Original Article

Developing a Prediction Model for Determination of Peanut Allergy Status in the Learning Early About Peanut Allergy (LEAP) Studies

Michelle L. Sever, PhD[a,b], Agustin Calatroni, MS[b], Graham Roberts, DM[c,d], George du Toit, MB BCh[e,f,g], Henry T. Bahnson, MS[h], Suzana Radulovic, MD[e,f,g], David Larson, PhD[i], Margie Byron, MS[b], Alexandra F. Santos, MD, PhD[j,k,l], Michelle F. Huffaker, MD[m], Lisa M. Wheatley, MD, MPH[n], and Gideon Lack, MB BCh[e,f,g] *Morrisville and Durham, NC; Seattle, Wash; Bethesda, Md; San Francisco, Calf; and Southampton, Isle of Wight, and London, United Kingdom*

What is already known about this topic? There is a need for diagnostic tools for predicting food allergy status in the absence of an oral food challenge (OFC). Previous work has largely used sequential algorithms that have limitations in specificity or result in a significant number of nonevaluable results.

What does this article add to our knowledge? A logistic regression model using skin test and serum IgE values outperforms a traditional clinical algorithm to predict peanut allergy in the absence of an OFC.

How does this study impact current management guidelines? Statistical models could be used in the future as an adjunct to help support clinical decision making, prioritization, and risk assessment before OFC.

[a]PPD Government and Public Health Services, Morrisville, NC
[b]Rho Federal Systems Division, Durham, NC
[c]University of Southampton and Southampton NIHR Biomedical Research Centre, Southampton, United Kingdom
[d]David Hide Centre, Isle of Wight, United Kingdom
[e]Peter Gorer Department of Immunobiology, School of Immunology & Microbial Sciences, London, United Kingdom
[f]Pediatric Allergy Group, Department of Women and Children's Health, School of Life Course Sciences, King's College London, London, United Kingdom
[g]Children's Allergy Service, Guy's and St Thomas' NHS Foundation Trust, London, United Kingdom
[h]The Immune Tolerance Network, Benaroya Research Institute at Virginia Mason, Seattle, Wash
[i]The Immune Tolerance Network, Bethesda, Md
[j]Department of Pediatric Allergy, School of Life Course Sciences, King's College London and Guy's and St Thomas' Hospital NHS Foundation Trust, London, United Kingdom
[k]Peter Gorer Department of Immunobiology, School of Immunology & Microbial Sciences, King's College London, London, United Kingdom
[l]MRC and Asthma UK Centre in Allergic Mechanisms of Asthma, London, United Kingdom
[m]Immune Tolerance Network, University of California San Francisco, San Francisco, Calf
[n]Division of Allergy, Immunology, and Transplantation, National Institute of Allergy and Infectious Diseases, Bethesda, Md

This research was performed as a project of the Immune Tolerance Network, an international clinical research consortium headquartered at the Benaroya Research Institute, and supported by the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under award no. UM1AI109565. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflicts of interest: G. Roberts reports grants from the National Institute of Allergy and Infectious Diseases (NIAID, National Institutes of Health [NIH]). G. du Toit reports grants from NIAID (NIH), Food Allergy & Research Education (FARE), MRC & Asthma UK Centre, UK Department of Health through the National Institute for Health and Care Research (NIHR), and Action Medical Research and National Peanut Board; is a scientific advisory board member of Aimmune; is an investigator on pharma-sponsored allergy studies (Aimmune and DBV Technologies); and is a scientific advisor to Aimmune, DBV, and Novartis. H. T. Bahnson reports contract work paid to the institution, Benaroya Research Institute, from DBV Technologies, MYOR, King's College London, and Stanford University; and additional salary support paid by King's College London and Stanford University. S. Radulovic reports salary support from grants from NIAID (NIH). D. Larson reports employment compensation from Horizon Therapeutics. A. F. Santos reports grants from the Medical Research Council, the National Institute for Health Research, and NIAID; grants pending with Asthma UK, Medical Research Council, Biotechnology and Biological Sciences Research Council, and Rosetrees Trust; consultancy from Allergy Therpeutics, Stallergenes, and IgGenix; paid speaker services for Thermofisher, Buhlmann, Infomed, Nutricia, and Nestle; and provision of reagents through collaboration with King's College London and Thermofisher and Buhlmann. G. Lack reports grants from NIAID (NIH) and other from FARE, MRC & Asthma UK Centre, UK Department of Health through NIHR, National Peanut Board (NPB), and the Davis Foundation, during the conduct of the study; is a shareholder in DBV Technologies and Mighty Mission Me; and reports personal fees from Novartis, Sanofi-Genyzme, Regeneron, ALK-Abello, and Lurie Children's Hospital, outside the submitted work. The rest of the authors declare that they have no relevant conflicts of interest.

Received for publication November 17, 2022; revised March 9, 2023; accepted for publication April 12, 2023.
Available online May 3, 2023.

Corresponding author: Michelle F. Huffaker, MD, Immune Tolerance Network, University of California, San Francisco Diabetes Center, 513 Parnassus Ave, HSW 11, Box 0534, San Francisco, CA 94143-0534. E-mail: mhuffaker@immunetolerance.org.

2213-2198

https://doi.org/10.1016/j.jaip.2023.04.032

Abbreviations used
AUC- Area under the curve
BEEP- Barrier Enhancement for Eczema Prevention
CI- Confidence interval
EAT- Enquiring About Tolerance
LEAP- Learning Early About Peanut Allergy Study
MICE- Multiple Imputation by Chained Equations
OFC- Oral food challenge
OR- Odds ratio
Pn-sIgE- Peanut-specific IgE
Pn-sIgG4- Peanut-specific IgG4
SPT- Skin prick test

BACKGROUND: The Learning Early About Peanut Allergy (LEAP) study team developed a protocol-specific algorithm using dietary history, peanut-specific IgE, and skin prick test (SPT) to determine peanut allergy status if the oral food challenge (OFC) could not be administered or did not provide a determinant result.
OBJECTIVE: To investigate how well the algorithm determined allergy status in LEAP; to develop a new prediction model to determine peanut allergy status when OFC results are not available in LEAP Trio, a follow-up study of LEAP participants and their families; and to compare the new prediction model with the algorithm.
METHODS: The algorithm was developed for the LEAP protocol before the analysis of the primary outcome. Subsequently, a prediction model was developed using logistic regression.
RESULTS: Using the protocol-specified algorithm, 73% (453/617) of allergy determinations matched the OFC, 0.6% (4/617) were mismatched, and 26% (160/617) participants were nonevaluable. The prediction model included SPT, peanut-specific IgE, Ara h 1, Ara h 2, and Ara h 3. The model inaccurately predicted 1 of 266 participants as allergic who were not allergic by OFC and 8 of 57 participants as not allergic who were allergic by OFC. The overall error rate was 9 of 323 (2.8%) with an area under the curve of 0.99. The prediction model additionally performed well in an external validation cohort.
CONCLUSION: The prediction model performed with high sensitivity and accuracy, eliminated the problem of nonevaluable outcomes, and can be used to estimate peanut allergy status in the LEAP Trio study when OFC is not available. **(J Allergy Clin Immunol Pract 2023;11:2217-27)**

Key words: *Food allergy; Peanut allergy; Prevention; LEAP; Diagnostic algorithm*

The Learning Early About Peanut Allergy (LEAP) study showed that sustained consumption of peanut beginning between age 4 and 11 months among infants at high risk for food allergy resulted in an 81% lower rate of peanut allergy at age 60 months than among children who avoided peanuts.[1] The subsequent Persistence of Oral Tolerance to Peanut (LEAP-On) study showed that the prevention of peanut allergy persisted at 72 months of age after 12 months of not eating peanuts.[2] This same population was followed to 144 months of age after a prolonged period of *ad lib* consumption, LEAP Trio study (ClinicalTrials.gov NCT03546413). The LEAP Trio study is evaluating the original LEAP study participants at 12 to 16 years of age after a further 6 to 10 years of *ad lib* peanut consumption to see whether the protective effect of early peanut consumption remains; LEAP Trio will also be examining the allergic status of the siblings and parents. Missing data for oral food challenge (OFC), peanut-specific IgE (Pn-sIgE), and skin prick test (SPT) to peanut are expected in LEAP Trio at a higher rate than observed in LEAP or LEAP-On, particularly in the context of the COVID-19 pandemic.

In both previous studies (LEAP and LEAP-On), the primary outcome of peanut allergy was determined using an OFC in more than 90% of participants. Participants not expected to be allergic based on the peanut SPT and food reaction history received 5 g of peanut protein in a single dose. A double-blind, placebo-controlled food challenge was conducted for other participants with incremental doses (total of 9.4 g of peanut protein) as previously described.[1,3] Among those who did not have an outcome determined by OFC, peanut allergy was determined by a clinical algorithm specified in the protocol. The algorithm was developed based on clinical expertise and the relevant literature at the time. The algorithm incorporated dietary history, peanut SPT wheal size, and Pn-sIgE values.

In this analysis, we investigated the accuracy of the original LEAP protocol-specified clinical algorithm in predicting peanut allergy assessed by the gold standard OFC. We then developed a new prediction model using statistical modeling techniques. The new prediction model is intended to be used to assign peanut allergy status when OFC results are not available, primarily in the LEAP Trio study. The prediction model will provide an *a priori* method for imputing peanut allergy status when an OFC outcome is not available. LEAP Trio completed enrollment in August 2022 and results are expected early 2023.

We validated the model with the LEAP-On study and an external cohort with differing characteristics to determine if the model can be applied in other studies with appropriate consideration of population differences.

METHODS

Study population, design, and procedures

This is a secondary analysis using data from the LEAP study, which has been previously published.[1] The LEAP study enrolled infants $\geq$4 to $<$11 months of age with severe eczema and/or egg allergy. Participants were randomly assigned to avoid or consume peanut. Assessments, including Pn-sIgE to peanut and SPT measurements, were undertaken at enrollment and at age 12, 30, and 60 months. Duplicate skin tests were used in the LEAP studies, and the SPT wheal diameter was averaged. If the difference exceeded an expected amount, a third SPT was performed and the highest 2 results were used in an interest of safety. Peanut allergy was assessed at 60 months of age and was determined in 617 participants through an OFC (Figure 1). Among 11 (1.7%) study participants for whom data from the OFC were either inconclusive or not available, peanut allergy status was determined using the original LEAP protocol–specified algorithm (Figure 2) based on clinical history, SPT, and values for Pn-sIgE. The LEAP study was approved by the institutional review board (the National Research Ethics Service Committee London–Fulham).

Evaluation of the protocol-specified algorithm

To evaluate the accuracy of the protocol-specified algorithm, we compared participants' allergy status as determined by the OFC at

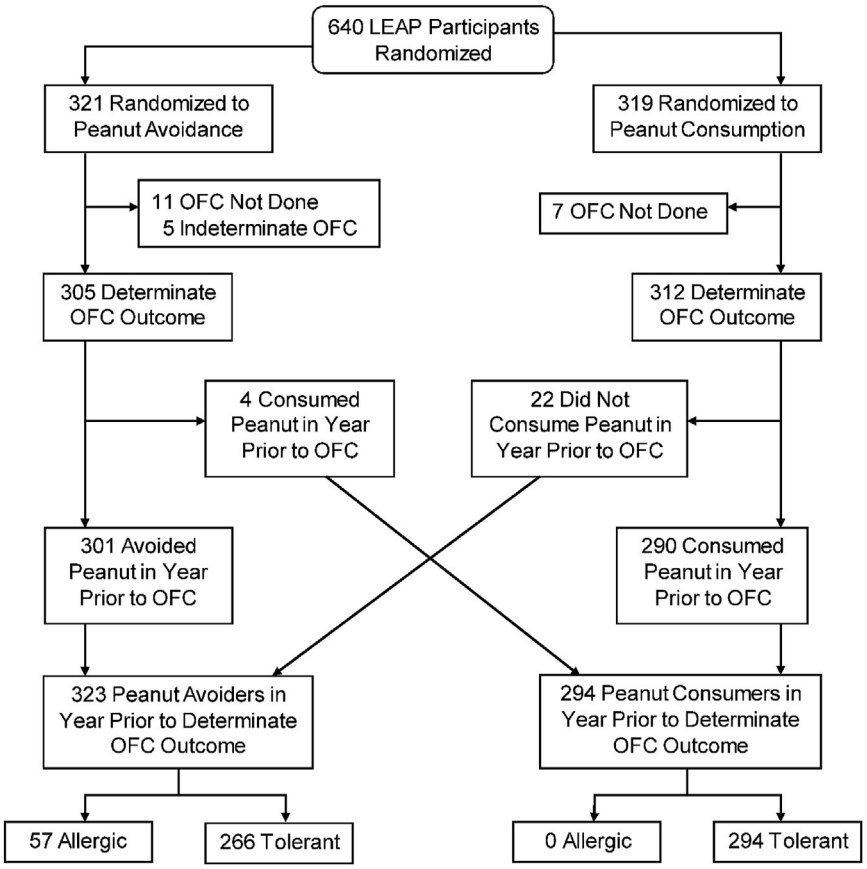

FIGURE 1. CONSORT (Consolidated Standards of Reporting Trials) diagram. *LEAP*, Learning Early About Peanut Allergy; *OFC*, oral food challenge.

the end of the LEAP study (n = 617) to the status of the same participant as determined by the protocol-specified algorithm. The first step in this algorithm (Figure 2) grouped participants by allergic symptoms related to peanut exposure at any time. Those who had not experienced allergic symptoms were then grouped by whether they had been exposed to no more than a trace amount of peanut protein (0.25 g in a single exposure) in the previous 12 months. Each group was then subdivided based on peanut SPT wheal size and then by Pn-sIgE to determine peanut allergy status as allergic, tolerant, or nonevaluable. For example, if a participant had not experienced symptoms related to peanut exposure, had no more than trace exposure in the previous 12 months, and had SPT wheal <3 mm and Pn-sIgE <0.35 kU/L, they would have been classified as tolerant (not allergic). If that same participant had Pn-sIgE >0.35 kU/L, they would have been classified as nonevaluable. The accuracy of this algorithm was also evaluated without the Pn-sIgE criteria.

Development of the prediction model

We developed a statistical model to predict the outcome (allergic/nonallergic) of a peanut OFC using 2 different populations: the 617 LEAP participants with a determinate OFC and the subset of 323 participants who did not consume peanut in the year before the OFC (Figure 1). This subset consisted of the 301 LEAP participants randomized to peanut avoidance who had a determinate OFC outcome at the end of LEAP as well as 22 LEAP participants randomized to the consumption group who did not consume peanut in the year before the OFC. In constructing the prediction model, all participants who were avoiding peanut were included, specifically those who did not consume 2 g of peanut per week or on any single occasion.

We considered biomarkers known to be diagnostic of peanut allergy that could be determined easily and reliably: SPT (mm), Pn-sIgE (kU/L), Ara h 1 (kU/L), Ara h 2 (kU/L), and Ara h 3 (kU/L). Peanut-specific IgG4 (Pn-sIgG4) was not included because of the lack of consistent evidence that it is diagnostic of peanut allergy[4-8] as well as the uncertainty of its longitudinal trajectory after 72 months of age (Figure E1, available in this article's Online Repository at www.jaci-inpractice.org). Moreover, IgG4 had the lowest relative importance for prediction compared with all other variables (Figure E2, available in this article's Online Repository at www.jaci-inpractice.org). All variables were treated as continuous variables. A square root transformation for SPT and log10 transformation for the other variables were applied to reduce the degree of skewness. Distributions with heavy tails, such as those observed in the LEAP data, can have an undue influence on the model coefficient estimates resulting in a worse fit. Among the subset (n = 323), 8 (2.5%) participants are missing data for Pn-sIgE, which was assigned a value using the MICE (Multiple Imputation by Chained Equations) predictive mean matching algorithm using the other model patient characteristics.[9] All 8 participants were randomized to peanut avoidance, had a 0 mm peanut SPT wheal, and had a negative OFC at the end of LEAP. Sixty-four (20%) participants did not have Ara h component measurements because specimens were

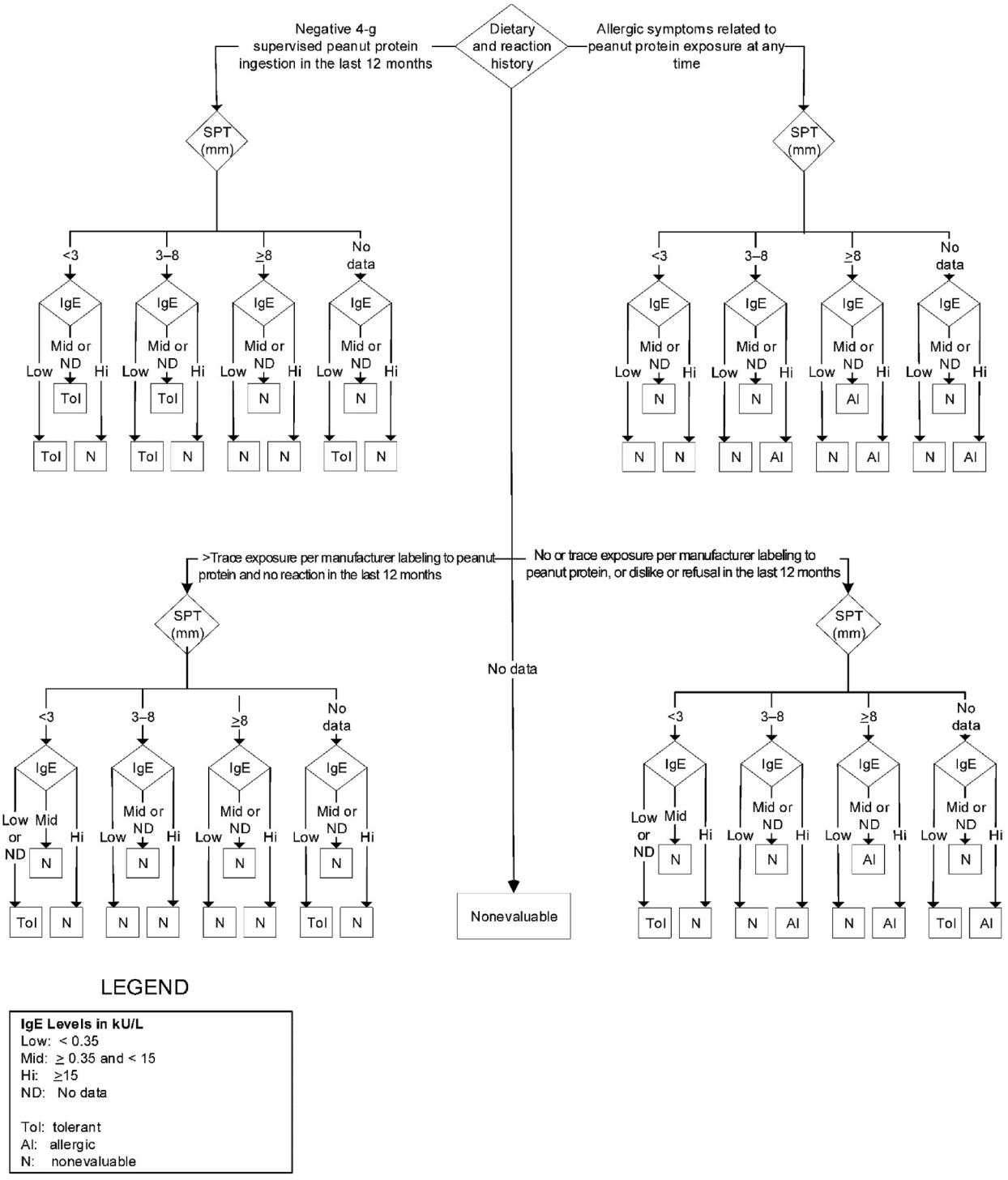

FIGURE 2. LEAP protocol-specified algorithm. *LEAP,* Learning Early About Peanut Allergy; *SPT,* skin prick test.

not assayed for peanut components if participants did not have at least 1 Pn-sIgE measurement greater than 0.1 kU/L. Values for Ara h 1, 2, and 3 for these participants were filled in as the lowest available value from the assay (0.01 kU/L) if Pn-sIgE was less than or equal to 0.1 kU/L. Finally, those participants with Pn-sIgE greater than 0.1 kU/L but with peanut-specific components missing (approximately 15%) were assigned a value using the MICE algorithm described above. The inclusion of this small number of these assigned values did not alter the median and ranges for Pn-sIgE or SPT, nor did it change the predictive behavior of the model (data not shown).

Logistic regression analysis was performed with OFC as a dichotomous outcome and the aforementioned variables in an additive model. The discriminatory power of the model was investigated by several metrics and calculated in a 5-fold cross-validation to control for

TABLE I. Characteristics of participants by oral food challenge results

		Peanut avoidance			
Characteristics	**All (N = 617)**	**Overall (N = 323)**	**Not allergic (N = 266)**	**Allergic (N = 57)**	***P* value***
Sex, n (%)					.25
Male	370 (60)	211 (65)	170 (64)	41 (72)	
Female	247 (40)	112 (35)	96 (36)	16 (28)	
Primary ethnicity, n (%)					.028
White	455 (74)	241 (75)	206 (77)	35 (61)	
Mixed	83 (13)	41 (13)	31 (12)	10 (18)	
Black	47 (7.6)	27 (8.4)	21 (7.9)	6 (11)	
Asian	24 (3.9)	10 (3.1)	6 (2.3)	4 (7.0)	
Chinese, middle eastern, or other group	7 (1.1)	3 (0.9)	2 (0.8)	1 (1.8)	
Missing	1 (0.2)	1 (0.3)	0 (0)	1 (1.8)	
Age at OFC (y)					.52
Median (IQR)	5.0 (4.8, 5.2)	5.0 (4.8, 5.2)	5.0 (4.8, 5.2)	5.0 (4.8, 5.2)	
Range	4.3-6.8	4.3-6.5	4.3-6.5	4.5-6.0	
Treatment group, n (%)					.007
Peanut avoidance	305 (49)	301 (93)	253 (95)	48 (84)	
Peanut consumption	312 (51)	22 (6.8)	13 (4.9)	9 (16)	
Peanut wheal size (mm)					<.001
Median (IQR)	0.0 (0.0, 0.0)	0.0 (0.0, 2.0)	0.0 (0.0, 0.0)	9.0 (6.0, 12.0)	
Range	0.0-18.0	0.0-18.0	0.0-8.0	0.0-18.0	
Missing	1				
Peanut-specific IgE (kU/L)					<.001
Median (IQR)	0.08 (0.02, 0.54)	0.07 (0.01, 1.00)	0.04 (0.01, 0.22)	3.75 (1.06, 36.30)	
Range	0.01-363	0.01-363	0.01-21	0.14-363	
Missing	24	8	8	0	
Peanut-specific IgG4 (μg/L)					.001
Median (IQR)	320 (70, 1240)	105 (70, 460)	80 (70, 352)	210 (80, 685)	
Range	70-50,300	70-22,700	70-22,700	70-19,500	
Missing	26	9	8	1	
IgG4 (μg/L)/IgE (ng/mL) ratio log					<.001
Median (IQR)	3.27 (2.51, 3.74)	2.99 (2.13, 3.46)	3.18 (2.56, 3.46)	1.46 (0.81, 2.02)	
Range	−0.38 to 5.31	−0.38 to 4.89	0.67 to 4.89	−0.38 to 3.03	
Missing	26	9	8	1	
Ara h 1 (kU/L)					<.001
Median (IQR)	0.01 (0.01, 0.03)	0.01 (0.01, 0.04)	0.01 (0.01, 0.02)	0.16 (0.01, 4.26)	
Range	0.01-279.0	0.01-279.0	0.01-1.3	0.01-279.0	
Missing	137	64	58	6	
Ara h 2 (kU/L)					<.001
Median (IQR)	0.01 (0.01, 0.04)	0.01 (0.01, 0.06)	0.01 (0.01, 0.02)	1.75 (0.28, 18.30)	
Range	0.01-456.0	0.01-456.0	0.01-0.5	0.01-456.0	
Missing	137	64	58	6	
Ara h 3 (kU/L)					<.001
Median (IQR)	0.01 (0.01, 0.04)	0.01 (0.01, 0.04)	0.01 (0.01, 0.03)	0.08 (0.01, 0.45)	
Range	0.01-92.1	0.01-92.1	0.01-5.4	0.01-92.1	
Missing	137	64	58	6	
Ara h 1 (kU/L) attributed					<.001
Median (IQR)	0.01 (0.01, 0.02)	0.01 (0.01, 0.03)	0.01 (0.01, 0.02)	0.19 (0.01, 4.05)	
Range	0.01-279.0	0.01-279.0	0.01-1.3	0.01-279.0	
Ara h 2 (kU/L) attributed					<.001
Median (IQR)	0.01 (0.01, 0.04)	0.01 (0.01, 0.05)	0.01 (0.01, 0.01)	1.75 (0.30, 38.50)	
Range	0.01-456.0	0.01-456.0	0.01-0.5	0.01-456.0	
Ara h 3 (kU/L) attributed					<.001
Median (IQR)	0.01 (0.01, 0.04)	0.01 (0.01, 0.04)	0.01 (0.01, 0.02)	0.08 (0.01, 0.46)	
Range	0.01-92.1	0.01-92.1	0.01-5.4	0.01-92.1	

Peanut wheal size, peanut-specific IgE, peanut-specific IgG4, and Ara h 1, 2, and 3 were all assessed at the time of the OFC.
IQR, Interquartile range; *OFC*, oral food challenge.
*Pearson χ^2 test, Fisher exact test, and Wilcoxon rank-sum test.

TABLE II. Comparison of protocol-specified algorithm determination of allergy status with peanut-specific IgE criteria to the OFC outcome

	Allergy status (determined by OFC)		
	Not allergic	Allergic	Total
Allergy status (protocol-specified algorithm), n (%)			
Not allergic	417 (74.5)	0 (0)	417 (68)
Allergic	4 (0.7)	36 (63.2)	40 (6)
Not evaluable	139 (24.8)	21 (36.8)	160 (26)
Total	560	57	617

OFC, Oral food challenge.

TABLE III. Comparison of protocol-specified algorithm determination of allergy status without peanut-specific IgE criteria to the OFC outcome

	Allergy status (determined by OFC)		
	Not allergic	Allergic	Total
Allergy status (protocol-specified algorithm removing IgE criteria), n (%)			
Not allergic	504 (90.0)	0 (0)	504 (81.7)
Allergic	25 (4.5)	55 (96.5)	80 (13.7)
Not evaluable	31 (5.5)	2 (3.5)	33 (5.3)
Total	560	57	617

OFC, Oral food challenge.

TABLE IV. Multivariate results for the prediction of the oral food challenge outcome

Variable	N	OR	95% CI	*P* value
Peanut wheal size (mm) Sqrt	323	10.4	3.92, 36.1	<.001
Peanut-specific IgE (kU/L) Log	323	0.62	0.15, 2.35	.49
Ara h 1 (kU/L) Log	323	0.57	0.13, 2.11	.42
Ara h 2 (kU/L) Log	323	11.5	3.22, 57.0	<.001
Ara h 3 (kU/L) Log	323	0.53	0.09, 2.24	.42

AIC: 67.6; BIC: 90.3.

This predictive model was developed for the LEAP Trio cohort, and its further use will need to be trialed in local populations before use in clinical practice. Formula: logit(OFC_Positive) = −4.03 + 2.35*√(skin test [mm]) − 0.48*log10(IgE [kU/L]) − 0.56*log10(Ara h 1 [kU/L]) + 2.44*log10(Ara h 2 [kU/L]) − 0.64*log10(Ara h 3 [kU/L]).

AIC, Akaike information criterion; *BIC*, Bayesian information criterion; *CI*, confidence interval; *LEAP*, Learning Early About Peanut Allergy; *OFC*, oral food challenge; *OR*, odds ratio.

overfitting of the data (see this article's Online Repository at www.jaci-inpractice.org). In addition, an internal and an independent external cohort of participants with peanut allergy status results were used for validation. Prediction model development and validation were constructed with logistic regression using h2o within R software version 3.4.4 (R Development Core Team, Vienna, Austria).[10,11]

Sensitivity analyses around missing data

We evaluated the prediction model under a complete case analysis based on available peanut-specific component data without assignment of values for missing data when Pn-sIgE was greater than 0.1 kU/L. The complete case data set consisted of 259 participants.

Training cross-validation

The logistic regression analysis classifies a participant as OFC positive if its probability is bigger than a cutoff. The optimal cutoff is obtained by maximizing the F1 score $\left(2 \times \frac{precison \times recall}{precision + recall}\right)$, a measure of test's accuracy, through 5-fold cross-validation of each prediction model on the training data (n = 323). The F1 score can be interpreted as the harmonic mean of the precision and recall, where the highest possible value is 1.0, indicating perfect precision and recall, and the lowest possible value is 0 if either the precision or the recall is zero.

Internal validation

The prediction model was validated with data from the same cohort obtained at 72 months of age after 12 months of peanut avoidance when the population was re-evaluated clinically for peanut allergy. The internal validation was conducted in n = 265 who had not consumed peanut and had repeat OFCs.

External validation

The prediction model was validated with data from an external cohort, previously evaluated for peanut allergy, at a tertiary pediatric allergy clinic.[12]

RESULTS

Demographics/baseline characteristics

Participants in LEAP who had a determinant OFC outcome (n = 617) and the subset who did not consume peanut during the year before the OFC (n = 323) are described in Table I. Overall, the subset was similar to the full study population; participants who were peanut allergic at 60 months of age had higher values for peanut SPT; Pn-sIgE, Pn-sIgG4, Ara h 1, 2, and 3; and a lower Pn-sIgG4:Pn-sIgE ($P < .01$).

Evaluation of the protocol-specified algorithm

Overall, 73% (453/617) of allergy determinations matched between the protocol-specified algorithm and the OFC, 0.6% (4/617) were mismatched, and 26% (160/617) participants were classified as nonevaluable by the algorithm (Table II). Of the 57 participants determined to be allergic by OFC, 36 (63%) were also allergic by the protocol-specified algorithm and 21 (37%) were nonevaluable. Of the 560 who were not allergic by OFC, 417 (75%) were also not allergic by the protocol-specified algorithm, 4 (1%) were misclassified as allergic, and 139 (25%) were nonevaluable. Further, while the false-positive rate of the algorithm was very low (0%), the false-negative rate was 10%.

When the Pn-sIgE criterion was removed from the protocol-specified algorithm, 91% (559/617) of peanut allergy determinations between the OFC and the algorithm matched, and the number of nonevaluable determinations decreased from 160 (26%) to 33 (5.3%) (Table III). However, when the Pn-sIgE criterion was removed, the number of participants who were not allergic by the OFC who would have been classified as allergic by the algorithm increased from 4 of 560 to 25 of 560.

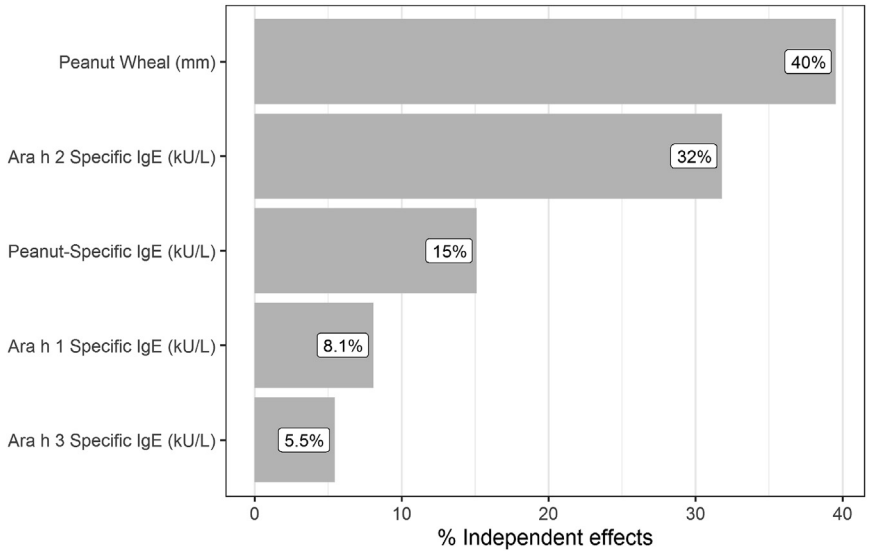

FIGURE 3. Relative importance of variables in the model for the prediction of the oral food challenge outcome.

TABLE V. Training cross-validation results of 5-variable logistic regression model for the prediction of the oral food challenge

Statistic	Result
Prediction results	
True positive	265
False positive	1
False negative	8
True negative	49
Overall	
Accuracy	0.97 (0.95, 0.99)
Balanced accuracy	0.93
Kappa	0.90
Classes	
Sensitivity (recall/true positive rate)	0.86 (0.74, 0.94)
Specificity (true negative rate)	1.00 (0.98, 1.00)
Positive predictive value (precision)	0.98 (0.89, 1.00)
Negative predictive value	0.97 (0.94, 0.99)
F1 score*	0.92 (0.80, 0.96)
Prevalence	0.18 (0.14, 0.22)
Detection prevalence	0.15 (0.12, 0.20)

The highest possible value of the F1 score is 1.0, indicating perfect precision and recall, and the lowest possible value is 0 if either the precision or the recall is zero.

*F1 score is the harmonic mean of the precision and recall $= 2 \times \frac{\text{precison} \times \text{recall}}{\text{precision} + \text{recall}}$

Prediction model development

The prediction model was initially developed using the full LEAP population (n = 617); however, all participants consuming peanut within 12 months before the OFC in LEAP passed the OFC and were classified as not allergic by the model. Future use of this model would likely only be necessary when peanut is not being consumed or consumption is unknown; therefore, the final model was developed with the subset of participants (n = 323) who did not consume peanut in the year before the OFC. All variables included in the model, square root of SPT (mm) and the log10 transformations of Pn-sIgE (kU/L), Ara h 1 (kU/L), Ara h 2 (kU/L), and Ara h 3 (kU/L), were each strongly associated with the OFC outcome ($P < .001$) (Table I). SPT (odds ratio [OR]: 10.4, 95% confidence interval [CI]: 3.92–36.1, $P < .001$) and Ara h 2 (OR: 11.5, 95% CI: 3.22-57.0, $P < .001$) (Table IV) accounted for 40% and 32% of the relative importance in predicting the OFC outcome, respectively (Figure 3).

The optimal cutoff for the classification of allergic versus nonallergic was determined to be 0.64, derived using 5-fold cross-validation and chosen for the maximal F1 (test accuracy) scores. The model with these 5 variables predicted the OFC outcome well (Table V; Figure 4). The overall error rate was 9 of 323 (2.8%) with an area under the curve (AUC) of 0.99 (Figure 4, *A*). The model inaccurately predicted 1 participant as allergic who was determined to be not allergic by the OFC representing a 0.4% false-positive rate (1/266) and 8 participants as not allergic who were determined to be allergic at OFC representing a 14% false-negative rate (8/57). The sensitivity or true positive rate was 0.86. The predicted probabilities for not allergic or allergic were clustered around 0 (not allergic) or 1 (allergic), respectively, with only 15 participants with probabilities between .25 and .75 (Figure 4, *B*), indicating a well-calibrated model. Furthermore, using bootstrap resampling, the model fit well over the range of predicted probabilities providing good calibration results (mean absolute error 0.013 based on 1000 bootstrap repetitions), especially among the OFC positive outcomes. The 9 participants whose OFC outcome was incorrectly predicted by the model are described in Table E1 (available in this article's Online Repository at www.jaci-inpractice.org), including the probabilities the model assigned for allergic and nonallergic status. These 9 participants tended to have low or discordant SPT and Ara h 2 results, and 8 of 9 were male. There were no other consistent clinical features among these 9 participants including eczema severity at the time of OFC or OFC reaction severity (data not shown).

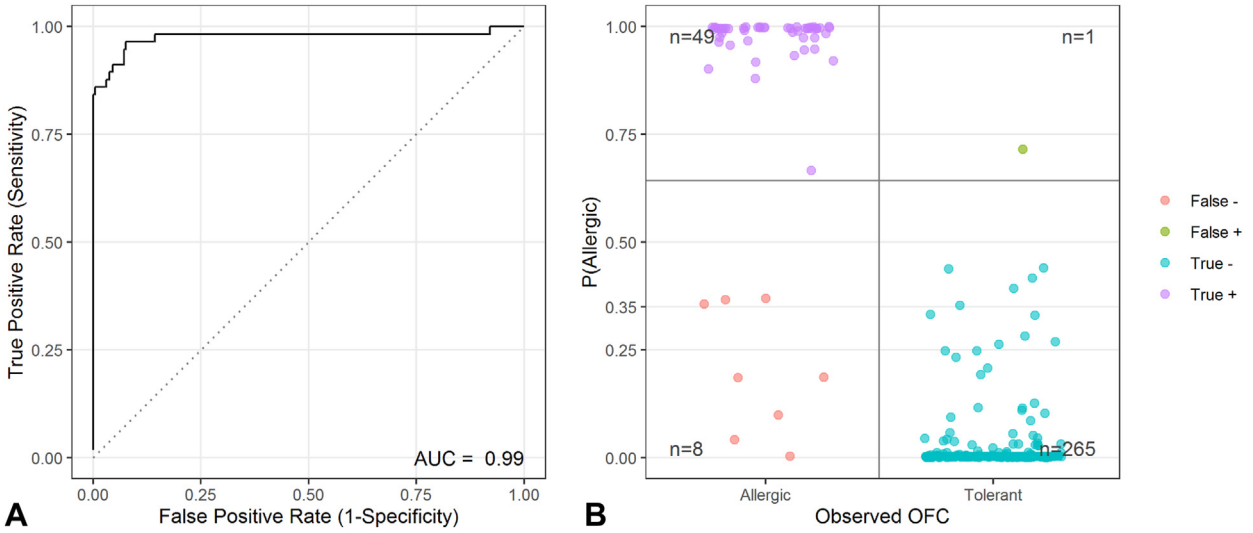

FIGURE 4. (**A**) Receiver operating characteristic curves for the prediction of oral food challenge results. (**B**) Predicted probability of the positive oral food challenge [*P*(Allergic)] versus observed oral food challenge (OFC) results. *AUC*, Area under the curve.

Sensitivity analyses

We found that including SPT, Ara h 2, or the combination of SPT, Ara h 1, Ara h 2, Ara h 3, and Pn-sIgE in the model provided similar sensitivity and accuracy (Table VI; Table E2, available in this article's Online Repository at www.jaci-inpractice.org). SPT had the highest relative importance in the full prediction model, and all results that included SPT had an accuracy of ≥0.96, a sensitivity of ≥0.84, and a specificity of ≥0.98. Adding additional variables improves the model fit. Model results that did not include SPT had slightly lower accuracy (0.95-0.96), sensitivity (0.79-0.82), and specificity (0.98-1.00). Without SPT, the model with Ara h 1, Ara h 2, Ara h 3, and Pn-sIgE performed well with an accuracy of 0.96, a sensitivity of 0.82, and a specificity of 0.98.

Prediction model internal validation

To assess internal validity, the prediction model was used with data from the LEAP-On study (assessment 12 months after the LEAP end point).[2] The characteristics of this cohort (n = 265) are described in Table E3 in this article's Online Repository at www.jaci-inpractice.org and are similar to the LEAP subset used to develop the model. The overall error rate was 2.6% (7/265), with an AUC of 0.99 (Figure E3 and Table E4, available in this article's Online Repository at www.jaci-inpractice.org). The model inaccurately predicted 1 participant as allergic who was determined to be not allergic by the OFC representing a 0.4% (1/231) false-positive rate and 6 participants as not allergic out of 34 who were determined to be allergic at OFC (18% false-negative rate). The sensitivity or true positive rate was 0.82.

Prediction model external validation

To assess external validity, the prediction model was run with data from a study conducted by Santos et al.[12] The characteristics of this cohort (n = 60), who were being evaluated for clinical suspicion of peanut allergy and had nonequivocal evaluations, are described in Table E5 in this article's Online Repository at www.jaci-inpractice.org, with this cohort having higher values for SPT, Pn-sIgE, and Ara h 2, and a lower Pn-sIgG4:Pn-sIgE ($P < .01$) than the LEAP subset used to develop the model. The overall error rate was 7 of 60 (12%), with an AUC of 0.98, a sensitivity of 0.88, and a specificity of 0.88 (Figure E4 and Table E6, available in this article's Online Repository at www.jaci-inpractice.org). The model inaccurately predicted 4 participants as allergic who were determined as not allergic representing an 11.7% (4/34) false-positive rate and 3 participants as not allergic who were determined to be allergic (11.5% false-negative rate).

Positive and negative predictive values in study populations

Positive predictive value provides information about the probability that a patient with a positive OFC is actually allergic by our prediction and resulted in 0.98, 0.97, and 0.85 for the training, internal, and external validation data, respectively. Similarly, negative predicted value is the probability that a participant with a negative OFC is defined as nonallergic by the algorithm and resulted in 0.97, 0.97, and 0.91 for training, internal, and external validation data, respectively.

DISCUSSION

The OFC has long been considered, and remains, the gold standard method for determining the presence of food allergy.[3,13-15] There are many barriers to obtaining an OFC, both in clinical practice and in the context of research studies, and alternative means of determining allergy status are needed. Statistical models using biomarkers can accurately predict allergic status when OFCs are not available.

Here we present a highly accurate prediction model that provides an *a priori* method for imputing peanut allergy status for study participants when an OFC outcome is not available. This model will be beneficial for the LEAP Trio study and potentially other peanut allergy studies. For the LEAP Trio study, no assignment of values will be made and only available variables (SPT, Pn-sIgE, Ara h 1, Ara h 2, and Ara h 3) for a given participant will be used in the model. Participants may be missing OFC outcome data for a variety of reasons including participant family relocation, scheduling issues, loss to follow-up,

TABLE VI. External validation results of sequential logistic regression models (with and without skin test) for the prediction of peanut allergy status

			Cross-validation (n = 323)							External (n = 60)					
No. of variables	**Variables**	**AIC**	**Cutoff**	**AUC**	**F1**	**FN − FP**	**Accuracy 95% CI**	**Sensitivity**	**Specificity**	**AUC**	**F1**	**FN − FP**	**Accuracy 95% CI**	**Sensitivity**	**Specificity**
Models without skin test															
1-VAR	Ara h 2	105.1	0.50	0.94	0.86	10 − 5	0.95 (0.92-0.97)	0.82	0.98	0.93	0.86	4 − 3	0.88 (0.77-0.95)	0.85	0.91
2-VAR	Ara h 2, sIgE	106.4	0.49	0.95	0.84	11 − 6	0.95 (0.92-0.97)	0.81	0.98	0.94	0.85	4 − 4	0.87 (0.75-0.94)	0.85	0.88
3-VAR	Ara h 2, sIgE, Ara h 3	93.0	0.69	0.96	0.88	12 − 0	0.96 (0.94-0.98)	0.79	1.00	0.95	0.92	4 − 0	0.93 (0.84-0.98)	0.85	1.00
4-VAR	Ara h 2, sIgE, Ara h 3, Ara h 1	94.6	0.60	0.96	0.87	10 − 4	0.96 (0.93-0.98)	0.82	0.98	0.95	0.88	3 − 3	0.90 (0.79-0.96)	0.88	0.91
Models with skin test															
1-VAR	Skin test	79.1	0.66	0.97	0.89	8 − 4	0.96 (0.94-0.98)	0.86	0.98	0.93	0.77	6 − 6	0.80 (0.68-0.89)	0.77	0.82
2-VAR	Skin test, Ara h 2	66.8	0.59	0.98	0.92	8 − 1	0.97 (0.95-0.99)	0.86	1.00	0.96	0.85	3 − 5	0.87 (0.75-0.94)	0.88	0.85
3-VAR	Skin test, Ara h 2, sIgE	65.3	0.88	0.98	0.91	9 − 0	0.97 (0.95-0.99)	0.84	1.00	0.96	0.90	4 − 1	0.92 (0.82-0.97)	0.85	0.97
4-VAR	Skin test, Ara h 2, sIgE, Ara h 3	66.3	0.51	0.98	0.92	7 − 2	0.97 (0.95-0.99)	0.88	0.99	0.97	0.85	3 − 5	0.87 (0.75-0.94)	0.88	0.85
5-VAR	Skin test, Ara h 2, sIgE, Ara h 3, Ara h 1	67.6	0.64	0.98	0.92	8 − 1	0.97 (0.95-0.99)	0.86	1.00	0.98	0.87	3 − 4	0.88 (0.77-0.95)	0.88	0.88

AIC, Akaike information criterion; *AUC*, area under the curve; *CI*, confidence interval; *FN*, false negative; *FP*, false positive.

a previous adverse event during OFC, or restrictions on travel due to the COVID-19 pandemic. In a research setting, assignment of values for outcomes benefits from an *a priori* specification of clear criteria to determine outcomes.

In the LEAP study protocol, the algorithm specified to determine a peanut allergy outcome for those who did not have an outcome determined by OFC was developed based on clinical expertise and the relevant literature at the time. It incorporated dietary history, peanut SPT wheal size, and Pn-sIgE values. In this analysis, we compared participants' allergy status as determined by the OFC at the end of the LEAP study (n = 617) with the status of the same participants as determined by the protocol-specified algorithm and found 73% agreement between the 2 measures with a low misclassification rate (0.6%); however, more than 25% of participants were classified as nonevaluable by the algorithm.

Removing the Pn-sIgE criteria from the protocol-specified algorithm decreased the number of nonevaluable participants but increased the number of misclassified participants. Most of the misclassification occurred with participants with discordant Pn-sIgE and SPT results. It is interesting to note that when cutoffs are applied to Pn-sIgE in a diagnostic algorithm, the diagnostic accuracy is decreased with a large number of indeterminate outcomes, consistent with previous publications on Pn-sIgE and SPT.[16] In contrast, when peanut Pn-sIgE was used as a continuous variable in the multivariate logistic regression prediction model, it enhanced diagnostic performance. The variables found to be predictive of the OFC outcome for peanut in the prediction model were SPT, Pn-sIgE, Ara h 1, Ara h 2, and Ara h 3. The accuracy of the prediction model was high at 0.97, with an overall error rate of 2.8%. The model is thus a useful research tool when OFC or ingestion history is not available. In a clinical setting, OFC remains a necessary practice, as any false-negative result from the prediction model has potentially dangerous consequences.

SPT remains the most influential variable within the model, followed by Ara h 2. Previous studies have shown similar results, with SPT or Ara h 2 as 2 of the strongest predictors of peanut allergy status.[17-20] Here, SPT is the only functional and *in vivo* assessment included in this model, which perhaps explains its predictive value. Differences in age at assessment may contribute to the relative importance of SPT versus Ara h 2 in this and other studies. The addition of Ara h 1 and Ara h 3 modestly improved the model performance. As noted, Pn-sIgG4 was not included in the final model due to the lack of consistent evidence that it is of diagnostic value in peanut allergy as well as uncertainty of its longitudinal trajectory after 72 months of age.[4-8] Further, the addition of Pn-sIgG4 did not significantly improve or contribute to model performance (Figure E2, available in this article's Online Repository at www.jaci-inpractice.org). Other *in vitro* assessments, including the basophil activation test[12] and epitope mapping,[21] have been shown to be superior to SPT in the prediction of peanut allergy status and may one day replace SPT in similar research and clinical models; however, these assays are not yet readily available for widespread use in clinical practice.

The model performed very well in internal validation, using the participants of LEAP-On, as well as in external validation.

It is difficult to predict how immunological characteristics will have changed in the LEAP participants over the 5- to 8-year interval between LEAP and LEAP Trio, although prior work indicates that allergic status is likely to be relatively stable in this age range.[1,2,22] We know, for example, that SPT wheal diameter (greater than 4 mm), which is highly predictive of peanut allergy in LEAP and LEAP-On, is highly predictive of a positive challenge in the Australian HealthNuts study at 12 months of age.[23] For this reason, we expect the model to perform well in the LEAP participants who continue in the LEAP Trio trial, likely similar to the internal validation that used data from participants of LEAP-On.

The siblings of the LEAP participants will also be invited to participate in the LEAP Trio trial. Certain biomarkers such as SPTs have been shown to be highly predictive of peanut allergy in lower-risk general population studies[23] as well as in the LEAP population and, therefore, are likely to perform well in an intermediate-risk population such as the LEAP siblings.

A number of groups have published algorithms to diagnosis food allergies that differ from the approach presented here. DunnGalvin et al[24] demonstrated how SPT, Pn-sIgE, total IgE minus Pn-sIgE, symptoms, sex, and age can be used to predict peanut, milk, or egg allergy across childhood and adolescence. In their algorithm, SPT and history of reaction were the first and second strongest predictors. Kelleher et al[25] developed an algorithm building on a consensus from the Integrated Approaches to Food Allergen and Allergy Management study,[26] which they tested using data from 2 preschool age controlled trials, the Barrier Enhancement for Eczema Prevention (BEEP) trial and the Enquiring About Tolerance (EAT) trial.[27,28] This algorithm was used to guide the deliberations of an expert group who reviewed the available data to define whether participants were or were not food allergic and included the SPT result and more detailed reaction history. This second algorithm had a high sensitivity of 94% (95% CI: 68-100) for BEEP and 90% (72-97) for EAT plus a moderate specificity of 67% (39-87) for BEEP and 67% (39-87) for EAT. The DunnGalvin[24] algorithm had similarly high sensitivity, but its specificity was much higher at approximately 90%. Thus, in these prior algorithms, the addition of total IgE, Pn-sIgE, sex, and age to SPT seems to improve specificity. In this study, we have not included history given that the model was developed using participants who had not consumed peanut in the year before the 60-month OFC in the LEAP trial, but do include peanut diagnostic approaches that have become available in the interim, including Ara h 1, Ara h 2, and Ara h 3. Sex was not found to influence the diagnostic performance of the model (data not shown). In addition, the influence of age on the predictive value of each of the included variables was not possible for the development of this model, as all participants were approximately the same age at the time of assessment.

This prediction model has several limitations. The methodologies used for SPT were standardized across the LEAP studies. Notably, SPT has been shown to have significant variability depending on the extract and device used, as well as the technique of the applicator and duration of time between application and measurement of the test.[29-31] Further, in the LEAP study, the SPT assessments were performed very proximate to the OFC, and we cannot predict how interval time between SPT and OFC may influence the accuracy of the model. Caution should be taken when applying this model to circumstances in which SPT is measured differently as the predictive importance of SPT may change when an alternate methodology is used.

Another limitation of the model is that it was developed using a predominantly Caucasian population (75%), which may limit generalizability. Subanalysis demonstrated that the model performed well within racial and ethnic minorities included in the

analysis; however, the number of participants in this subanalysis was small (data not shown).

A third limitation is that LEAP participants were all enrolled as infants and had severe atopic dermatitis and/or egg allergy. Although the model performed reasonably well on external validation in an older pediatric population, the model has not yet been validated in other populations with varying degrees of atopy.

The prediction model performs extremely well in LEAP participants and was designed specifically for use in the LEAP Trio study. The model performs well in external validation as well, suggesting that similar models may be used in other trials. Given the multitude of barriers to performing an OFC both in clinical research and patient care, further tools such as this model are needed to predict allergy status. Although the application of this model to clinical practice is not currently advised because of the risk of false-negative predictions, statistical models could be used in the future as an adjunct to help support clinical decision-making, prioritization, and risk assessment before the OFC. We would be happy to share this model with other centers and readers wishing to validate this model in their clinics.

Acknowledgments

We thank the many nurses, dietitians, doctors, and members of the administrative staff at St. Thomas' Hospital Children's Allergy Service for clinical and logistic assistance for the LEAP, LEAP-On, and LEAP Trio studies, as well as all the children and their families who took part in these studies. We also thank Kaitie Lawson, MS, for her contributions during the early discussions of this manuscript and her work conducting the evaluation of the protocol-specified algorithm.

REFERENCES

1. Du Toit G, Roberts G, Sayre PH, Bahnson HT, Radulovic S, Santos AF, et al. Randomized trial of peanut consumption in infants at risk for peanut allergy. N Engl J Med 2015;372:803-13.
2. Du Toit G, Sayre PH, Roberts G, Sever ML, Lawson K, Bahnson HT, et al. Effect of avoidance on peanut allergy after early peanut consumption. N Engl J Med 2016;374:1435-43.
3. Lieberman JA, Sicherer SH. Diagnosis of food allergy: epicutaneous skin tests, in vitro tests, and oral food challenge. Curr Allergy Asthma Rep 2011;11:58-64.
4. Hong X, Caruso D, Kumar R, Liu R, Liu X, Wang G, et al. IgE, but not IgG4, antibodies to Ara h 2 distinguish peanut allergy from asymptomatic peanut sensitization. Allergy 2012;67:1538-46.
5. Lopes de Oliveira LC, Aderhold M, Brill M, Schulz G, Rolinck-Werninghaus C, Clare Mills EN, et al. The value of specific IgE to peanut and its component Ara h 2 in the diagnosis of peanut allergy. J Allergy Clin Immunol Pract 2013;1:394-8.
6. Suarez-Farinas M, Suprun M, Bahnson HT, Raghunathan R, Getts R, duToit G, et al. Evolution of epitope-specific IgE and IgG4 antibodies in children enrolled in the LEAP trial. J Allergy Clin Immunol 2021;148:835-42.
7. Datema MR, Eller E, Zwinderman AH, Poulsen LK, Versteeg SA, van Ree R, et al. Ratios of specific IgG4 over IgE antibodies do not improve prediction of peanut allergy nor of its severity compared to specific IgE alone. Clin Exp Allergy 2019;49:216-26.
8. Glaumann S, Nilsson C, Asarnoj A, Moverare R, Johansson SG, Borres MP, et al. IgG4 antibodies and peanut challenge outcome in children IgE-sensitized to peanut. Pediatr Allergy Immunol 2015;26:386-9.
9. van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. J Stat Softw 2011;45:67.
10. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. R version 4.2.0. 2022. Accessed April 22, 2022. https://www.R-project.org/
11. h2o: R interface for the 'H2O' scalable machine learning platform. R package version 3.30.0.1. 2020. Accessed April 22, 2022. https://CRAN.R-project.org/package=h2o
12. Santos AF, Douiri A, Becares N, Wu SY, Stephens A, Radulovic S, et al. Basophil activation test discriminates between allergy and tolerance in peanut-sensitized children. J Allergy Clin Immunol 2014;134:645-52.
13. May CD. Objective clinical and laboratory studies of immediate hypersensitivity reactions to foods in asthmatic children. J Allergy Clin Immunol 1976;58:500-15.
14. Bernstein M, Day JH, Welsh A. Double-blind food challenge in the diagnosis of food sensitivity in the adult. J Allergy Clin Immunol 1982;70:205-10.
15. Bird JA, Leonard S, Groetch M, Assa'ad A, Cianferoni A, Clark A, et al. Conducting an oral food challenge: an update to the 2009 adverse reactions to foods committee work group report. J Allergy Clin Immunol Pract 2020;8:75-90.e17.
16. Greenhawt M, Shaker M, Wang J, Oppenheimer JJ, Sicherer S, Keet C, et al. Peanut allergy diagnosis: a 2020 practice parameter update, systematic review, and GRADE analysis. J Allergy Clin Immunol 2020;146:1302-34.
17. Klemans RJ, Otte D, Knol M, Knol EF, Meijer Y, Gmelig-Meyling FH, et al. The diagnostic value of specific IgE to Ara h 2 to predict peanut allergy in children is comparable to a validated and updated diagnostic prediction model. J Allergy Clin Immunol 2013;131:157-63.
18. Klemans RJ, van Os-Medendorp H, Blankestijn M, Bruijnzeel-Koomen CA, Knol EF, Knulst AC. Diagnostic accuracy of specific IgE to components in diagnosing peanut allergy: a systematic review. Clin Exp Allergy 2015;45:720-30.
19. Keet C, Plesa M, Szelag D, Shreffler W, Wood R, Dunlop J, et al. Ara h 2-specific IgE is superior to whole peanut extract-based serology or skin prick test for diagnosis of peanut allergy in infancy. J Allergy Clin Immunol 2021;147:977-983.e2.
20. Santos AF, Du Toit G, O'Rourke C, Becares N, Couto-Francisco N, Radulovic S, et al. Biomarkers of severity and threshold of allergic reactions during oral peanut challenges. J Allergy Clin Immunol 2020;146:344-55.
21. Suprun M, Sicherer SH, Wood RA, Jones SM, Leung DYM, Henning AK, et al. Early epitope-specific IgE antibodies are predictive of childhood peanut allergy. J Allergy Clin Immunol 2020;146:1080-8.
22. Peters RL, Allen KJ, Dharmage SC, Koplin JJ, Dang T, Tilbrook KP, et al. Natural history of peanut allergy and predictors of resolution in the first 4 years of life: a population-based assessment. J Allergy Clin Immunol 2015;135:1257-1266.e1-2.
23. Koplin JJ, Peters RL, Dharmage SC, Gurrin L, Tang MLK, Ponsonby AL, et al. Understanding the feasibility and implications of implementing early peanut introduction for prevention of peanut allergy. J Allergy Clin Immunol 2016;138:1131-1141.e2.
24. DunnGalvin A, Daly D, Cullinane C, Stenke E, Keeton D, Erlewyn-Lajeunesse M, et al. Highly accurate prediction of food challenge outcome using routinely available clinical data. J Allergy Clin Immunol 2011;127:633-639.e1-3.
25. Kelleher MM, Jay N, Perkin MR, Haines RH, Batt R, Bradshaw LE, et al. An algorithm for diagnosing IgE-mediated food allergy in study participants who do not undergo food challenge. Clin Exp Allergy 2020;50:334-42.
26. Grabenhenrich LB, Reich A, Bellach J, Trendelenburg V, Sprikkelman AB, Roberts G, et al. A new framework for the documentation and interpretation of oral food challenges in population-based and clinical research. Allergy 2017;72:453-61.
27. Perkin MR, Logan K, Tseng A, Raji B, Ayis S, Peacock J, et al. Randomized trial of introduction of allergenic foods in breast-fed infants. N Engl J Med 2016;374:1733-43.
28. Chalmers JR, Haines RH, Bradshaw LE, Montgomery AA, Thomas KS, Brown SJ, et al. Daily emollient during infancy for prevention of eczema: the BEEP randomised controlled trial. Lancet 2020;395:962-72.
29. Maleki SJ, Casillas AM, Kaza U, Wilson BA, Nesbit JB, Reimoneqnue C, et al. Differences among heat-treated, raw, and commercial peanut extracts by skin testing and immunoblotting. Ann Allergy Asthma Immunol 2010;105:451-7.
30. Bernstein IL, Li JT, Bernstein DI, Hamilton R, Spector SL, Tan R, et al. Allergy diagnostic testing: an updated practice parameter. Ann Allergy Asthma Immunol 2008;100(Suppl 3):S1-148.
31. Sicherer SH, Wood RA. Advances in diagnosing peanut allergy. J Allergy Clin Immunol Pract 2013;1:1-13. quiz: 4.

ONLINE REPOSITORY

METHODS

Five-fold cross-validation is a widely used technique for evaluating the performance of machine learning models, including logistic regression models (Figure E1). The idea behind 5-fold cross-validation is to split the entire data set into 5 equally sized folds or subsets.

The process works as follows:

(1) Split the data set into 5 folds: the original data set is divided into 5 folds, each with roughly equal number of samples.
(2) Train the model on 4 folds: the logistic regression model is trained on 4 of the 5 folds (or 80% of the data) while holding out 1 fold (or 20% of the data).
(3) Evaluate the model on the held-out fold: the trained model is then used to make predictions on the held-out fold (the one not used for training). The F1 score is calculated for the predictions made on this fold.
(4) Repeat steps 2-3 for all 5 folds: the process of training the model on 4 folds and evaluating it on 1 fold is repeated 5 times, each time with a different fold held out for evaluation. The F1 score is calculated for each of the 5 evaluations.
(5) Average the 5 F1 scores: finally, the 5 F1 scores are averaged to obtain a single overall performance metric for the logistic regression model.

By using 5-fold cross-validation, the model is trained and evaluated on multiple different subsets of the data, which provides a better estimate of the model's generalization performance compared with using a single training and testing split. The F1 score is a good performance metric to use in this context because it balances precision and recall, making it suitable for imbalanced data sets where some classes have many more examples than others.

The F1 score provides a measure of how well the logistic regression model can classify positive oral food challenge (OFC) (given a threshold value). The F1 score is calculated from the harmonic mean of the precision and recall. An F1 score of 1 means both precision and recall are perfect, and the model correctly identified all the positive OFC and did not make an error predicting a negative OFC as positive:

$$\mathrm{F1} = 2 \times \frac{\mathrm{precision} \times \mathrm{recall}}{\mathrm{precision} + \mathrm{recall}}.$$

Precision (positive predicted value) is the positive OFC (true positives) the model correctly identified from all the observations it labeled as positive (the true positives + the false positives).

Recall (sensitivity) is the positive observations (true positives) the model correctly identified from all the actual positive cases (the true positives + the false negatives).

The F1 score intends to strike a balance between Precision and Recall. Because Precision and Recall each have strengths and weaknesses, the F1 score is often a good metric to optimize. F1 scores range from 0 to 1. Values closer to 1 are preferred.

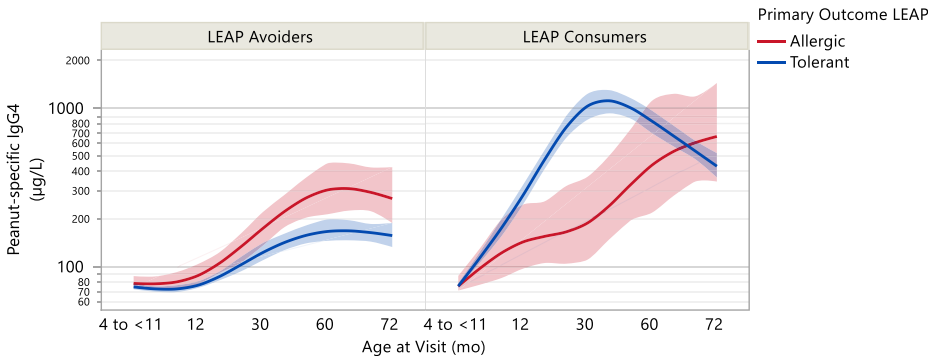

FIGURE E1. Peanut-specific IgG4 in Learning Early About Peanut Allergy (LEAP).

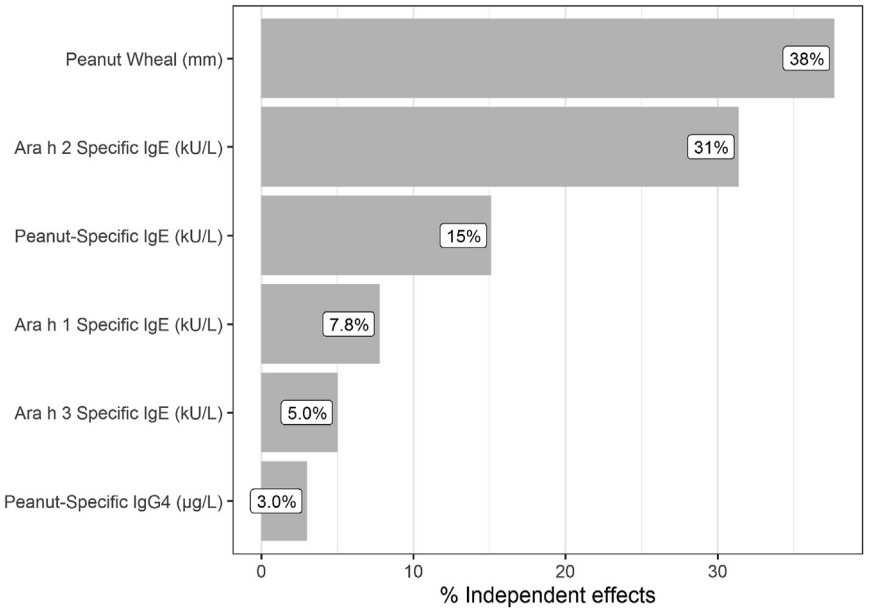

FIGURE E2. Relative importance of variables in the model, including peanut-specific IgG4, for prediction of the oral food challenge outcome.

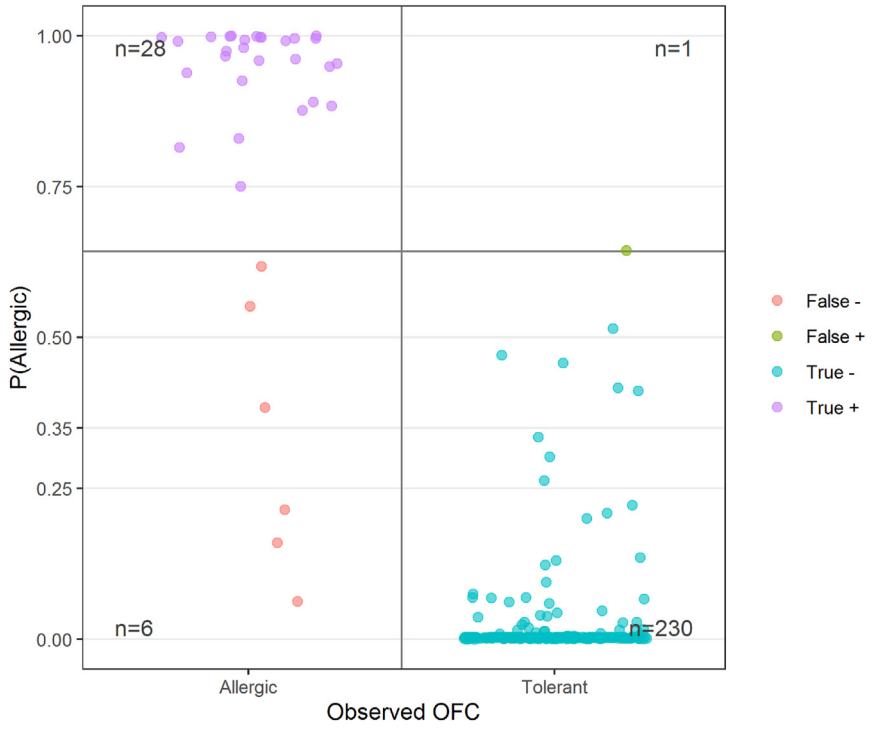

FIGURE E3. Internal validation (LEAP-On n = 265) predicted probability of positive oral food challenge [P(Allergic)] versus observed allergic oral food challenge (OFC) results. *LEAP*, Learning Early About Peanut Allergy.

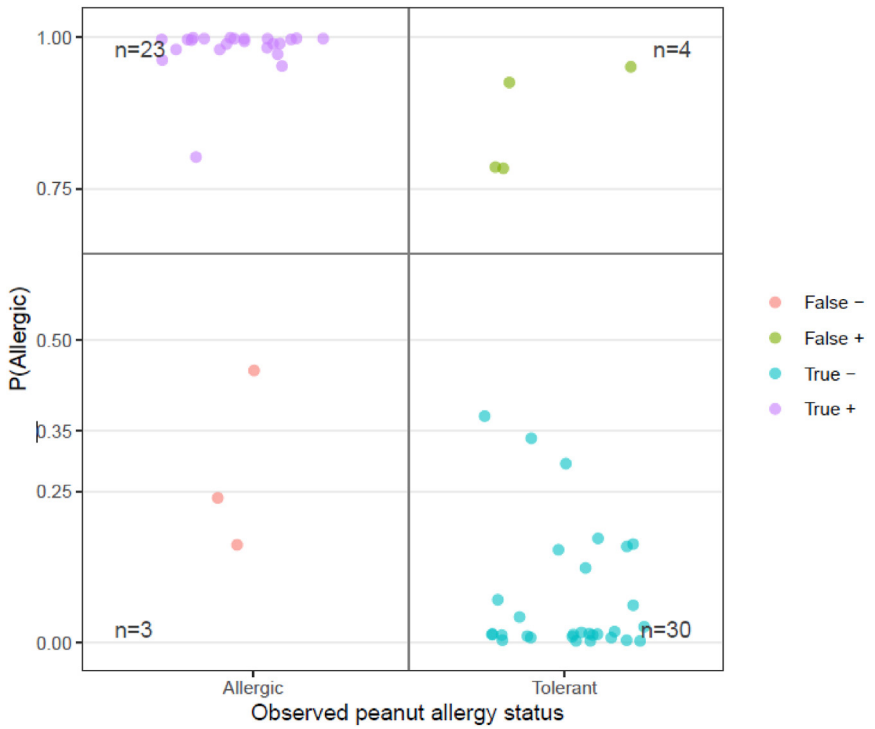

FIGURE E4. External validation (n = 60) predicted probability of positive oral food challenge [P(Allergic)] versus observed peanut allergy status results.

TABLE E1. Mismatches between observed and predicted oral food challenge (OFC) results

	Observed								Predicted		
Sex	Wheal size (mm)	sIgE (kU/L)	IgG4 (μg/L)	IgG4/IgE	Ara h1 (kU/L)	Ara h2 (kU/L)	Ara h3 (kU/L)	OFC	OFC	P(Nonallergic) probability	P(Allergic) probability
M	5.0	0.81	900	463.0	0.01	0.09	0.02	Not allergic	Allergic	0.28	0.72
M	0.0	0.29	750	1077.6	0.02	0.03	0.01	Allergic	Not allergic	1.00	0.00
M	4.0	0.44	70	66.3	0.08	0.01	0.34	Allergic	Not allergic	0.96	0.04
F	3.0	0.32	80	104.2	0.01	0.01	0.01	Allergic	Not allergic	0.90	0.10
M	4.0	2.17	190	36.5	0.01	0.03	0.11	Allergic	Not allergic	0.81	0.19
M	4.0	0.20	70	145.8	0.01	0.01	0.01	Allergic	Not allergic	0.81	0.19
M	2.0	0.43	170	164.7	0.05	0.26	0.10	Allergic	Not allergic	0.64	0.36
M	5.0	1.26	740	244.7	0.19	0.07	0.08	Allergic	Not allergic	0.63	0.37
M	3.0	0.66	70	44.2	0.01	0.08	0.04	Allergic	Not allergic	0.63	0.37

TABLE E2. Internal validation results of sequential logistic regression models (with and without the skin test) for the prediction of the oral food challenge

			Cross-validation (n = 323)							LEAP-On (n = 265)					
No. of variables	Variables	AIC	Cutoff	AUC	F1	FN − FP	Accuracy 95% CI	Sensitivity	Specificity	AUC	F1	FN − FP	Accuracy 95% CI	Sensitivity	Specificity
Models without skin test															
1-VAR	Ara h2	105.1	0.50	0.94	0.86	10 − 5	0.95 (0.92-0.97)	0.82	0.98	0.93	0.70	12 − 7	0.93 (0.89-0.96)	0.65	0.97
2-VAR	Ara h2, sIgE	106.4	0.49	0.95	0.84	11 − 6	0.95 (0.92-0.97)	0.81	0.98	0.95	0.70	12 − 7	0.93 (0.89-0.96)	0.65	0.97
3-VAR	Ara h2, sIgE, Ara h3	93.0	0.69	0.96	0.88	12 − 0	0.96 (0.94-0.98)	0.79	1.00	0.94	0.72	13 − 3	0.94 (0.90-0.97)	0.62	0.99
4-VAR	Ara h2, sIgE, Ara h3, Ara h1	94.6	0.60	0.96	0.87	10 − 4	0.96 (0.93-0.98)	0.82	0.98	0.94	0.75	11 − 4	0.94 (0.91-0.97)	0.68	0.98
Models with skin test															
1-VAR	Skin test	79.1	0.66	0.97	0.89	8 − 4	0.96 (0.94-0.98)	0.86	0.98	0.99	0.87	6 − 2	0.97 (0.94-0.99)	0.82	0.99
2-VAR	Skin test, Ara h2	66.8	0.59	0.98	0.92	8 − 1	0.97 (0.95-0.99)	0.86	1.00	0.99	0.87	6 − 2	0.97 (0.94-0.99)	0.82	0.99
3-VAR	Skin test, Ara h2, sIgE	65.3	0.88	0.98	0.91	9 − 0	0.97 (0.95-0.99)	0.84	1.00	0.99	0.83	10 − 0	0.96 (0.93-0.98)	0.71	1.00
4-VAR	Skin test, Ara h2, sIgE, Ara h3	66.3	0.51	0.98	0.92	7 − 2	0.97 (0.95-0.99)	0.88	0.99	0.99	0.92	4 − 1	0.98 (0.96-0.99)	0.88	1.00
5-VAR	Skin test, Ara h2, sIgE, Ara h3, Ara h1	67.6	0.64	0.98	0.92	8 − 1	0.97 (0.95-0.99)	0.86	1.00	0.99	0.89	6 − 1	0.97 (0.95-0.99)	0.82	1.00

AIC, Akaike information criterion; *AUC*, area under the curve; *CI*, confidence interval; *FN*, false negative; *FP*, false positive; *LEAP*, Learning Early About Peanut Allergy.

TABLE E3. Characteristics of the internal validation cohort

	All			Allergic		
Characteristics	**Training (N = 323)**	**LEAP-On (N = 265)**	***P* value***	**Training (N = 57)**	**LEAP-On (N = 34)**	***P* value***
Oral food challenge outcome, n (%)			.11			
Not allergic	266 (82)	231 (87)				
Allergic	57 (18)	34 (13)				
Age at OFC (y)			<.001			<.001
Median (IQR)	5.0 (4.8, 5.2)	6.2 (6.0, 6.6)		5.0 (4.8, 5.2)	6.3 (6.0, 6.6)	
Range	4.3-6.5	5.5-8.8		4.5-6.0	5.6-7.6	
Peanut wheal size (mm)			.33			.58
Median (IQR)	0.0 (0.0, 2.0)	0.0 (0.0, 1.0)		9.0 (6.0, 12.0)	9.0 (6.0, 10.8)	
Range	0.0-18.0	0.0-17.0		0.0-18.0	4.0-17.0	
Peanut-specific IgE (kU/L)			.77			.020
Median (IQR)	0.07 (0.01, 1.00)	0.08 (0.02, 0.63)		3.75 (1.06, 36.30)	1.35 (0.46, 3.85)	
Range	0.01-363	0.01-210		0.14-363	0.18-210	
Missing	8	9		0	1	
Peanut-specific IgG4 (μg/L)			.21			.23
Median (IQR)	105 (70, 460)	80 (70, 322)		210 (80, 685)	130 (70, 430)	
Range	70-22,700	70-15,900		70-19,500	70-13,000	
Missing	9	9		1	1	
IgG4 (μg/L)/IgE (ng/mL) log			.96			.025
Median (IQR)	2.99 (2.13, 3.46)	2.99 (2.08, 3.46)		1.46 (0.81, 2.02)	1.81 (1.45, 2.11)	
Range	−0.38 to 4.89	−0.75 to 5.02		−0.38 to 3.03	−0.75 to 2.69	
Missing	9	9		1	1	
Ara h1 (kU/L)			.13			.029
Median (IQR)	0.01 (0.01, 0.04)	0.01 (0.01, 0.03)		0.16 (0.01, 4.26)	0.02 (0.01, 0.15)	
Range	0.01-279.0	0.01-97.3		0.01-279.0	0.01-97.3	
Missing	64	62		6	2	
Ara h2 (kU/L)			.089			.027
Median (IQR)	0.01 (0.01, 0.06)	0.01 (0.01, 0.04)		1.75 (0.28, 18.30)	0.54 (0.12, 1.49)	
Range	0.01-456.0	0.01-91.9		0.01-456.0	0.01-91.9	
Missing	64	63		6	2	
Ara h3 (kU/L)			.24			.006
Median (IQR)	0.01 (0.01, 0.04)	0.01 (0.01, 0.04)		0.08 (0.01, 0.45)	0.54 (0.12, 1.49)	
Range	0.01-92.1	0.01-91.9		0.01-92.1	0.01-91.9	
Missing	64	63		6	2	
Ara h1 (kU/L) attributed			.008			.017
Median (IQR)	0.01 (0.01, 0.03)	0.01 (0.01, 0.02)		0.19 (0.01, 4.05)	0.02 (0.01, 0.14)	
Range	0.01-279.0	0.01-97.3		0.01-279.0	0.01-97.3	
Ara h2 (kU/L) attributed			.009			.011
Median (IQR)	0.01 (0.01, 0.05)	0.01 (0.01, 0.02)		1.75 (0.30, 38.50)	0.50 (0.13, 1.47)	
Range	0.01-456.0	0.01-91.9		0.01-456.0	0.01-91.9	
Ara h3 (kU/L) attributed			<.001			.19
Median (IQR)	0.01 (0.01, 0.04)	0.02 (0.01, 0.06)		0.08 (0.01, 0.46)	0.04 (0.01, 0.13)	
Range	0.01-92.1	0.01-7.4		0.01-92.1	0.01-4.7	

IQR, Interquartile range; LEAP, Learning Early About Peanut Allergy; *OFC*, oral food challenge.
*Pearson χ^2 test, Fisher exact test, and Wilcoxon rank-sum test.

TABLE E4. Internal (LEAP-On) validation results of the 5-variable logistic regression model for the prediction of the oral food challenge

Statistic	Result
Prediction results	
True positive	230
False positive	1
False negative	6
True negative	28
Overall	
Accuracy	0.97 (0.95, 0.99)
Balanced accuracy	0.91
Kappa	0.87
Classes	
Sensitivity (recall/true positive rate)	0.82 (0.65, 0.93)
Specificity (true negative rate)	1.00 (0.98, 1.00)
Positive predictive value (precision)	0.97 (0.82, 1.00)
Negative predictive value	0.97 (0.95, 0.99)
F1 score*	0.89 (0.78, 0.96)
Prevalence	0.13 (0.09, 0.17)
Detection prevalence	0.11 (0.07, 0.15)

The highest possible value of the F1 score is 1.0, indicating perfect precision and recall, and the lowest possible value is 0 if either the precision or the recall is zero.

*F1 score is the harmonic mean of the precision and recall $= 2 \times \frac{\text{precison} \times \text{recall}}{\text{precision} + \text{recall}}$.

TABLE E5. Characteristics of external validation cohort

	All			Allergic		
Characteristics	**Training (N = 323)**	**External (N = 60)**	***P* value***	**Training (N = 57)**	**External (N = 26)**	***P* value***
Outcome, n (%)			<.001			
Not allergic	266 (82)	34 (57)				
Allergic	57 (18)	26 (43)				
Sex, n (%)			.34			.39
Male	211 (65)	43 (72)		41 (72)	21 (81)	
Female	112 (35)	17 (28)		16 (28)	5 (19)	
Age (y)			<.001			<.001
Median (IQR)	5.0 (4.8, 5.2)	7.7 (4.0, 10.6)		5.0 (4.8, 5.2)	8.0 (5.9, 11.4)	
Range	4.3-6.5	0.5-16.9		4.5-6.0	1.7-16.9	
Peanut wheal size (mm)			<.001			.82
Median (IQR)	0.0 (0.0, 2.0)	4.0 (0.0, 8.0)		9.0 (6.0, 12.0)	8.0 (7.0, 12.8)	
Range	0.0-18.0	0.0-19.0		0.0-18.0	1.0-19.0	
Peanut-specific IgE (kU/L)			<.001			.006
Median (IQR)	0.07 (0.02, 1.00)	2.17 (0.06, 17.4)		3.75 (1.06, 36.30)	72.10 (2.84, 191.5)	
Range	0.01-363	0.01-568		0.14-363	0.15-568	
Missing	8	0				
Peanut specific IgG4 (μg/L)			.14			.66
Median (IQR)	105 (70, 460)	160 (70, 515)		210 (80, 685)	290 (130, 770)	
Range	70-22,700	70-4300		70-19,500	70-2700	
Missing	9	1		1	1	
IgG4 (μg/L)/IgE (ng/mL) ratio log			<.001			.003
Median (IQR)	2.99 (2.13, 3.46)	1.63 (0.59, 3.16)		1.46 (0.81, 2.02)	0.51 (0.09, 1.25)	
Range	−0.38 to 4.89	−0.79 to 4.52		−0.38 to 3.03	−0.79 to 3.01	
Missing	9	1		1	1	
Ara h1 (kU/L)			<.001			.032
Median (IQR)	0.01 (0.01, 0.04)	0.03 (0.01, 11.6)		0.16 (0.01, 4.26)	15.40 (0.04, 68.7)	
Range	0.01-279.0	0.01-199.0		0.01-279.0	0.01-199.0	
Missing	64	1		6	0	
Ara h2 (kU/L)			<.001			.035
Median (IQR)	0.01 (0.01, 0.06)	0.12 (0.04, 4.97)		1.75 (0.28, 18.30)	45.45 (1.38, 87.6)	
Range	0.01-456.0	0.01-278.0		0.01-456.0	0.01-278.0	
Missing	64	1		6	0	
Ara h3 (kU/L)			<.001			.045
Median (IQR)	0.01 (0.01, 0.04)	0.04 (0.02, 0.71)		0.08 (0.01, 0.45)	0.71 (0.02, 26.5)	
Range	0.01-92.1	0.01-89.6		0.01-92.1	0.01-89.6	
Missing	64	1		6	0	
Ara h1 (kU/L) attributed			<.001			.027
Median (IQR)	0.01 (0.01, 0.03)	0.03 (0.01, 11.5)		0.19 (0.01, 4.05)	15.40 (0.04, 68.7)	
Range	0.01-279.0	0.01-199.0		0.01-279.0	0.01-199.0	
Ara h2 (kU/L) attributed			<.001			.047
Median (IQR)	0.01 (0.01, 0.05)	0.12 (0.04, 3.80)		1.75 (0.30, 38.50)	45.45 (1.38, 87.6)	
Range	0.01-456.0	0.01-278.0		0.01-456.0	0.01-278.0	
Ara h3 (kU/L) attributed			<.001			.047
Median (IQR)	0.01 (0.01, 0.04)	0.04 (0.02, 0.62)		0.08 (0.01, 0.46)	0.71 (0.02, 26.5)	
Range	0.01-92.1	0.01-89.6		0.01-92.1	0.01-89.6	

IQR, Interquartile range.
*Pearson χ^2 test, Fisher exact test, and Wilcoxon rank-sum test.

TABLE E6. External validation results of the 5-variable logistic regression model for the prediction of peanut allergy status

Statistic	Result
Prediction results	
True positive	30
False positive	4
False negative	3
True negative	23
Overall	
Accuracy	0.88 (0.77, 0.95)
Balanced accuracy	0.88
Kappa	0.76
Classes	
Sensitivity (recall/true positive rate)	0.88 (0.70, 0.98)
Specificity (true negative rate)	0.88 (0.73, 0.97)
Positive predictive value (precision)	0.85 (0.66, 0.96)
Negative predictive value	0.91 (0.76, 0.98)
F1 score*	0.87 (0.68, 0.96)
Prevalence	0.43 (0.31, 0.57)
Detection prevalence	0.45 (0.32, 0.58)

The highest possible value of the F1 score is 1.0, indicating perfect precision and recall, and the lowest possible value is 0 if either the precision or the recall is zero.

*F1 score is the harmonic mean of the precision and recall $= 2 \times \frac{\text{precison} \times \text{recall}}{\text{precision} + \text{recall}}$.