These Labels are Nuts: Challenges to Safe Product Identification for Nut-Allergic Consumers

Lauren Parikhal,[1] Hillary Abraham,[1] Alea Mehler, [1] Thomas McWilliams,[1] Jonathan Dobres,[1] Nadine Chahine, & Bryan Reimer[1]
1. Massachusetts Institute of Technology AgeLab

Allergen information on food labels is not standardized, making allergen avoidance difficult for consumers. This study investigated the speed and accuracy of allergen identification on commercial packaging across different types of warning labels. The results identified packaging label characteristics significantly correlated with faster and more accurate identification of allergens. Standardizing warning and safe-to-consume labels may reduce risk of accidental allergen exposure for consumers managing food allergies.

INTRODUCTION

Managing food allergies requires consumers to make various dietary and lifestyle changes. The most prevalent advice for preventing serious allergic reactions is to read "every label, every time" (Food Allergy Research & Education [FARE], 2015, p. 6). While thoroughly reading food labels can be effective, even educated and cautious consumers make mistakes (Joshi, Mofidi, & Sicherer, 2002).

The Food Allergen Labeling and Consumer Protection Act (FALCPA) attempted to ameliorate the process of repeatedly examining food labels. It defines the eight major allergens accounting for 90% of allergic reactions in the US and requires manufacturers to label products containing any of these allergens with their common name (e.g., "milk" instead of "casein"). However, the allergens can be listed either in the full ingredients list or in a separate "Contains" statement (Food & Drug Administration [FDA], 2004). This lack of standardization confuses consumers, who frequently interpret a lack of a "Contains" statement as an indication that the product is safe to consume (Barnett et al., 2011).

Allergen advisory statements are also unstandardized (Pieretti, Chung, Pacenza, Slotkin, & Sicherer, 2009). These statement variations do not correlate with risk of exposure to an allergen; for example, "may contain" statements do not consistently indicate a higher, lower, or equal level of risk as "shared equipment" statements. Yet consumers interpret these statements as reliable gauges of risk and avoid products with these statements at different rates (Hefle et al., 2007).

Little research has explored whether FARE's advice of reading "every label, every time" is easier, faster, or more effective with certain kinds of labels. A study was designed to explore how different labels correlated with accurate and rapid identification of at-risk products when labels were examined.

METHODS

Participants

Forty-four individuals between the ages of 18-24 and 55-69 participated. None of the participants reported having a nut allergy. All participants were in self-reported reasonably good health for their age, fully comfortable handling products that contained nuts, and had normal or corrected to normal vision. The final dataset of 32 participants was equally balanced by age group and gender, such that 8 participants fell into each demographic quadrant. Of the 12 excluded participants, 4 were due to hardware difficulties, 4 for misinterpreting instructions, and 4 for procedural errors.

Products

Fifty-two products were originally chosen as stimuli. Products were selected based on a combination of ingredients, presence or absence of a warning label, visual style of warning (i.e. ingredients and/or allergen advisory statements), type of package, and without regard for a particular brand or food type. The products were equally divided into 6 categories by the amount and type of nut-related information on each product's packaging. Two products were removed after researchers determined that they did not adequately fit the category criteria, and 1 was removed due to severe package deterioration over the course of the study caused by frequent handling by participants, leaving 49 products for analysis.

Apparatus

Video was collected using a GoPro Hero 3. The GoPro was positioned in front of the participant to capture the participant's face, hand movements, verbal responses, and the products themselves during inspection. The products were placed at random within a large bin on its side with the opening facing the participant. A multi-panel curtain was draped over the opening to block sight of the products while allowing participants to reach in and retrieve a product. An empty bin was provided in which participants placed products after they had interacted with it.

Procedure

Upon arrival, participants were provided an outline of the study, read and signed an informed consent form approved by the institutional review board, completed a pre-experimental questionnaire, and had their near vision assessed using the Federal Aviation Administration's Form 8500-1 for near acuity. Participants were then asked to imagine they were shopping for foods that would be safe for someone with a severe broad-spectrum nut allergy (i.e. allergic to both tree nuts and peanuts) to consume. In line with guidance commonly provided by physicians, participants were instructed to avoid all products containing a wide range of nuts, specifically “peanuts, almonds, walnuts, pecans, macadamia nuts, tree nuts, etc.,” and to avoid products “that (have) the possibility of cross contamination with nuts.”

Participants were instructed to reach in and remove one product at random from the covered bin. This method of item selection ensured that product order was randomized for each participant. After removing an item, participants determined if the product was safe or unsafe for purchase given the scenario presented earlier. They were allowed to handle and turn over the product freely to examine all sides of the packaging. Once they had determined if the product was safe or unsafe, they gave a verbal rating of “yes” or “no”, indicating a safe or unsafe product, respectively. After making a determination, participants placed the product in the spare bin and repeated the task for the remaining products. After all products were inspected, participants completed a post-experimental questionnaire.

ANALYSIS

Video and audio of each experimental session were used to determine the duration a participant spent examining a product, the number of times each product was turned while examined, and the accuracy of each response. This dataset was double coded by two trained researchers and mediated by a third, if discrepancies in coding were found. Abbreviated coding guidelines defining the beginning and end of a task period, scoring, and turns are outlined in (Table 1).

Duration	Scoring	Turning
Start time: the product was roughly halfway out of the box **End time:** the subject gave a final verbal response of yes or no	**1:** Final verbal response was correct with no errors. **0.5:** Subject initially made an error, but corrected themselves **0:** Subject’s final answer was incorrect	**Turn:** rotation of the product in order to look at a different section **Not a turn:** slight rotations in order to read the same section of the package **Not a Turn:** re-orientation immediately after taking it from the bin or placing it in the disposal bin as or after they give a verbal response.

Table 1: Abbreviated coding guidelines

General Characterization

To establish an initial characterization of the data, dependent measures were averaged across products per participant, which were then used to compute group statistics. Across all products and participants, an average of 20.2 seconds was spent examining each product (SE 1.49), with an average of 2.15 turns (SE 0.10). Overall decision accuracy was 90.37% (SE 1.17%). Duration and number of turns were found to be significantly correlated ($R = 0.73$, $t(30) = 5.92$, $p < 0.001$), while duration and accuracy were not ($R = -0.06$, $t(30) = -0.33$, $p = 0.745$). Given the strong correlation between search duration and number of turns, subsequent analyses do not examine number of turns separately.

Examining differences between demographic groups, tests show that search duration was significantly affected by age group (Older $M = 25.6$s, Younger $M = 14.7$s; $F(1, 28) = 27.9$, $p < 0.001$) and by gender (Female $M = 17.7$s, Male $M = 22.6$s; $F(1, 28) = 5.25$, $p = 0.030$). These factors did not interact significantly ($F(1, 28) = 0.33$, $p = 0.571$). Decision accuracy was not affected by age ($F(1,28) = 0.04$, $p = 0.842$) or gender ($F(1, 28) = 0.763$, $p = 0.390$) as main effects, but age and gender did interact significantly ($F(1, 28) = 6.26$, $p = 0.018$). Examining differences between genders within each age group, decision accuracy was not significantly affected by gender among younger participants ($t(10) = 1.06$, $p = 0.312$). However, gender did significantly affect decision accuracy among older participants ($t(12) = 2.62$, $p = 0.022$), with older women having higher accuracy than older men (94.4% and 87.0%, respectively).

Owing to the large number of within-subject categories (49 products) relative to the sample size (32 participants), omnibus testing of differences between products would not be statistically sound nor particularly informative. Instead, dependent measures can be averaged per item, and the resulting correlations between dependent measures tested. Since several products were associated with overall nut identification accuracies of 100%, the relationship between search duration and decision accuracy is nonlinear, and Spearman’s rho is used to calculate correlations between the two. Results show that search duration is negatively correlated with decision accuracy ($rho = -0.83$, $p < 0.001$), suggesting that longer search times are associated with lower accuracy. The strong correlation between these measures observed among per-item aggregates, combined with the lack of significant correlation among per-participant averages (above), suggests that the relationship between search time and decision accuracy has less to do with participant idiosyncrasies and more to do with the products themselves.

Presence of Warnings

Several broad categories representing the presence or lack of warnings on products were identified during the design phase of this study. A number of products were selected for each category to obtain a degree of balance across trials (Table 2). The study included 29 products that advertised at least some risk of nut allergen content. Eight products contained an explicit warning (“Only Warning” category), 7 products included “nut” in the product name and an explicit warning (“Name & Warning” category), 7 products included a warning indicating the potential for cross-contamination (“Allergen Advisory” category), and 7 products included no explicit warning about nut content (“No Warning” category). The study also included 20 products that contained no nut ingredients. Four of these products included “nut” in the product name despite being nut-free (“Negative” category). The remaining 16 are part of the “No Trace” category.

Figure 1 shows a violin plot for distributions of search durations averaged per participant and product category. Violins represent smoothed histograms of the resulting distributions, and the horizontal lines represent their median values. Repeated-measures ANOVA for the “safe” products

Product Group	Safety Status	Nuts in name	Nuts in ingredients	Warning Present	# of Products
No Trace	Safe	No	No	No	16
Negative	Safe	Yes	No	No	4
Name & Warning	Unsafe	Yes	Yes	Yes	7
Only Warning	Unsafe	No	Yes	Yes	8
No Warning	Unsafe	No	Yes	No	7
Allergen Advisory	Unsafe	No	No	Yes	7

Table 2: Product categories and characteristics

shows that there was no significant difference in search times for “No Trace” products and “Negative” products ($F(1, 31) = 0.67$, $p = 0.418$). Among “unsafe” products, there were significant differences between the product categories ($F(3, 93) = 71.91$, $p < 0.001$). Post-hoc testing indicates that all of the unsafe product categories had search times significantly different from each other ($p < 0.003$, paired t-tests). Generally speaking, safe categories were searched for longer than unsafe categories.

Figure 2 shows distributions of decision accuracies averaged per participant and product category. Patterns of decision accuracy are, generally speaking, the inverse of the search time patterns. A Friedman test between the safe products shows no significant difference between the “Negative” and “No Trace” categories ($\chi^2(1) = 1.00$, $p = 0.317$). Among the unsafe products, decision accuracy differed significantly by product category ($\chi^2(3) = 41.25$, $p < 0.001$). Posthoc testing shows that the “Name & Warning” and “Only Warning” categories had the highest accuracies, and were similar to each other ($V = 7.0$, $p = 0.577$). The “No Warning” and “Allergen Advisory” categories shared similar accuracies ($V = 101.0$, $p = 0.504$), and were lower than the other two categories. Comparisons between products outside these two subcategories were all statistically significant (all $p < 0.002$).

Warning Types

If a product either contains nuts or could have been exposed to nuts during production, the manufacturer has several options for notifying the consumer. Nine of the products examined included a specific statement about nut content, usually at the end of the ingredient list, such as “Contains almonds” (“Contains” category). Seven products containing nuts did not include an explicit warning, other than advertising the nut content on the front of the package or as part of the list of ingredients (“None” category). Five products included nuts in a specially demarcated listing of potential allergens (“Sectioned” category; products in this category may also have included a “Contains” statement and/or Allergen Advisory statement, but are not counted in those categories). Four products included Allergen Advisory statements such as, “This product may contain nuts” (“May Contain” category). Four products included Allergen Advisory statements such as, “Manufactured in a facility that also produced nut products” (“Facility” category). Among the products examined that did not contain nuts, two included a special “Nut Free” label, while eighteen included no notice about nut content (“None”).

Figure 3 and Figure 4 show violin plots for search duration and decision accuracy, respectively, for these categories. To reduce statistical noise, data points are averaged per each participant and combination of warning type and safe/unsafe category. The resulting data were analyzed using 1-way repeated-measures ANOVAs (duration data) or Friedman tests to control for non-normality (accuracy data). Results show that among unsafe products, warning type significantly affects search duration ($F(4, 124) = 54.79$, $p < 0.001$), with Allergen warnings having the shortest search times, followed by Contains, Facility, and May Contain statements. Unsafe products that did not include an explicit warning statement had search times similar to products with Facility statements (posthoc t-test between None and Facility nonsignificant; all other posthoc comparisons significant, $p < 0.007$). Among nut-free products, those that included a nut-free label were searched significantly faster than those without ($F(1, 31) = 45.21$, $p < 0.001$).

Decision accuracy for unsafe products was significantly affected by warning type ($\chi^2(4) = 31.88$, $p < 0.001$). Allergen and Contains products were associated with the highest accuracy, and Facility and None the lowest. Among nut-free products, those with a nut-free label were categorized more accurately than those without ($\chi^2(1) = 9.0$, $p = 0.003$).

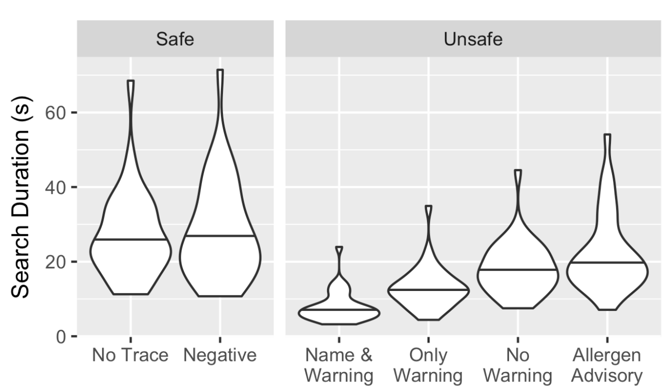

Figure 1: Violin plots of search duration for products in each broad category (products within each category were averaged to produce a single score for each category per participant).

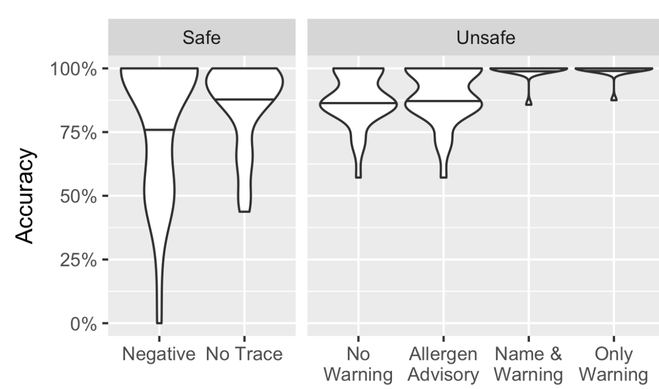

Figure 2: Violin plots of decision accuracy for products in each broad category. Aggregation and labeling as in Figure 1.

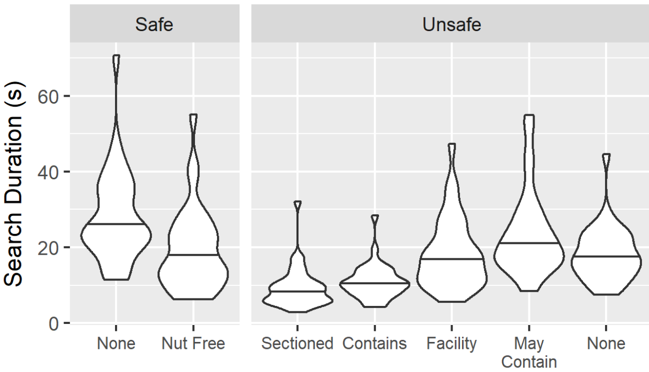

Figure 3: Violin plots of search duration for products in each warning label category. Aggregation and labeling as in Figure 1.

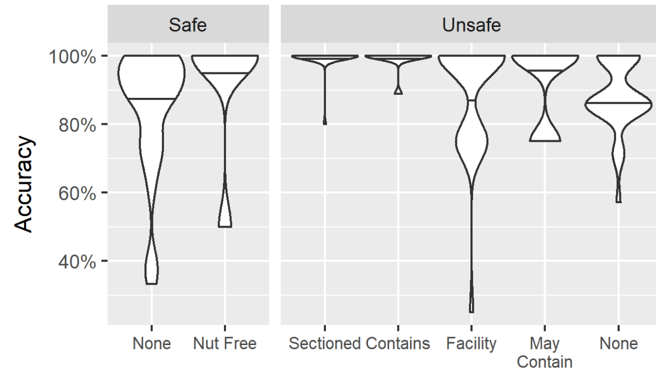

Figure 4: Violin plots of decision accuracy for products in each warning label category. Aggregation and labeling as in Figure 1.

Signal Detection Analysis

Signal detection theory can be applied to the present data set, as it involves classifying the presence or absence of a signal (in this case, nut content) (Stanislaw & Todorov, 1999). Rather than simply classifying responses into correct and incorrect categories, signal detection theory divides decisions into hits (product contained nuts and was identified as such), misses (product contained nuts but was identified as safe), correct rejections (product did not contain nuts and was identified as such), and false positives (product did not contain nuts but was identified as unsafe). This allows for a participant's sensitivity to the task to be separated from his or her decision bias.

Figure 5 (left panel) shows mean search durations for each of the four signal detection categories. Search durations were significantly affected by detection category ($F(3, 82.1) = 29.71$, $p < 0.001$). Follow-up testing shows that most categories had significantly different search times from each other (all $p < 0.001$), except that false positive search durations were not significantly different from misses nor correct rejections. Hits were searched the fastest and misses the slowest.

Given the high overall accuracy in this data set, the measure of sensitivity is not of particular interest. However, the measure of decision bias can provide insight into participants' detection strategies. The majority of participants had a negative bias, indicating a strategy that produced more false positives and incorrectly classified safe products as unsafe. The distribution of bias values (Figure 5, right panel) was significantly different from 0 ($t(31) = -3.29$, $p = 0.003$).

Self-Report Data

In a post-experiment questionnaire, participants were asked to respond to the open-ended question "Is there anything you learned or became more aware of as a result of participating in this study?" In line with previous research, 47% of participants reported aesthetic challenges to label legibility, such as glossy packaging, extremely small print, long lists of ingredients, contrast between text and background, and text printed in crevices of a package rather than on a flat surface. Thirty-one percent of participants noted allergen label inconsistencies as contributing to difficulty in identifying safe to consume products. In particular, participants expressed having a designated area for allergens (i.e. "Contains" statements) made determining product safety much easier than for those products that did not have such statements. Additionally, participants were confused as to why some products had "Contains" statements while other packages only listed allergens in the ingredients list. Participants also found the negative condition (nuts in product name, but not in ingredients) to be misleading. Finally, 44% of respondents expressed empathy for individuals with food allergies, specifically with regards to the amount of time it would take the individual to grocery shop.

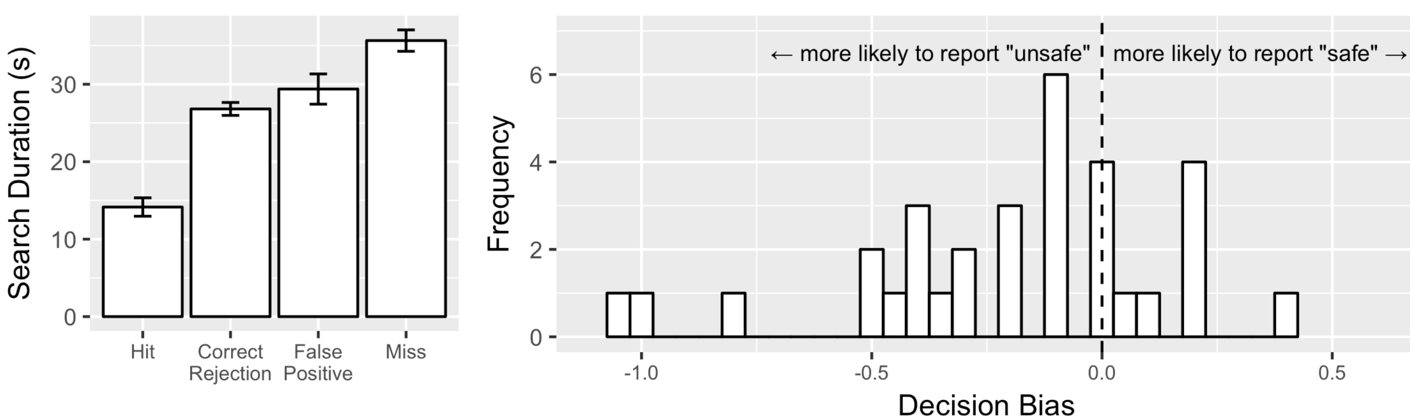

Figure 5: Signal detection metrics. Left Panel: Mean search durations for each of the four signal detection categories. Error bars represent ±1 mean-adjusted standard error. Right Panel: Histogram of decision biases across sample participants. Negative values indicate a decision bias in favor of reporting a product as unsafe.

DISCUSSION

Examining Food Labels as a Strategy for Allergen Avoidance

When products were examined carefully, individuals were generally able to accurately identify both safe and unsafe products. However, ensuring a product was safe, rather than eliminating unsafe products, took significantly more time and led to more errors than identifying a product as unsafe. Older consumers in particular required additional time to safely categorize each item compared to younger consumers. Participants seemed to adopt a "better safe than sorry" mentality; if they were unsure of safety, after a period of time they gave up on searching and defaulted to avoiding the product. Participants in this study were compensated and may have spent more time examining each product than they would have in a grocery store. Higher rates of erroneous product avoidance may take place in real-world settings than occurred in this experiment, which may lead to lost sales of otherwise safe-to-consume products.

Because manufacturers have the ability to change packaging or ingredients, current recommendations advise allergic consumers to examine food labels not just the first time they purchase a product, but every time. Approximately nine months after the initial product purchases, seven products used in this study had completely redesigned packaging or reformatted ingredients sections. This suggests that at-risk individuals may need to expend more time shopping—on every trip—compared to non-allergic customers. Importantly, this study only examined non-food allergic individuals; it is unclear whether consumers with food allergies develop strategies or faster search methods while maintaining high accuracy of safe product identification. Nevertheless, study participants took note of the burden of reading food labels placed on food-allergic consumers and expressed empathy for the additional difficulties nut-allergic individuals experience while grocery shopping.

Characteristics of Difficult-to-Read Labels

Allergen advisory statements are intended to alert consumers to possible cross contamination; however, this study found products with these labels required more time to classify as unsafe as opposed to other unsafe products, including those that had nuts only in the ingredients and no warning label, and were incorrectly categorized as safe just as frequently as products with nuts in the ingredients but no warning label. Variation in allergen advisory statements was also a challenge. Participants noted phrasing and warning labels were different between products and expressed that the inconsistency led to difficulties when classifying certain products. As a result, "May Contain" statements took longer to examine, but "Facility" statements were the least accurate. Consistent with Hefle et al. (2007), it seems consumers are unclear on how different phrasing in allergen advisory warnings implicates product safety.

Products that have "nuts" in the title, but not in the ingredients, were reported as especially challenging because participants expected nuts to appear in the ingredient list. It took participants more time and turns to make a determination, and they were less accurate in their decisions.

Participants described aesthetic challenges to easily classifying products, such as glossy packaging, poor contrast between font print and packaging background, and small font size. While this analysis did not focus on typographic or design characteristics of packaging, further research may quantify how different contrasts, formatting and font size may play a role in safe product identification.

Characteristics of Easy-to-Read Labels

This study identified warning types and styles that make identification of safe and unsafe products faster and more accurate. Safe products with a nut-free label were examined significantly faster and were significantly more likely to be classified as safe than those without a nut-free label. Unsafe products following FALCPA recommendations to include a "Contains" statement were identified faster and more accurately than unsafe products with nuts in the ingredient and no "Contains" statement. Standardizing product labelling, including safe products, unsafe products, and products at risk of cross-contamination may benefit consumers by reducing the amount of time needed to examine a product and clarifying whether or not a product presents a risk of allergic reaction.

CONCLUSIONS

This study found that the phrasing and content of allergen labels on food packaging significantly affected allergen identification accuracy and search times. The lack of standardization in warning labels may create a high burden for allergy sufferers, who must re-assess packages whenever they shop. Labeling safe products as "nut-free" and labeling unsafe products with the FALCPA recommended "Contains" statement improves the speed and accuracy of allergen presence identification. Future research is needed to explore the role of typographic factors and label location on accurate product classification.

REFERENCES

Barnett, J., Leftwich, J., Muncer, K., Grimshaw, K., Shepherd, R., Raats, M. M., … Lucas, J. S. (2011). How do peanut and nut-allergic consumers use information on the packaging to avoid allergens? *Allergy, 66*(7), 969-978. DOI:10.1111/j.1398-9995.2011.02563.x

Food Allergy Research & Education. (2015). *Your Food Allergy Field Guide.* Retreived from www.foodallergy.org

Food and Drug Administration. (2004). Food Allergen Labeling and Consumer Protection Act of 2004. (Title II of Public Law 108-282).

Hefle, S. L., Furlong, T. J., Niemann, L., Lemon-Mule, H., Sicherer, S., & Taylor, S. L. (2007). Consumer attitudes and risks associated with packaged foods having advisory labeling regarding the presence of peanuts. *J Allergy Clin Immunol*, *120*(1), 171-176.

Joshi, P., Mofidi, S., & Sicherer, S. H. (2002). Interpretation of commercial food ingredient labels by parents of food-allergic children. *J Allergy Clin Immunol*, *109*(6), 1019-1021.

Pieretti, M. M., Chung, D., Pacenza, R., Slotkin, T., & Sicherer, S. H. (2009). Audit of manufactured products: Use of allergen advisory labels and identification of labeling ambiguities. *J Allergy Clin Immunol*, *124*(2), 337-341. DOI:10.1016/j.jaci.2009.05.032

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, *31*(1), 137–149. DOI:10.3758/BF03207704