Pediatr Allergy Immunol 2009: 20: 347–352
DOI: 10.1111/j.1399-3038.2008.00791.x

© 2009 The Authors
Journal compilation © 2009 Blackwell Munksgaard

PEDIATRIC ALLERGY AND IMMUNOLOGY

Association analysis of food allergens

Kanagawa Y, Matsumoto S, Koike S, Imamura T. Association analysis of food allergens.
Pediatr Allergy Immunol 2009: 20: 347–352.
© 2009 The Authors
Journal compilation © 2009 Blackwell Munksgaard

Food allergy patients are known to present with allergic reactions to multiple allergens, but extrapolating these associations is difficult. Data mining, a procedure that analyzes characteristic combinations among large amounts of information, is often used to analyze and predict consumer purchasing behaviour. We applied this technique to the extrapolation of food allergen associations in allergy patients. We sent 1510 families our 'Questionnaire survey for the prevention of food allergies'. Responses noting 6549 allergens came from 878 families with 1383 patients, including 402 with anaphylaxis. Some results of the survey have already been published and here we presented the results of our association analysis of combinations of food allergens. Egg, milk, wheat, peanuts, and buckwheat are the most common food allergens. The most common simultaneous combinations of these allergens were 'egg-milk', 'egg-wheat', and 'milk-wheat'. The occurrence probability of a combination (i.e. one person suffering from a certain allergen also suffers from another) is called 'confidence'. Confidence was higher for 'chicken-egg', 'abalone-salmon eggs', and 'matsutake mushroom-milk'. As well, the combinations of 'crab-shrimp', 'squid-shrimp', and 'squid-crab' also indicated higher values in a statistical examination of the occurrence probabilities of these allergen combinations (Z-score). From the results of the association analysis, we speculated that some food allergens, such as abalone, orange, salmon, chicken, pork, matsutake mushroom, peach and apple did not independently induce food allergies. We also found that combinations, such as 'crab-shrimp', 'squid-shrimp', 'squid-crab', 'chicken-beef', and 'salmon-mackerel' had strong associations.

Yoshiyuki Kanagawa[1], Shinya Matsumoto[1], Soichi Koike[1] and Tomoaki Imamura[2]

[1]Department of Medical Informatics and Economics, Division of Social Medicine, School of Medicine, The University of Tokyo, Tokyo, [2]Department of Public Health Policy, Nara Medical University School of Medicine, Nara, Japan

Key words: food allergy; food allergen; multiple allergens; data mining; association analysis

Dr Yoshiyuki Kanagawa, Department of Medical Informatics and Economics, Division of Social Medicine, School of Medicine, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8655, Japan
Tel.: +81 3 5800 8716
Fax: +81 3 5800 8765
E-mail: kanagawa-tky@umin.ac.jp

Accepted 7 January 2008

Patients with food allergies are known to present with allergy symptoms to multiple food allergens (1–7). From the results of our 2003 questionnaire survey, one patient indicated about five kinds of allergic substances. Among patients who had experienced anaphylaxis, each patient presented with anaphylaxis to two or three kinds of antigens. Food allergens are known to vary with countries. In Japan, the most common allergen involved in anaphylaxis is milk, followed by eggs and wheat in order of frequency; whereas in the USA, peanuts are the most common allergen, followed by eggs, fruits/vegetables, wheat, and cow's milk (8). The French Allergy Vigilance Network survey performed by Moneret-Vautrin et al. (9) in France, demonstrated that peanuts, other nuts, shellfish, and lupine and wheat flours were the main allergens in severe anaphylaxis. Thus, it would be fair to assume that allergens vary by region. With regard to food allergies, causative foods for anaphylaxis are different from those for food allergies. However, there have been few detailed investigations into combinations of food allergens. Recent advances in analysis techniques now allow for large amounts of data, including data from medical examinations and questionnaire survey results, to be obtained through medical practices. Conventional data analysis has emphasized the comparison of a single symptom against its cause, the accompanying test results, and other factors. This method cannot be applied to the analysis of the combinations of multiple causes. However,

recent data analysis methods, like data mining, have advanced remarkably (10, 11). Association analysis is a particularly well-known analytical method that can comprehensively process combinations.

Materials and methods

Analysis object data

In July 2003, we used sent our 'Questionnaire survey for the prevention of food allergies' by mail to 1510 Allergy Nationwide Network member families; a nation-wide group of patients with food allergies. In the questionnaire, the following questions concerning the causative agents and so forth of individual patients were posed.

1 At what age did you first develop anaphylactic shock?
2 What were the allergens that caused anaphylaxis?
3 Where did the symptoms of anaphylaxis develop?
4 What was the treatment given for anaphylaxis and did it include hospitalization?
5 Did a doctor diagnose a food allergy? What were the symptoms?

We obtained answers from 878 families (response rate 58%; 878/1510). As a result, responses were obtained from a total of 1403 individual members. Except for 20 members whose attributes and other question fields remained unanswered, 1383 answers were considered valid and were used to summarize food allergens. In this survey, the presence or absence of symptoms of food allergy that respondees experienced (urticaria, asthma, cough, decreased blood pressure, dyspnea, etc.) was checked. Using the results, we selected analysis subjects in this study from those with allergic symptoms. Patient were 778 males, 529 females (gender unknown for 76); mean age: 11.25 yr (youngest: 0; oldest: 93; mode: 5). Many patients with food allergies presented with allergic reactions to multiple food allergens and thus answers were obtained for a total of 6549 food allergens.

Association analysis

Association analysis, also known as 'market basket analysis', is a method used to analyze large amounts of data (10). It is often applied in the retail industry to analyze from cash register records, and to then evaluate the tendency for consumers to simultaneously purchase multiple items. For example, it can be used as a method to evaluate the possibility of an item being contained in the same shopping basket. In association analysis, two items in a shopping basket, such as 'disposable diapers' and 'beer', are considered as individual data items, and the basket is used as a group to bundle these items. This makes it possible to calculate the occurrence ratio of combinations of all items from the object data (groups of many shopping baskets), and to provide the results of calculations to ascertain specific combinations. The extrapolation of multiple food allergen combinations from individual patients usually requires a huge time input. However, the use of association analysis methods makes it possible to analyze associations among large numbers of combinations, and to then conduct a quantitative ranking. Association analysis was applied by positioning each food allergen as an item, and patients as a data group that bundle items (basket).

The basic concept of association analysis is presented in Fig. 1. For example, there are total 1383 patients, including 1055 patients with egg and 830 with milk allergies. There are 757 patients with a milk and egg allergy.

1 Left support (antecedent probability): ratio of meeting the first condition of the total. For example, the ratio of patients with egg allergy is 0.762 = 1055/1383.
2 Right support (prior probability): ratio of meeting the second condition of the total. For example, the ratio of patients with milk allergy is 0.600 = 830/1383.
3 Support (simultaneous probability): ratio of meeting both first and second conditions of the total. For example, the ratio of patients with milk and egg allergy is 0.547 = 757/1383.
4 Confidence (conditional probability): 'support'/'left support'. For example, the ratio of patients with milk allergy among those with egg allergy is 0.717 = 0.547/0.762.
5 Lift value (improvement ratio): 'confidence'/'right support'. For example, the ratio of change when patients with milk allergy are limited to those with a milk and egg allergy is 1.196 = 0.717/0.600.
6 *Z*-score: a statistical score of the occurrence probability of conditions in combination with other conditions.

Analytical method

Number of food allergens per patient. The average number of food allergens per patient was

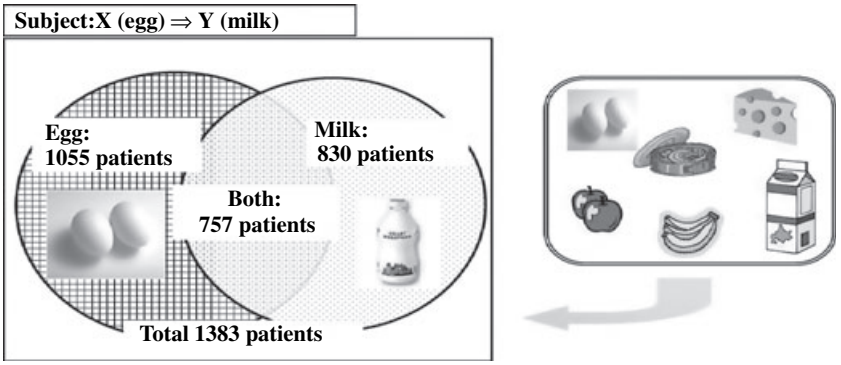


Fig. 1. Concept of association analysis. The occurrence ratio of combination patterns is calculated from relevant data to ascertain 'specific combinations'. Basket: Items bundled as the same group unit; Association: X (pre-conditions) → Y (post-conditions) probability.

calculated by counting food allergens among valid responders to obtain the number of food allergens per patient.

Food allergens. The number of patients presenting with an allergy to a certain food allergen was counted by the respective food allergen.

Combinations of food allergens in patients with multiple food allergens. The number of patients who gave only one or two answers for types of food allergens was counted. For those reported to have two allergens, combinations of these food allergens were then investigated.

As well, combinations were investigated in patients who gave more than two answers for types of food allergens, and the combinations for these were compared with those of patients who gave only two answers. Furthermore, to investigate the strength of the association of two food allergens in patients with two or more food allergens, the strength of the combinations was examined using indexes of L-support, R-support, support, confidence, lift and *Z*-score.

Results

Number of food allergens per patient with food allergies

The number of food allergens per patient is presented in Table 1. The average number of food allergens per patient was 4.87 (minimum: one allergen; maximum: 25 allergens). About 80% of all the patients (1115) presented with allergic reactions to multiple food allergens. Thirty-eight patients did not indicate which allergens.

Food allergens

Table 2 shows the numbers of patients by food allergen among all patients with food allergies, and patients who gave only one or two answers

Table 1. Number of food allergens per patient with food allergies

No. food allergens	No. patients
0	38
1	230
2	263
3	183
4	152
5	94
≥6	423
Total	1383

Table 2. Food allergens among patients with food allergies

Food allergen	Total	Independent	Only two
Egg	1055	99	181
Milk	830	24	131
Wheat	519	18	42
Peanuts	351	2	16
Buckwheat	331	8	13
Salmon eggs	314	3	10
Soybean	293	3	14
Shrimp	285	5	15
Crab	247	1	4
Mackerel	238	6	10
Chicken	225	0	5
Beef	192	2	2
Kiwifruit	186	10	8
Squid	169	2	2
Walnut	168	1	4
Yam	167	2	8
Gelatin	149	2	3
Salmon	141	0	3
Abalone	81	0	0
Pork	77	0	1
Apple	55	0	6
Peach	46	0	3
Orange	43	0	0
Matsutake mushroom	25	0	0
Others	362	42	45
Total	6549	230	526

Total – total number of questionnaire responders. Independent – total patients who gave just one response for types of food allergens. Two – total patients with this food allergen among those with just two food allergens.

regarding food allergens. Thirty-eight patients did not respond about food allergens. Egg, milk, wheat, peanuts, and buckwheat, were the most common, in that order.

Comparing responses, egg, milk, wheat, kiwifruits, and buckwheat were more common among the 230 patients who gave only one response of a causative agent. On the other hand, none gave a response of abalone, orange, salmon, chicken, pork, matsutake mushrooms, peaches, or apples as a causative agent.

Furthermore, 263 patients gave only two responses for food allergens. From these responses, chicken, pork, apples, and peaches, which were not found to be food allergens independently, were identified as food allergens.

Combinations of food allergens in food allergy patients with multiple food allergens

From the 263 patients with only two food allergens, the combinations of the food allergens are presented in Table 3. Among combinations of causative agents of food allergies, combinations with egg, the most common causative agent, were more frequently seen. The following combinations are listed as examples of the most common: egg-milk, followed by egg-wheat, egg-peanuts, milk-wheat, and egg-soybean.

Table 3. Combinations of food allergens among patients with just two food allergens

Food allergen 1	Food allergen 2	No. patients
Egg	Milk	108
Egg	Wheat	18
Egg	Peanuts	9
Milk	Wheat	8
Egg	Soybean	6
Soybean	Wheat	5
Salmon eggs	Egg	5
Buckwheat	Egg	5
Egg	Chicken	4
Yam	Egg	4
Other combinations		91
Total		263

Association analysis was used to evaluate 1115 patients with two or more food allergens to examine the ratio (support value; simultaneous probability) of patients simultaneously having two food allergens. The combinations of egg-milk, egg-wheat, milk-wheat, egg-peanuts, egg-buckwheat, and egg-salmon eggs indicated higher values and proved the most common among the combinations (Table 4).

A person who presents with allergy symptoms to a certain allergen, will simultaneously present with allergy symptoms to another allergen was higher for chicken-egg, abalone-salmon eggs, Matsutake mushroom-milk, and Matsutake mushroom-egg, in that order (Table 5).

Furthermore, when *Z*-scores were calculated, combinations of crab-shrimp, squid-shrimp, squid-crab, chicken-beef, and salmon-mackerel indicated higher values (Table 6).

Discussion

Patients with food allergies are known to often present with allergic reactions to multiple food allergens (1–3, 12). In Japan, egg, milk, and wheat top the list of food allergens. In Korea, among patients 6- to 12-yr old, egg, milk, soybean, peanuts, buckwheat, and wheat are most common allergens, in that order. However, among patients 12- to 15-yr old, milk, egg, soybean, buckwheat, peanuts, and wheat are the most common allergens, in that order. This suggests that there are age-related differences among food allergens (13). Among American children, milk, egg, and peanuts are the main food allergens (5–7). In France, rosaceous fruits are the most common (14%), followed by vegetables in general, milk, crustaceans, shellfish, other fruits (such as avocado and kiwifruit), and eggs (14). Among adults in Germany, nuts, fruits, milk, wine, and vegetables are more common as food allergens (9). Thus, food allergens vary by region and age.

As well, antigenically-similar substances are known to cause similar allergic symptoms in food allergy patients. For example, shrimp or crab-

Table 4. Combinations of food allergens (support value)

Food allergen 1	Food allergen 2	Left support	Right support	Support	Confidence	Lift value	*Z*-score
Egg	Milk	0.7614	0.5980	0.5459	0.7170	1.1990	6.7667
Egg	Wheat	0.7614	0.3738	0.3276	0.4302	1.1508	3.5372
Milk	Wheat	0.5980	0.3738	0.2914	0.4873	1.3036	6.0571
Egg	Peanuts	0.7614	0.2531	0.2336	0.3067	1.2121	3.8532
Egg	Buckwheat	0.7614	0.2386	0.2169	0.2849	1.1940	3.3993
Egg	Salmon eggs	0.7614	0.2263	0.2097	0.2754	1.2169	3.6801

Table 5. Occurrence probability (confidence) that a person with a certain allergen has that allergen in combination with another

Food allergen 1	Food allergen 2	Left support	Right support	Support	Confidence	Lift value	*Z*-score
Chicken	Egg	0.1627	0.7614	0.1584	0.9733	1.2784	3.8923
Abalone	Salmon eggs	0.0586	0.2263	0.0564	0.9630	4.2548	14.0293
Matsutake mushroom	Milk	0.0181	0.5980	0.0174	0.9596	1.6048	2.3513
Matsutake mushroom	Egg	0.0181	0.7614	0.0174	0.9596	1.2604	1.1439
Abalone	Crab	0.0586	0.1786	0.0557	0.9507	5.3228	16.5288
Abalone	Egg	0.0586	0.7614	0.0550	0.9382	1.2322	1.8657
Beef	Egg	0.1388	0.7614	0.1302	0.9375	1.2313	2.9569
Pork	Egg	0.0557	0.7614	0.0521	0.9350	1.2280	1.7840
Orange	Egg	0.0311	0.7614	0.0289	0.9302	1.2217	1.2839
Salmon eggs	Egg	0.2263	0.7614	0.2097	0.9265	1.2169	3.6801

Table 6. Combinations with statistically high probability of occurrence (*Z*-score)

Food allergen 1	Food allergen 2	Left support	Right support	Support	Confidence	Lift value	*Z*-score
Shrimp	Crab	0.2061	0.1786	0.1417	0.6877	3.8507	20.7227
Squid	Shrimp	0.1222	0.2061	0.1063	0.8698	4.2209	19.2518
Squid	Crab	0.1222	0.1786	0.0969	0.7929	4.4394	19.1058
Beef	Chicken	0.1388	0.1627	0.0983	0.7084	4.3540	18.9607
Salmon	Mackerel	0.102	0.1714	0.0817	0.8015	4.6768	18.2337

scallop, peanuts-almonds, peanuts- macadamia nuts, salmon caviar-salmon, and salmon caviar-cod roe are indicated as showing cross-reactivity (14–18). However, it is difficult to performed a survey on cross-reactivity with many food allergy patients and causative agents because of the low precision of sampling of patients' sera and the limited number of samples. Therefore, it is thought useful to investigate combinations of causative agents to prevent the development of symptoms of food allergies. However, such a survey has not been reported before. Here, data mining techniques were applied for the extrapolation of combinations of food allergens in food allergy patients. We compared food allergens among all patients and those with just one food allergen. For food allergens in all patients, egg was 1.27 times more common than milk; however, if patients were limited to those with just one food allergen, egg was about four times more common than milk. The order of frequency of food allergens also differed. For example, kiwifruit ranked 13th among all patients; however, it ranked 4th in patients with just one food allergen. In patients with just one food allergen, abalone, orange, salmon, chicken, pork, matsutake mushroom, peach and apple were not included as food allergens. Thus, we considered that these allergens did not independently induce a food allergy reaction.

We also compared combinations of two food allergens between patients with just two food allergens and all those patients with two or more food allergens. In these cases, combinations mainly with egg (the most common food allergen) were seen in both groups, although the combinations consisted of different food allergens. In patients with just two food allergens, we speculated that milk, wheat, peanut and soybean were likely to be the second food allergen, in that order. When patients presenting with allergy symptoms to two or more food allergens have two food allergens simultaneously, milk, wheat, peanut, buckwheat, salmon eggs were considered likely to be the second food allergen, in that order.

As well, when combinations of food allergens were examined in terms of confidence, chicken-egg, abalone-salmon eggs, matsutake mushroom-milk, matsutake mushroom-egg and abalone-crab showed higher values. Among them, as described above, abalone, matsutake mushroom and chicken were not independent food allergens. Thus, we speculated that these three foods were unlikely to be the first allergen, but were likely to be consumed as the second or later allergen.

'*Z*-score', a statistical score of the occurrence probability of conditions in combination with other conditions, allows not only the evaluation of the numbers of food allergens but also statistical examination of the occurrence probability of food allergen combinations. It allows statistical examination of the association strength between food allergen combinations. As a result of this analysis, *Z*-scores indicated higher values for 'crab-shrimp', 'squid-shrimp', 'squid-crab', 'chicken-beef', and 'salmon-mackerel'. The

combinations were few, but were considered to be strongly associated food allergen combinations. Indeed, cross-reactivity between food allergens such as 'crab-shrimp' has been demonstrated (15, 16). Traditionally, a follow-up study on cross-reactivity between these food allergens would have had no choice other than to depend on clinical experience. However, *Z*-scores allowed us to speculate on combination strengths using association analysis. It is also probable that the cause was not cross-reactivity, but rather the existence of identical proteins, or that the second food allergen was a substance easily acquired. In either case, we consider that extracting combinations with statistically strong associations may be useful for future studies of food allergies.

Conclusion

From our association analysis, we speculated that food allergens, such as abalone, orange, salmon, chicken, pork, matsutake mushroom, peach and apple, do not independently induce food allergy reactions. We also demonstrated that some food allergen combinations, such as 'crab-shrimp', 'squid-shrimp', 'squid-crab', 'chicken-beef', and salmon-mackerel' were strongly associated. Knowing not only the number of food allergens of an allergy but also the kinds of food allergens indicating strong associations allows speculation about items likely to be a second food allergen after the first allergen has been determined. It is usually difficult to perform a survey on cross-reactivities among many foods; however, presenting possible combinations of causative agents by a statistical method allows allergy patients to pay attention to possible second food allergens. This method is expected to be useful for preventing the induction of food allergy symptoms.

Acknowledgments

We deeply appreciate the support given to this study through Grants-in-Aid for Scientific Research 2004 from the Japanese Ministry of Education, Culture, Sports, Science and Technology. We would also like to express our gratitude for the cooperation given by the members of the 'Parent's Association for Food Allergy Patients'. The authors thank Assistant Professor Dr Hideo Yasunaga of the Department of Planning Information and Management, The University of Tokyo Hospital, for critically reviewing the manuscript.

References

1. Iikura Y, Imai Y, Imai T, Akasawa A, Fujita K, Hoshiyama K, Nakura H, Kohno Y, Koike K, Okudaira H, Iwasaki E. Frequency of immediate-type food allergy in children in Japan. Int Act Allergy Immunol 1999: 118: 251–2.
2. Imai T, Iikura Y. The national survey of immediate type of food allergy. Jpn J Allergol 2003: 52: 1006–13 (Japanese).
3. Ebisawa M, Ikematsu K, Imai T, Tachimoto H. Food allergy in Japan. Allergy Clin Immunol Int 2003: 15: 214–7.
4. Sicherer SH, Sampson HA. Food allergy. J Allergy Clin Immunol 2006: 117 (2 Suppl. Mini-Primer):S470–5.
5. Sicherer SH. Food allergy. Lancet 2002: 360: 701–10.
6. Sampson HA. Epidemiology of food allergy. Pediatr Allergy Immunol 1996: 9 (Suppl.): 42–50.
7. Sampson HA. Update on food allergy. J Allergy Clin Immunol 2004: 113: 805–19.
8. Weiss C, Munoz-Furlong A, Furlong TJ, Arbit J. Impact of food allergies on school nursing practice. J Sch Nurs 2004: 20: 268–78.
9. Moneret-Vautrin DA, Kanny G, Morisset M, Rance F, Fardeau MF, Beaudouin E. Severe food anaphylaxis: 107 cases registered in 2002 by the Allergy Vigilance Network. Allergy Immunol 2004: 36: 46–51.
10. Agrawal R, Imielinski T, Swami A. Mining association rules between sets of items in large databases. ACM SIGMOD International Conference on Management of Data 1993: 207–216.
11. Imamura T, Matsumoto S, Kanagawa Y, Tajima B, Matsuya S, Furue M, Oyama H. A technique for identifying three diagnostic findings using association analysis. Med Biol Eng Comput 2007: 45: 51–9.
12. Sicherer SH. Clinical update on peanut allergy. Ann Allergy Asthma Immunol 2002: 88: 350–61.
13. Goh DL, Lau YN, Chew FT, Shek LP, Lee BW. Pattern of food-induced anaphylaxis in children of an Asian community. Allergy 1999: 54: 84–6.
14. Kanny G, Moneret-Vautrin DA, Flabbee J, Beaudouin E, Morisset M, Thevenin F. Population study of food allergy in France. J Allergy Clin Immunol 2001: 108: 133–40.
15. Yan Z, Hiroaki M, Eishin M. Cross-reactivity among shrimp, crab and scallops in a patient with a seafood allergy. J Dermatol 2006: 33: 174–7.
16. Kondo Y, Kakami M, Koyama H, Yasuda T, Nakajima Y, Kawamura M, Tokuda R, Tsuge I, Urisu A. IgE cross-reactivity between Fish Roe (Salmon, Herring and Pollock) and chicken egg in patients anaphylactic to Salmon Roe. Allergol Int 2005: 54: 317–23.
17. Goetz DW, Whisman BA. Occupational asthma in a seafood restaurant worker: cross-reactivity of shrimp and scallops. Ann Allergy Asthma Immunol 2000: 85: 461–6.
18. Ebisawa M, Aihara O, Tamaki J, Urisu A, Kondo N, Oshima Y, Shibata R, Akiyama H, Akazawa A. A study to establish a determination method of causative agents of anaphylaxis caused by foods, etc. and preventive and predictive methods. Report of Research Supported by Health and Labour Sciences Research Grants 2005. (Japanese)