicle control. In the challenge phase, Subject 46 showed Grade 1 effects from exposure to the chemical and the vehicle control in the original and adjacent sites 48 and 96 hours of exposure. The effects observed were described as mild circumscribed erythema.

Subject 66 was exposed to 55.6 µg/cm^2 and did not show any induction response at or higher than Grade 1, rather, they were all Grade 0 throughout the induction. However, Subject 66 did show a response of Grade 1 during the challenge phase in the original site after 48 hours and Grade 2 with spreading of the effect in the original and adjacent sites 96 hours after exposure. Subject 66 also showed a reaction of Grade 1 in the original site 96 hours after exposure and Grade 2 with spreading of the effect in the adjacent site 96 hours after exposure to the vehicle control. The effects were described as mild erythema at the 48-h time point. By the 96 h timepoint there was a papular response to the test substance and the vehicle.

Subject 96 was exposed to 55.6 µg/cm^2 BIT and had a Grade 1 score only during the challenge phase 48 hours after exposure in the original site of exposure. All the other scores were Grade 0 during the induction phase and other challenge phase timepoints and application location with the test chemical. The subject did not have any reaction above Grade 0 to the vehicle control at any time point or application location during the induction or challenge phase. The effects were described as mild erythema response to the test substance by the study authors.

Rechallenge Test Results

Subject 66 underwent a rechallenge test with 55.6 µg/cm^2 and did not show any dermal sensitization 48 h or 96 h after exposure.

Statistical Analysis

Statistical analyses performed did not identify a statistically significant difference in response to challenge application of BIT from control.

Table 1. Subjects that showed reactions to the vehicle or BIT during the induction and challenge phases.

Rhoplex AC-64																															
	1	2		3			4			5			6			7			8			9			MU			Challenge			
Subject	O	O	M	O	M	M-1	O	M	M-1	O	M	M-1	O	M	M-1	O	M	M-1	O	M	M-1	O	M	M-1	O	M	M-1	O	A	O'	A'
12										2										1											
35				1																											
46		1		1			1			1			1			1			1			1						1	1	1	1
66																														1	2S
67																						1									
85				1																											
115																1															
Proxel GXL (BIT) (27.8 µg/cm^2)																															
12										1			1			1															
20										1																					
46		1		1			1			1			1			1			1			1						1	1	1	1
Proxel GXL (55.6 µg/cm^2)																															
63																1			1			1			1						
66																												1		2S	2S
67																						1									
75																1															
80																			1												
96																												1			
103																			1												
109																1						1									

Numbers and letters depict reaction grade level. For a description see Evaluations section.
O - Original Site
MU - If an application was missed, the application was done at a make up date.
M - First adjacent site
M-1 - Second adjacent site
O, A - First scoring of original and adjacent challenge sites (48 h)
O', A' - Second scoring of challenge sites (96 h)

III. REVIEWER'S CONCLUSIONS:

The current study was conducted to examine doses of BIT that cause skin sensitization in a repeat insult patch test (RIPT).

The present RIPT study utilized a study population of 111 test subjects (10 dropped out) to determine if BIT is a dermal sensitizer at two doses. All test subjects were exposed to Rhoplex AC-64, a negative control vehicle. The RIPT study was designed to mimic repeated dermal exposure to a BIT product.

The results of the RIPT study showed that 6 of the 111 test subjects showed a Grade 1 reaction to the vehicle control and one of them also showed a Grade 2 reaction to the vehicle control. Three subjects showed consecutive or sporadic Grade 1 reaction(s) to BIT at 27.8 $\mu g/cm^2$ and 6 subjects showed sporadic or consecutive Grade 1 reaction(s) to 55.6 $\mu g/cm^2$ during the induction phase. Two subjects showed positive reactions to the vehicle control during the challenge phase. One subject who showed positive reactions to Rhoplex AC-64 during the challenge phase also showed positive reactions to 27.8 $\mu g/cm^2$ during the challenge phase. Another subject who showed a positive reaction to the vehicle control during the challenge phase, also showed positive reactions to BIT at 55.6 $\mu g/cm^2$ during the challenge phase. Another subject showed a Grade 1 reaction to 55.6 $\mu g/cm^2$ during the challenge phase but did not show any effects during the induction phase.

It is concluded that the study is appropriate for quantitative use as part of a weight of evidence determination.

Based on the study results and statistical analysis, BIT is not a dermal sensitizer at the concentrations tested in this study and did not cause an elicitation response at any dose tested. The No Observed Adverse Effect Level (NOAEL) for this study is 55.6 $\mu g/cm^2$; there was no LOAEL established in this study.

IV. STUDY LIMITATIONS/DEFICIENCIES:

- Only two doses of BIT were used in the study. This limits the ability to identify a clear dose response.
- A dose or doses at which adverse effects could have been observed were not tested.
- Analyses for solubility and stability on the test substance or dilutions with the vehicle were not reported in the study.
- A detailed description of the vehicle was not provided.
- A source for the scale used to grade irritation and sensitization was not provided.

V. REFERENCES

Memorandum. Cohen, J. Statistical Review of the BIT dermal sensitivity study. January 18, 2022, to Judy Facey, Jorge Muniz-Ortiz, Shawn Garred, Timothy McMahon, and Kathryn Korthauer.

APPENDIX I

Statistical Analysis Report

MEMORANDUM

To:	Judy Facey, EPA; Jorge Muniz-Ortiz, EPA; Shawn Garred, EPA; Timothy McMahon, EPA; Kathryn Korthauer, EPA
From:	Jonathan Cohen, ICF
Date:	January 28, 2022
Re:	Statistical Review of the BIT dermal sensitivity study

Introduction and Summary

In May 2020, DuPont Nutrition & Biosciences submitted their report "BIT: Repeated insult patch test" (Ladics 2020). in support of the dermal sensitization risk assessment for 1,2-Benzoisothiazolin-3-one (BIT). This report is based on a dermal patch study conducted for Rohm and Haas Company by Hill Top Research Inc in the year 1990. In this study there were originally 121 subjects, but only 111 subjects completed the study. The induction phase consisted of nine 24-hour applications at the same patch site over a period of three weeks. (In some cases where there was a strong reaction to the test material, the patch location for future applications was then changed to an adjacent site.) This was followed by a rest period of about two weeks and then a challenge patch test at the original site and at a new adjacent "naïve") site which were scored after 48 and 96 hours. One subject, number 66, did not respond to the induction tests, but did respond to the challenge test, and so this subject was given a rechallenge patch test at a naive site on their lower back. The first group of 60 subjects, out of which 56 subjects completed the study, received patches on their upper arms with a concentration of 500 ppm of BIT (27.8 $\mu g/m^3$) diluted in the vehicle Rhoplex AC-64, the Low dose. The second group of 61 subjects, out of which 55 subjects completed the study and I subject dropped out before any of the induction tests, received patches on their upper arms with a concentration of 1,000 ppm of BIT (55.6 $\mu g/m^3$) diluted in the vehicle Rhoplex AC-64, the High dose. All 121 subjects, out of which 111 subjects completed the

study and I subject dropped out before any of the induction tests, were also tested with patches containing the undiluted Rhoplex AC-64 vehicle only, which corresponds to a dose of 0 $\mu g/m^3$, the Vehicle or Zero dose. The Nonzero dose includes both the Low and High doses. The study report is not totally clear on this point, but it appears that these vehicle-only induction, challenge, and rechallenge tests were carried out simultaneously with the low and high dose BIT patch tests, but at a different body site.

Each patch test was scored according to the following scale (summary copied from the DER):

0 - No evidence of any reaction
1- Reaction readily visible but mild unless letter grade appended (see grade E and F below). Mild reactions include weak but definite erythema, and weak superficial skin responses such as glazing, cracking or peeling.
2 - Definite papular response (append E or F or S if appropriate).
3 - Definite edema (append E or F or S if appropriate).
4 - Definite edema and papules (append E or F or S if appropriate).
5 - Vesicular/bullous eruption.
E - If the letter E is appended to the numerical grade it documents the presence of strong erythema at the patch site.
F - If the letter F is appended to the numerical grade it documents the presence of strong effects on superficial layers of the skin. These include fissures, a film of dried serous exudate, small petechial erosions and/or scabs.
S - If the letter S is appended to the numerical grade it documents the presence of a reaction spreading beyond test site.

The results of the study showed that the only observed non-zero grades were 1, 2, and 2S.

ICF was asked to perform a statistical review of the study results. In section 2, we present the main statistical analysis of the 111 subjects that completed the study, excluding any subjects that dropped out at the beginning (subject 82), during the induction period, or before or during the challenge testing. We will refer to these data as “Complete data” since these data are from the subjects with complete data. In section 3, for completeness, we present a supplementary statistical analysis using the same statistical methods applied to all 120 subjects that started the study, excluding the subject 82 that dropped out at the beginning. We will refer to these data as “All data”.

The main purpose of the study was to evaluate clinical sensitization which they defined in the study as having no response (reaction) to the induction, a response to the challenge at the naïve site, and a confirmatory response to the rechallenge at another naïve site. Since only one subject (subject 66) showed no response to the induction, responded at grade 1 or higher to the challenge, but that subject showed no response to the rechallenge, there is no evidence of clinical sensitization. These statistical analyses therefore focus on clinical irritation, defined in the report as having a response during the induction or challenge testing but no response to a rechallenge.

Most of the grades were 0 and the only observed non-zero grades were 1, 2, and 2S. Only two subjects had grades 2 or higher: subject 12 for one patch during induction using the vehicle only; subject 66 during the challenge using the vehicle only and using the High dose. For these reasons, the statistical analyses do not distinguish between the non-zero grade responses. The statistical analyses evaluate clinical irritation separately for induction and challenge tests and summarize the induction and challenge test results as follows: A negative induction result (summary score 0) for a given dose is if all the induction tests were grade 0. A positive induction result for a given dose (summary score 1) is if at least one induction test was grade 1 or higher. A negative challenge result for a given dose (summary score 0) is if all the challenge tests (at both sites) were grade 0. A positive challenge result for a given dose (summary score 1) is if at least one challenge test (at one or both sites) was grade 1 or higher. We will define the test type as "induction" for induction tests and "challenge" for challenge tests. All statistical tests were carried out at the 5% significance level.

We began the analysis by summarizing the percentages of positive results for each combination of test type and dose. For the main analyses (excluding any subjects with dropped tests) the induction tests at doses High, Low, Nonzero, and Vehicle had 5 positive results out of 55 tests (9.1%), 3 positive results out of 56 tests (5.4%), 8 positive results out of 111 tests (7.2%), and 6 positive results out of 111 tests (5.4%), respectively. Also for the main analyses, the challenge tests at the same doses had 2 positive results out of 55 tests (3.6%), 1 positive results out of 56 tests (1.8%), 3 positive results out of 111 tests (2.7%), and 2 positive results out of 111 tests (1.8%), respectively. Thus the percentages of positive results are almost the same for the Vehicle and Low doses and higher for the High and Nonzero doses. The supplementary analyses for all the data show a similar pattern. For all the data, the induction tests at doses High, Low, Nonzero, and Vehicle had 6 positive results out of 60 tests (10.0%), 3 positive results out of 60 tests (5.0%), 9 positive results out of 120 tests (7.5%), and 6 positive results out of 120 tests (5.0%), respectively. Also for all the data, the challenge tests at the same doses had 2 positive results out of 58 tests (3.4%), 1 positive results out of 56 tests (1.8%), 3 positive results out of 114 tests (2.6%), and 2 positive results out of 114 tests (1.8%), respectively.

The second statistical analysis was a McNemar test for each dose level for whether the probabilities of a positive result are the same for the induction and challenge tests, among people taking both tests. For both the main and supplementary analyses, and for each dose level, the proportions of positive results for the induction and challenge tests were not statistically significantly different.

The third statistical analysis was a Fisher's exact test for each test type that compared the positivity rates for the High and Low dose levels. For both the main and supplementary analyses, this test showed that for each test type, the proportions of positive results for each dose level were not statistically significantly different.

The fourth statistical analysis was a Fisher's exact test for each test type that compared the positivity rates for the High, Low, and Zero (Vehicle only) dose levels. This test is of limited statistical validity because that analysis assumes that the results at all three dose levels are statistically independent, even though the same persons were tested at the High and Zero doses and the same persons were tested at the Low and Zero doses. For both the main and supplementary analyses, this test showed that for each test type, the proportions of positive results for each of the three dose levels were not statistically significantly different.

The fifth statistical analysis was fitting a set of non-linear mixed statistical models for each test type to predict the probability of a positive result as a function of the dose and a random effect for each person. A logistic model is used to model the log-odds of a positive result as a linear or quadratic function of the dose plus the normally distributed random effect. This method avoids the above-mentioned problem with the fourth statistical analysis that results for the same person using different doses should not be expected to be statistically independent. The detailed equations and parameter estimates are presented below together with plots of the observed and median predicted probability against each dose level, together with a 90% uncertainty interval. Consistent with the earlier findings, the linear functions show that for both test types, the dose effect is not statistically significant. The quadratic models were not supported because the quadratic coefficients were not statistically significant.

In summary, all the statistical analyses show no statistically significant dose effects on clinical irritation.

The attached SAS program file bitsascode15.sas was used for the statistical analysis. The attached input file alldata.xlsx contains the summary scores (0 or 1) for the induction and challenge dose by subject and dose, where the "Nodrops" tab has the data for subjects with complete data without any drops (analyzed in section 2) and the "All" tab has the data for All subjects (analyzed in section 3). In the All tab, subjects that did at least 1 induction test but no challenge tests have summary scores of either 1 or 0 for induction and missing summary scores for challenge. The attached file bitsascode15.lst contains the SAS output listing. The output graphs are included in this memorandum.

Statistical Analyses for Subjects with Complete Data

The analyses in this section only use data from the 111 subjects with complete data (no drops) for the induction and challenge tests. As noted above, for each subject and dose, an induction test result is positive (coded as 1) if and only if one or more of the induction patch tests is at grade 1 or higher, and an induction test result is negative (coded as 0) if and only if all of the induction patch tests are at grade 0. Similarly, for each subject and dose, a challenge test result is positive (coded as 1) if and only if one or more of the challenge patch tests is at grade 1 or higher, and a challenge test result is negative (coded as 0) if and only if all of the challenge patch tests are at grade 0.

Proportions of Positive Test Results

Table 1 shows the number and proportion of positive test results for each combination of dose level and test type. The dose levels are High (55.6 μg/m^3 BIT), Low (27.8 μg/m^3 BIT), Non-zero (High or Low), or Vehicle (Rhoplex AC-64, i.e., 0 μg/m^3 BIT).

Table 1. Numbers and proportions of positive test results by dose level and test type. Complete data.

Dose Level	Induction: Number of Positive Tests	Induction: Number of Subjects	Induction: Positivity Percentage	Challenge: Number of Positive Tests	Challenge: Number of Subjects	Challenge: Positivity Percentage
High	5	55	9.1%	2	55	2.6%
Low	3	56	5.4%	1	56	1.8%

Dose Level	Induction: Number of Positive Tests	Induction: Number of Subjects	Induction: Positivity Percentage	Challenge: Number of Positive Tests	Challenge: Number of Subjects	Challenge: Positivity Percentage
Non-zero	8	111	7.2%	3	111	2.7%
Vehicle	6	111	5.4%	2	111	1.8%

McNemar Test for the Agreement Between the Induction and Challenge Tests

Table 2 shows the number and proportion of positive test results for each dose level and test type for the subjects that did both induction and challenge tests. Table 2 also shows the p-value for the McNemar test that the marginal probabilities of a positive test result are the same for both test types. The dose levels are High (55.6 μg/m^3 BIT), Low (27.8 μg/m^3 BIT), Non-zero (High or Low), or Vehicle (Rhoplex AC-64, i.e., 0 μg/m^3 BIT). This table is the same as Table 1 because Section 2 only analyzes data for subjects with complete results. The corresponding pair of Tables 9 and 10 in Section 3 are not the same since Section 3 includes subjects that failed to complete the study, and, for example, may have done induction tests but not challenge tests.

Table 2. McNemar test P-values and the numbers and proportions of positive test results by dose level and test type for subjects that did both induction and challenge tests. Complete data.

Dose Level	Induction: Number of Positive Tests	Induction: Number of Subjects	Induction: Positivity Percentage	Challenge: Number of Positive Tests	Challenge: Number of Subjects	Challenge: Positivity Percentage	P-Value
High	5	55	9.1%	2	55	2.6%	0.45
Low	3	56	5.4%	1	56	1.8%	0.50
Non-zero	8	111	7.2%	3	111	2.7%	0.18
Vehicle	6	111	5.4%	2	111	1.8%	0.22

The McNemar test for each dose level compares the positivity probabilities for the induction and challenge test. For example, for the High dose level, the positivity percentage was 9.1% for the induction and 2.6% for the challenge. The p-values are shown in the final column. Since all the p-values are greater than 0.05, for each dose level there is insufficient evidence that the probabilities of a positive result for the induction and challenge tests are different. A statistically significant result could have been interpreted as showing evidence that the probability of a reaction during the induction is not the same as the probability of a reaction during the challenge that followed the induction.

Fisher's Exact Test Comparing the Positivity Probabilities Between the High and Low Doses

For the induction tests, as shown in Table 1, the positivity percentages were 9.1% at the High dose and 5.4% at the Low dose. Table 1 also shows the results for the challenge tests. Fisher's exact test can be used to test whether the corresponding probabilities of a positive result are the same at both doses. Note that the results for the Low and High doses can be assumed to be statistically independent because those two doses were tested on different subjects. The p-values for the induction and challenge tests are shown in the following table.

Table 3. P-values from Fisher's exact test for differences in the probability of a positive result between the High and Low dose levels. By test type. Complete data.

Test Type	P-value (Fisher's Exact Test)
Induction	0.49
Challenge	0.62

Since both p-values are greater than 0.05, for both test types there is insufficient evidence that the probabilities of a positive result for the Low and High dose levels are different.

Fisher's Exact Test Comparing the Positivity Probabilities Between the High, Low, and Zero (Vehicle Only) Doses

For the induction tests, as shown in Table 1, the positivity percentages were 9.1% at the High dose, 5.4% at the Low dose, and 5.4% for the Vehicle dose (i.e., 0% BIT). Table 1 also shows the results for the challenge tests. Under the assumption that the results for all three doses are statistically independent, Fisher's exact test can be used to test whether the corresponding probabilities of a positive result are the same at all three doses. Unfortunately, it is unrealistic to assume that the results for those three doses are statistically independent because all the subjects tested at the Low dose were also tested at the Vehicle dose, and all the subjects tested at the High dose were also tested at the Vehicle dose. The independence assumption would require that the response is essentially due to the dose and not to the person, which is at best an approximation. The statistical models developed in the next subsection more properly address this question by modeling both subject and dose effects. The p-values for the induction and challenge tests are shown in the following table.

Table 4. P-values from Fisher's exact test for differences in the probability of a positive result between the High , Low, and Vehicle dose levels. By test type. Assumes independence. Complete data.

Test Type	P-value (Fisher's Exact Test)
Induction	0.63

Test Type	P-value (Fisher's Exact Test)
Challenge	0.72

Since both p-values are greater than 0.05, for both test types there is insufficient evidence that the probabilities of a positive result for the High, Low and Vehicle dose levels are different.

Statistical Models for the Probability of a Positive Result

In this final subsection we develop and fit statistical models for the probability of a positive result (i.e., a response), separately for the induction and challenge tests. The probability is a function of both the dose (a fixed effect) and the subject (a normally distributed random effect).

Let exp denote the exponential function and log denote the natural logarithm. Let the dose be measured in $\mu g/m^3$ so dose is either 27.8 (Low dose), 55.6 (High dose), or 0 (Vehicle).

Linear dose model

For each test type, the fitted linear dose model is of the form:

Eta = alpha + beta × dose + u

The random variable u varies across subjects and is normally distributed with a mean of zero and a variance s2u. To improve convergence of this model, instead of directly using s2u as the parameter, we defined the new parameter logs2u so that

Variance(u) = s2u = exp(logs2u)

P(response) = exp(eta) / {1 + exp(eta)}

From this model, we get

Log-odds of a response = Logit (response) = log {P(response) / [1 − P(response)] } = eta

Thus we have a logistic mixed model for the probability of a response.

This model was fitted using SAS's NLMIXED procedure.

Quadratic dose model

For each test type, the fitted quadratic dose model is of the form:

Eta = alpha + beta × dose + gamma100/100 × $dose^2$ + u

The quadratic coefficient gamma100 was divided by 100 to improve convergence of the model.

As before, the random variable u varies across subjects and is normally distributed with a mean of zero and a variance s2u. To improve convergence of this model, instead of using s2u as the parameter, we defined the new parameter logs2u so that

Variance(u) = s2u = exp(logs2u)

P(response) = exp(eta) / {1 + exp(eta)}

From this model, we get

Log-odds of a response = Logit (response) = log { P(response) / [1 – P(response)] } = eta

Thus we have another logistic mixed model for the probability of a response.

This model was also fitted using SAS's NLMIXED procedure.

The models were fitted using the SAS options tech= nrridg, lis = 3, maxiter = 1000000, as defined in the SAS documentation. An array of suitable starting points was chosen by a trial-and-error approach to find a model that converged. The model fitting of the random effects requires an integration method. In this section, the analysis is for the complete data, excluding subjects with dropped tests, and the integration method chosen was the method of adaptive quadrature for the induction tests (qmax = 100) but using 2 quadrature points for the challenge tests (qpoints = 2). Although the adaptive quadrature integration method is preferred, the models fitted using adaptive quadrature failed to converge for the challenge tests. For the analysis of all the data in Section 3 below, we used the preferred integration method of adaptive quadrature with qmax = 100 for both the induction and challenge tests.

Fitted models

The estimated parameters of the fitted models for induction together with their standard errors and p-values are given in Tables 5 and 6 for the linear dose and quadratic dose models, respectively. The estimated parameters of the fitted models for challenge together with their standard errors and p-values are given in Tables 7 and 8 for the linear dose and quadratic dose models, respectively.

Table 5. Parameter estimates for the linear dose mixed model for the probability of an induction response. Complete data.

Parameter	Estimate	Standard Error	P-value
alpha	-9.9198	1.8869	< 0.0001
beta	0.0297	0.0242	0.2232
logs2u	4.6466	0.6974	< 0.0001

Table 6. Parameter estimates for the quadratic dose mixed model for the probability of an induction response. Complete data.

Parameter	Estimate	Standard Error	P-value
alpha	-9.8267	1.9158	< 0.0001
beta	-0.0094	0.1254	0.9401
gamma100	0.0718	0.2268	0.7523
logs2u	4.6452	0.7006	< 0.0001

Table 7. Parameter estimates for the linear dose mixed model for the probability of a challenge response. Complete data.

Parameter	Estimate	Standard Error	P-value
alpha	-34.3301	19.1656	0.0760
beta	0.1058	0.0885	0.2346
logs2u	6.3167	1.3449	< 0.0001

Table 8. Parameter estimates for the quadratic dose mixed model for the probability of a challenge response. Complete data.

Parameter	Estimate	Standard Error	P-value
alpha	-34.4577	19.1140	0.0742
beta	0.1496	0.4215	0.7233
gamma100	-0.0792	0.7322	0.9140
logs2u	6.3245	1.3439	< 0.0001

For the linear dose models tabulated in Tables 5 and 7, the p-values for the slope parameter beta are 0.22 and 0.23, respectively, which implies that the dose effect is not statistically significant (at the 5% level) for both induction and challenge. For the quadratic dose models tabulated in Tables 6 and 8, the p-values for the quadratic parameter gamma100 are 0.75 and 0.91, respectively, which implies that the quadratic dose effects are not statistically significant (at the 5% level) for both induction and challenge, so that the linear dose models should be preferred.

The predicted and observed probabilities of a response to induction are plotted in Figures 1 and 2 for the linear dose and quadratic dose models, respectively. The predicted and observed probabilities of a

response to the challenge are plotted in Figures 3 and 4 for the linear dose and quadratic dose models, respectively. The probability for induction = 1 is the probability of a positive result for at least one induction test. The probability for challenge = 1 is the probability of a positive result for at least one challenge test. The probabilities are shown on a logarithmic scale since they vary over a wide range. The observed probabilities are the observed proportions of subjects with a positive result. The red curve shows the predicted median probability, calculated by replacing u by zero in the equation for eta, since 0 is the median of a normal random variable with mean 0 and variance exp(logs2u). The brown curve (with very small probabilities) shows the predicted 5th percentile of the probability, calculated by replacing u by probit(0.05) × exp(logs2u/2) in the above equation for eta, where probit is the inverse of the cumulative distribution function for a standard normal random variable, since this gives the 5th percentile of the random effect u. The blue curve (with probabilities close to 1) shows the predicted 95th percentile of the probability, calculated by replacing u by probit(0.95) × exp(logs2u/2) in the equation for eta, since this gives the 95th percentile of the random effect u. Thus the interval between the brown and blue curves is a 90% uncertainty interval for the probability of a positive result. Note that for the linear dose models the fitted curves may appear to be exactly linear using the logarithmic scale for the probability, but in fact those lines are only approximately linear.

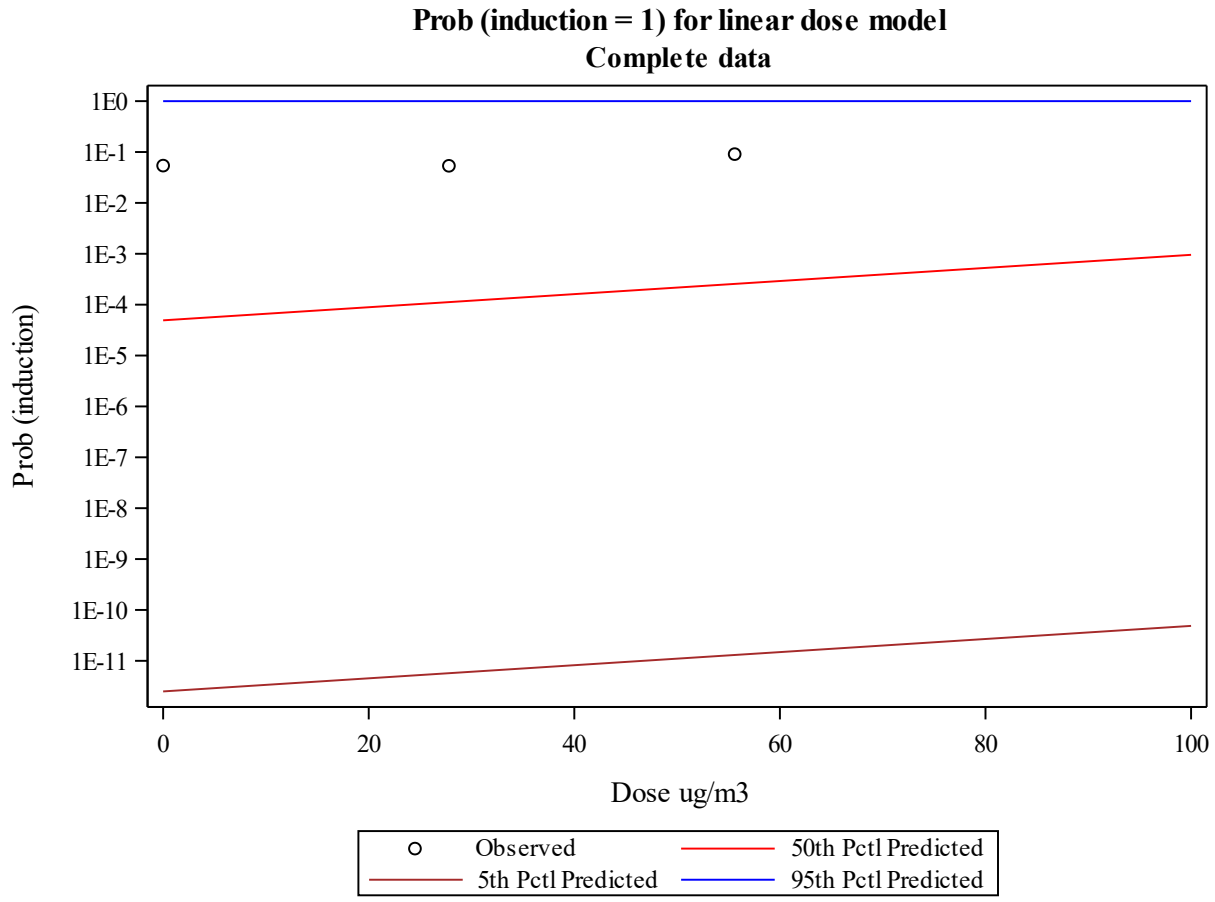


Figure 1. Predicted median and observed probability of a positive induction response based on the linear dose model, together with a 90% uncertainty interval. Plotted on a logarithmic scale. Complete data.

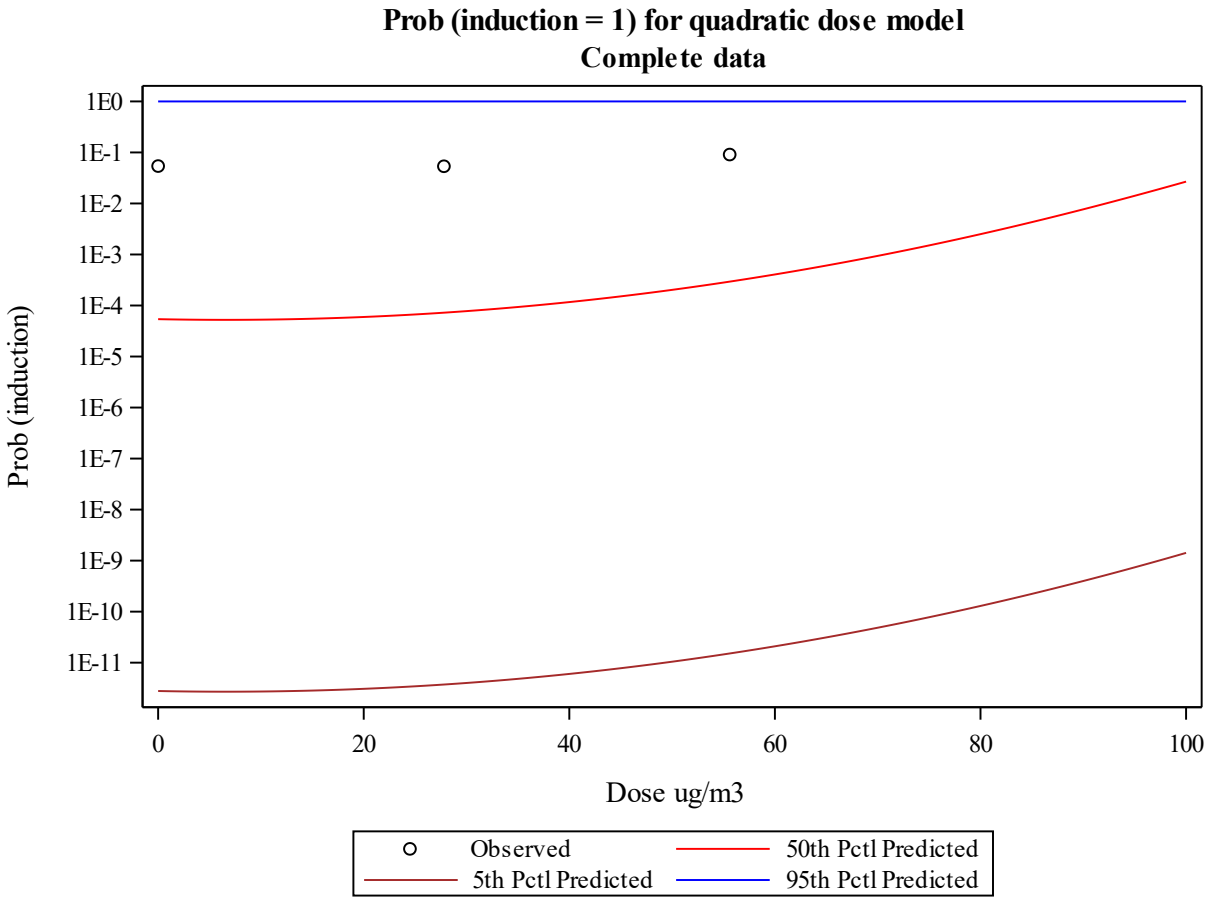


Figure 2. Predicted median and observed probability of a positive induction response based on the quadratic dose model, together with a 90% uncertainty interval. Plotted on a logarithmic scale. Complete data.

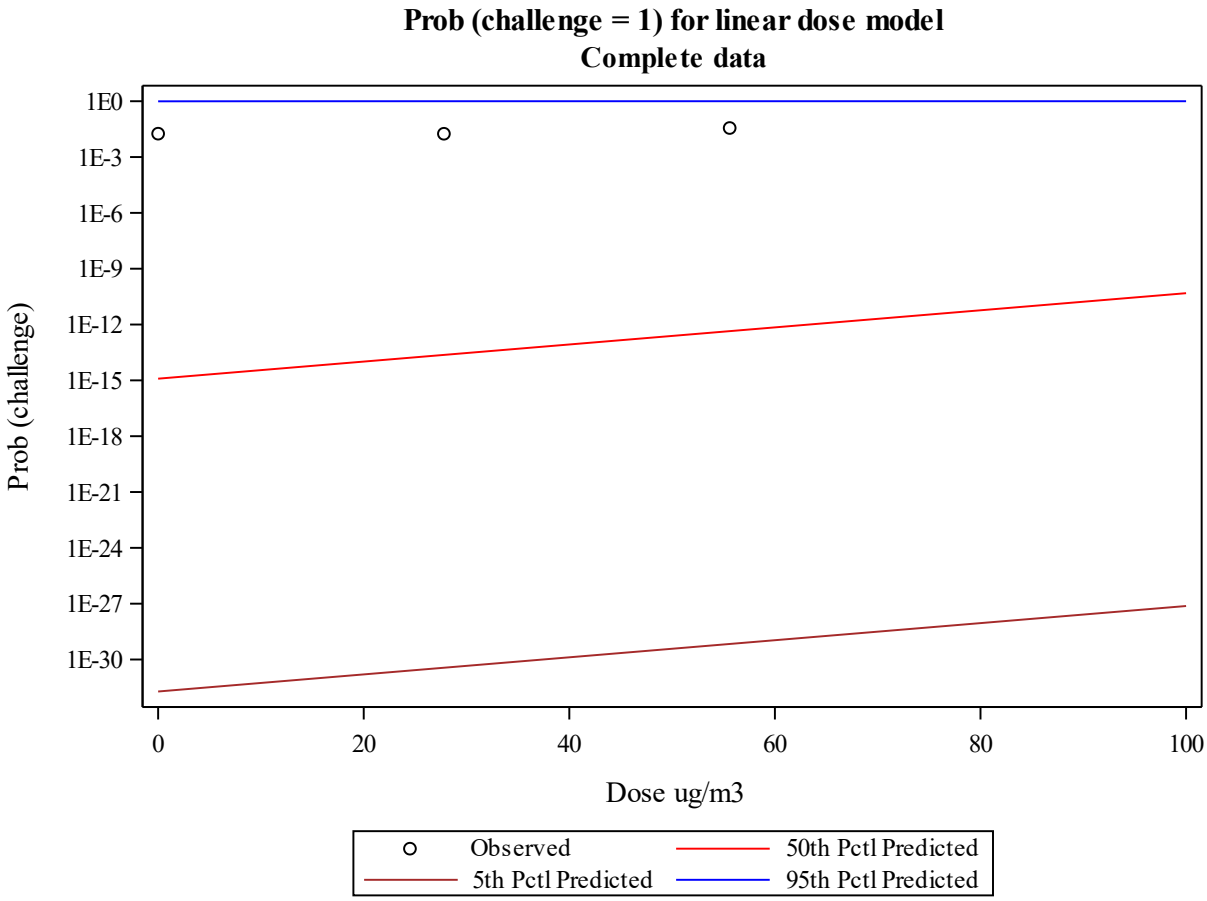


Figure 3. Predicted median and observed probability of a positive challenge response based on the linear dose model, together with a 90% uncertainty interval. Plotted on a logarithmic scale. Complete data.

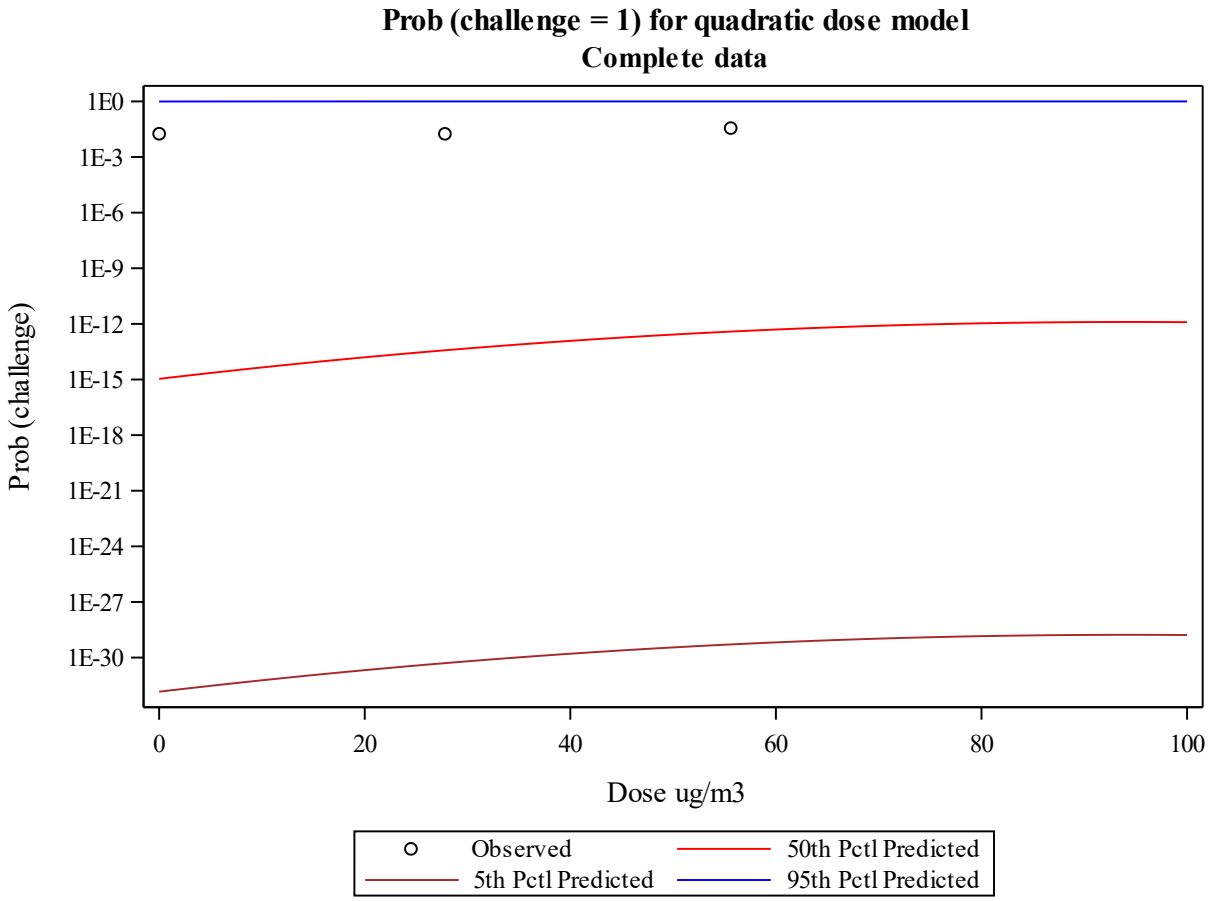


Figure 4. Predicted median and observed probability of a positive challenge response based on the quadratic dose model, together with a 90% uncertainty interval. Plotted on a logarithmic scale. Complete data.

Statistical Analyses for All Subjects

The analyses in this section use data from all 120 subjects with any data, thus excluding the subject that did not have any induction or challenge test results. Some of these subjects did not complete all the induction tests and some did not complete any or all of the challenge tests. The details of the statistical methods are exactly the same as in section 2 and therefore are not repeated in this section. The detailed results are, of course, different.

Proportions of Positive Test Results

Table 9 shows the number and proportion of positive test results for each combination of dose level and test type.

Table 9. Numbers and proportions of positive test results by dose level and test type. All subjects.

Dose Level	Induction: Number of Positive Tests	Induction: Number of Subjects	Induction: Positivity Percentage	Challenge: Number of Positive Tests	Challenge: Number of Subjects	Challenge: Positivity Percentage
High	6	60	10.0%	2	58	3.4%
Low	3	60	5.0%	1	56	1.8%
Non-zero	9	120	7.5%	3	114	2.6%
Vehicle	6	120	5.0%	2	114	1.8%

McNemar Test for the Agreement Between the Induction and Challenge Tests

Table 10 shows the number and proportion of positive test results for each dose level and test type for the subjects that did both induction and challenge tests. Table 10 also shows the p-value for the McNemar test that the probabilities of a positive test result are the same for both test types. The tabulated results in Tables 9 and 10 are not the same.

Table 10. McNemar test P-values and numbers and proportions of positive test results by dose level and test type for subjects that did both induction and challenge tests. All subjects.

Dose Level	Induction: Number of Positive Tests	Induction: Number of Subjects	Induction: Positivity Percentage	Challenge: Number of Positive Tests	Challenge: Number of Subjects	Challenge: Positivity Percentage	P-Value
High	6	58	10.3%	2	58	3.4%	0.29
Low	3	56	5.4%	1	56	1.8%	0.50
Non-zero	9	114	7.9%	3	114	2.6%	0.11
Vehicle	6	114	5.3%	2	114	1.8%	0.22

The McNemar test for each dose level compares the positivity probabilities for the induction and challenge test. For example, for the High dose level, the positivity percentage was 10.3% for the induction and 3.4% for the challenge. The p-values are shown in the final column. Since all the p-values are greater than 0.05, for each dose level there is insufficient evidence that the probabilities of a positive result for the induction and challenge tests are different.

Fisher's Exact Test Comparing the Positivity Probabilities Between the High and Low Doses

For the induction tests, as shown in Table 9, the positivity percentages were 10.0% at the High dose and 5.0% at the Low dose. Table 9 also shows the results for the challenge tests. Fisher's exact test can be

used to test whether the corresponding probabilities of a positive result are the same at both doses. Note that the results for the Low and High doses can be assumed to be statistically independent because those two doses were tested on different subjects. The p-values for the induction and challenge tests are shown in the following table.

Table 11. P-values from Fisher's exact test for differences in the probability of a positive result between the High and Low dose levels. By test type. All subjects.

Test Type	P-value (Fisher's Exact Test)
Induction	0.49
Challenge	1.00

Since both p-values are greater than 0.05, for both test types there is insufficient evidence that the probabilities of a positive result for the Low and High dose levels are different.

Fisher's Exact Test Comparing the Positivity Probabilities Between the High, Low, and Zero (Vehicle Only) Doses

For the induction tests, as shown in Table 9, the positivity percentages were 10.0% at the High dose, 5.0% at the Low dose, and 5.0% for the Vehicle dose (i.e., 0% BIT). Table 9 also shows the results for the challenge tests. Under the assumption that the results for all three doses are statistically independent, Fisher's exact test can be used to test whether the corresponding probabilities of a positive result are the same at all three doses. The p-values for the induction and challenge tests are shown in the following table.

Table 12. P-values from Fisher's exact test for differences in the probability of a positive result between the High , Low, and Vehicle dose levels. By test type. Assumes independence. All subjects.

Test Type	P-value (Fisher's Exact Test)
Induction	0.42
Challenge	0.84

Since both p-values are greater than 0.05, for both test types there is insufficient evidence that the probabilities of a positive result for the High, Low and Vehicle dose levels are different.

Statistical Models for the Probability of a Positive Result

In this final subsection we develop and fit statistical models for the probability of a positive result (i.e., a response), separately for the induction and challenge tests. The probability is a function of both the dose (a fixed effect) and the subject (a normally distributed random effect).

For the analysis of all the data in this Section 3, we used the preferred integration method of adaptive quadrature with qmax = 100.

Fitted models

The estimated parameters of the fitted models for induction together with their standard errors and p-values are given in Tables 13 and 14 for the linear dose and quadratic dose models, respectively. The estimated parameters of the fitted models for challenge together with their standard errors and p-values are given in Tables 15 and 16 for the linear dose and quadratic dose models, respectively.

Table 13. Parameter estimates for the linear dose mixed model for the probability of an induction response. All subjects.

Parameter	Estimate	Standard Error	P-value
alpha	-10.3539	1.9819	< 0.0001
beta	0.0410	0.0249	0.1020
logs2u	4.6495	0.6726	< 0.0001

Table 14. Parameter estimates for the quadratic dose mixed model for the probability of an induction response. All subjects.

Parameter	Estimate	Standard Error	P-value
alpha	-10.2543	2.0331	< 0.0001
beta	-0.0073	0.1288	0.9552
gamma100	0.0879	0.2323	0.7057
logs2u	4.6500	0.6791	< 0.0001

Table 15. Parameter estimates for the linear dose mixed model for the probability of a challenge response. All subjects.

Parameter	Estimate	Standard Error	P-value
alpha	-30.4867	13.2035	0.0228
beta	0.2731	0.1651	0.1009
logs2u	9.0766	3.2942	0.0068

Table 16. Parameter estimates for the quadratic dose mixed model for the probability of a challenge response. All subjects.

Parameter	Estimate	Standard Error	P-value
alpha	-30.4862	13.2048	0.0228
beta	0.2394	12.2777	0.9845
gamma100	0.0607	22.0795	0.9978
logs2u	9.0766	3.2942	0.0068

For the linear dose models tabulated in Tables 13 and 15, the p-values for the slope parameter beta are 0.10 and 0.10, respectively, which implies that the dose effect is not statistically significant (at the 5% level) for both induction and challenge. For the quadratic dose models tabulated in Tables 14 and 16, the p-values for the quadratic parameter gamma100 are 0.71 and 0.998, respectively, which implies that the quadratic dose effects are not statistically significant (at the 5% level) for both induction and challenge, so that the linear dose models should be preferred.

The predicted and observed probabilities for induction are plotted in Figures 5 and 6 for the linear dose and quadratic dose models, respectively. The predicted and observed probabilities for challenge are plotted in Figures 7 and 8 for the linear dose and quadratic dose models, respectively.

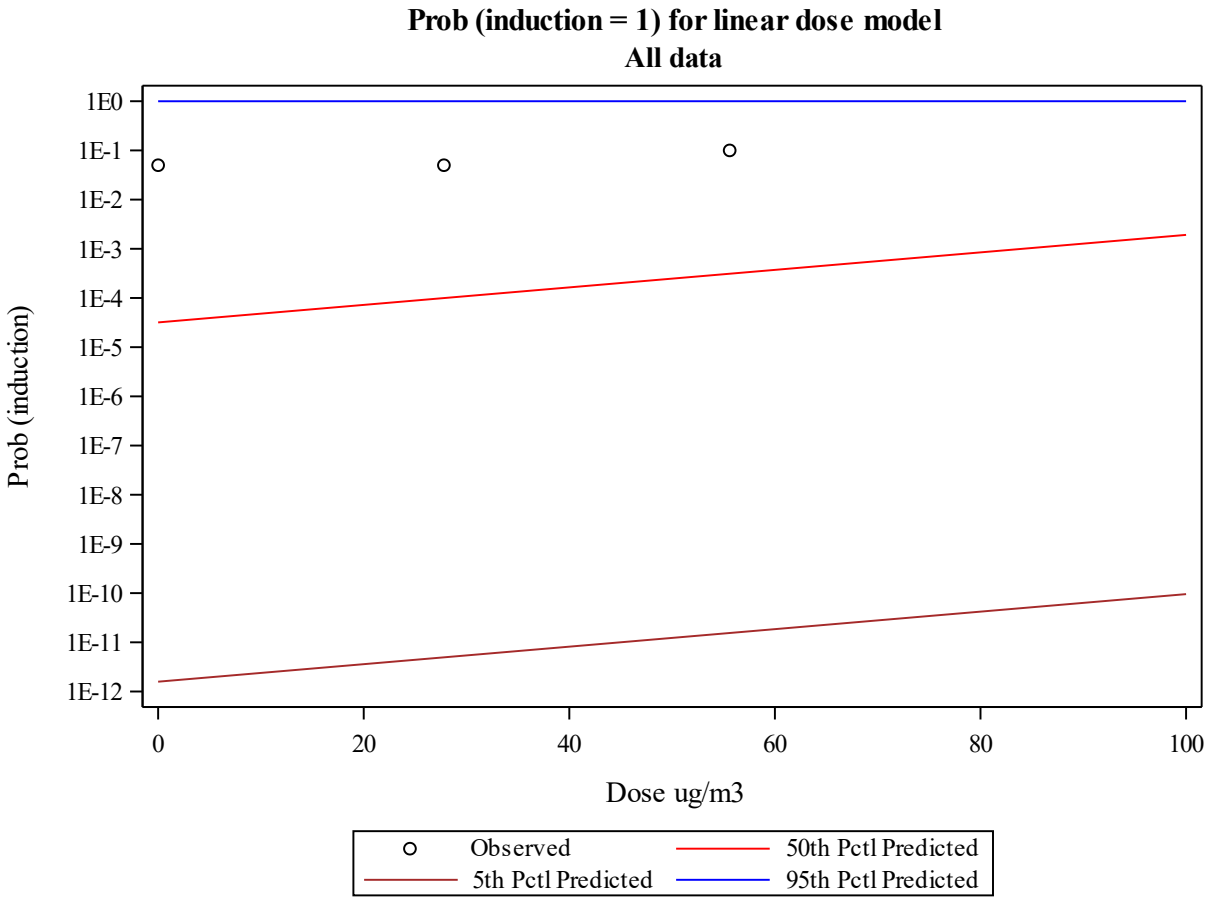


Figure 5. Predicted median and observed probability of a positive induction response based on the linear dose model, together with a 90% uncertainty interval. Plotted on a logarithmic scale. All subjects.

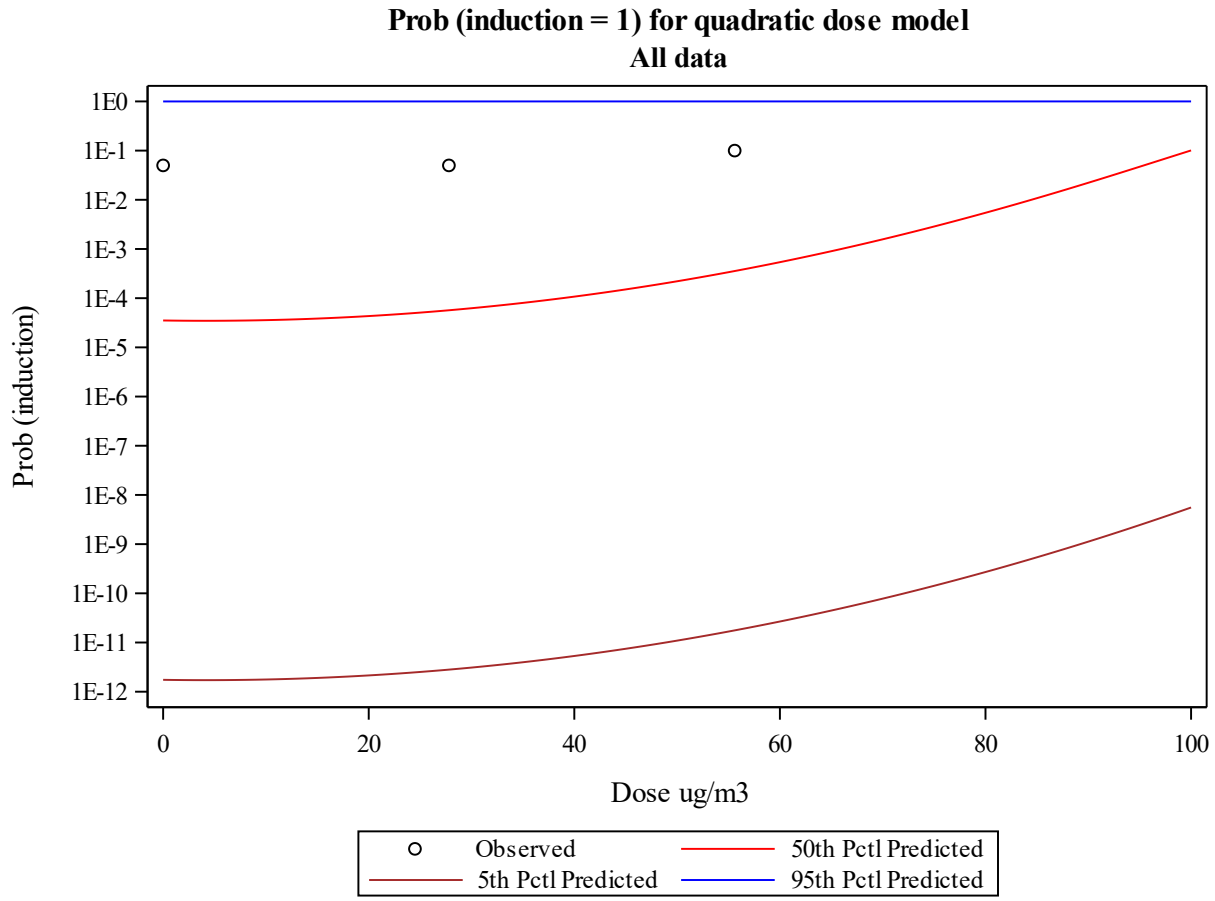


Figure 6. Predicted median and observed probability of a positive induction response based on the quadratic dose model, together with a 90% uncertainty interval. Plotted on a logarithmic scale. All subjects.

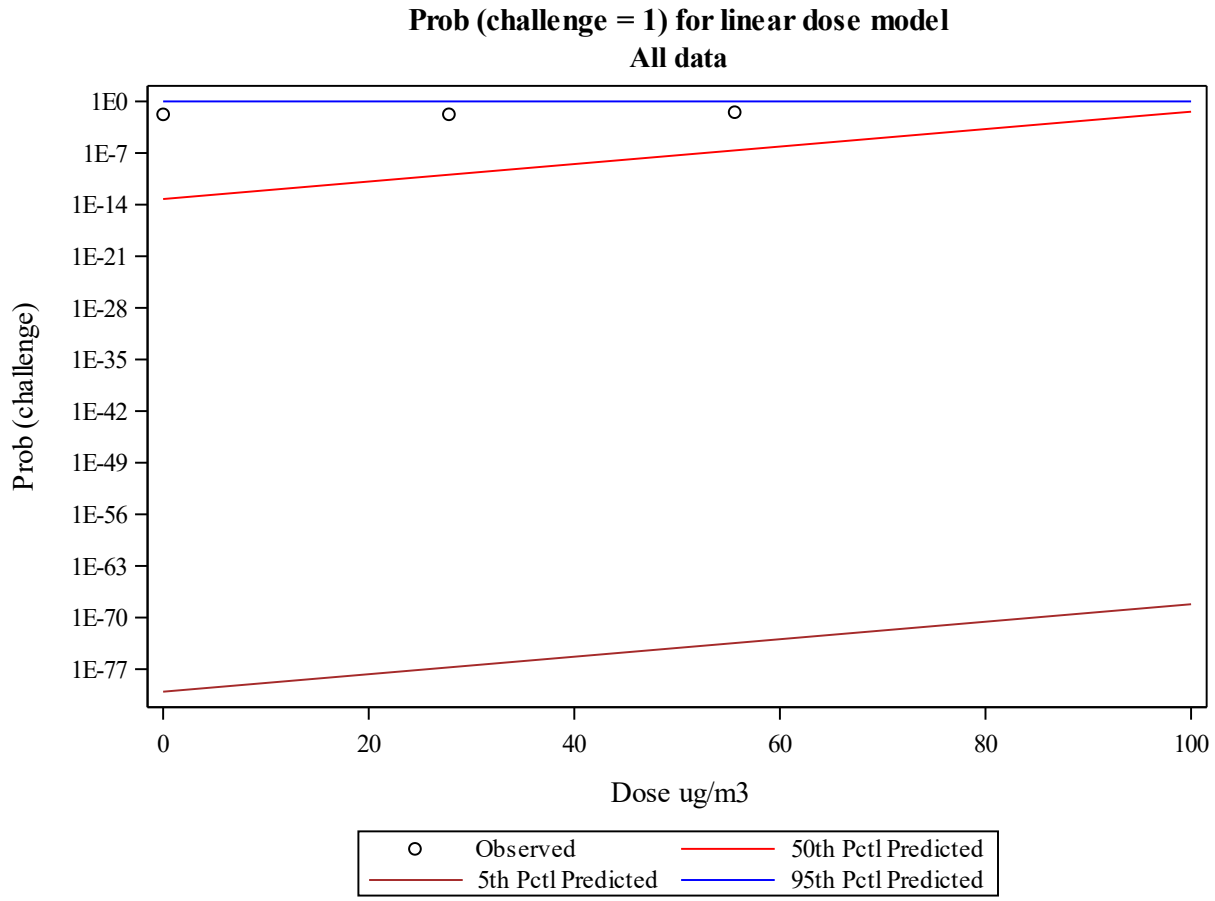


Figure 7. Predicted median and observed probability of a positive challenge response based on the linear dose model, together with a 90% uncertainty interval. Plotted on a logarithmic scale. All subjects.

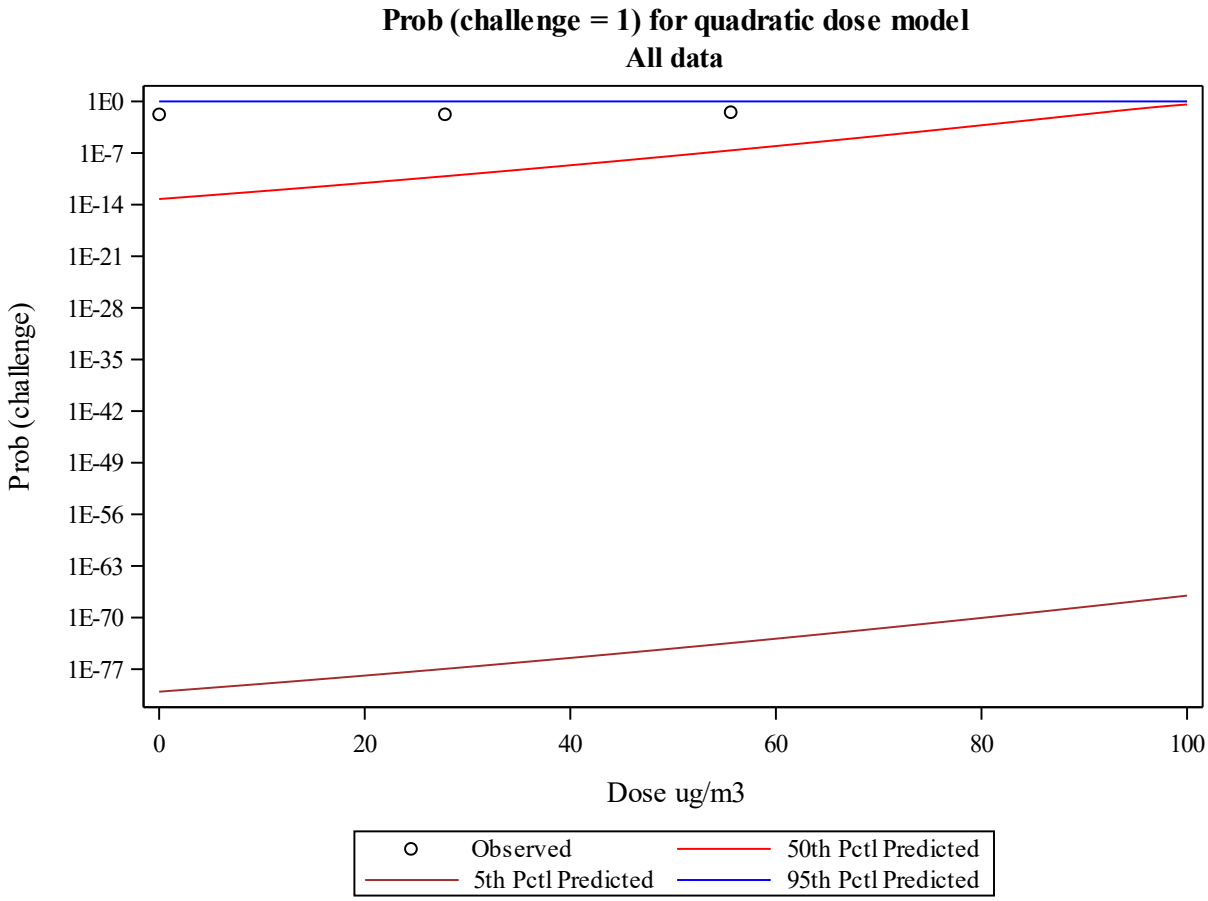


Figure 8. Predicted median and observed probability of a positive challenge response based on the quadratic dose model, together with a 90% uncertainty interval. Plotted on a logarithmic scale. All subjects.

References

Ladics, G. S (2020). "BIT: Repeated insult patch test". Dupont Nutrition & Biosciences Report NB-2000034-1730, MRID 51171302, May 29, 2020.