

Differences between observers in interpreting double-blind placebo-controlled food challenges: A randomized trial

Paul L. P. Brand^{1,2} & Marlouk A. Landzaat-Berghuizen¹

¹Princess Amalia Children's Centre, Isala Hospital, Zwolle, The Netherlands; ²UMCG Postgraduate School of Medicine, University Medical Centre and University of Groningen, Groningen, The Netherlands

To cite this article: Brand PLP, Landzaat-Berghuizen MA. Differences between observers in interpreting double-blind placebo-controlled food challenges: A randomized trial. *Pediatr Allergy Immunol* 2015; **25**: 755–759.

Keywords

food allergy; double-blind placebo-controlled food challenge; priming

Correspondence

Professor P.L.P. Brand, Princess Amalia Children's Centre, Isala Hospital, PO Box 10400, 8000 GM Zwolle, the Netherlands
Tel.: +31 38 4245050
Fax: +31 38 4247660
E-mail p.l.p.brand@isala.nl

Accepted for publication 14 November 2014

DOI:10.1111/pai.12313

Abstract

Background: Interpretation of double-blind placebo-controlled food challenges (DBPCFC) can be difficult, particularly with ambiguous subjective symptoms. Early opening of the challenge key (which day is verum and which placebo) may influence the clinician's interpretation of the DBPCFC result.

Methods: Fifty-one clinicians reviewing results of 19 DBPCFCs with ambiguous clinical symptoms were randomized into a key first strategy (opening the DBPCFC key before reviewing the symptoms on both challenge days and deciding on the DBPCFC result) or a symptoms first strategy (reviewing symptoms and interpreting both test days as positive or negative before opening the key and deciding on the DBPCFC result).

Results: The proportion of DBPCFCs considered inconclusive was comparable between the two strategy groups ($p = 0.791$). Participants in the symptoms first group were more likely to consider a DBPCFC as positive (in 14 tests, 73.7%) than subjects in the key first group (four tests, 21.1%). The number of positive tests was higher in the symptoms first group (median 50.0%, interquartile range [IQR] 23.1–76.9%) than in the key first groups (44.0%, IQR 12.0–68.0%, $p = 0.031$). This was independent of the participant's profession (pediatrician or other), age, gender, or years of experience ($p > 0.3$).

Conclusions: Clinicians differ in their interpretation of DBPCFC results when symptoms are ambiguous. Opening the key of a DBPCFC before reviewing and interpreting symptoms significantly reduces the likelihood of the challenge being interpreted as positive. Guidelines for performing DBPCFCs should standardize the moment of opening the challenge key.

Despite their undisputed status as the gold standard in the diagnosis of food allergy (1–3), interpretation of double-blind placebo-controlled food challenges (DBPCFCs) is not always straightforward. Reported potential pitfalls include the difficulty to maintain an appropriate strict diet throughout the oral challenge procedure, augmentation factors such as viral infections, and inadvertent drug use during the procedure (4). The presence of subjective symptoms such as fussing and crying in infants, or abdominal discomfort and irritable behavior in older children, also may hamper the interpretation of DBPCFCs (5, 6). In general, a DBPCFC is considered positive (food allergy confirmed) if the patient has symptoms when exposed to verum (either the same symptoms as the patient presented with, or clear objective allergic symptoms) and not to placebo (7, 8). If the symptoms that the patient originally presented with include such subjective symptoms, parents may recognize this during DBPCFC as suggestive of their child's suspected food allergy (6). If such symptoms occur during a

DBPCFC, it is recommended to continue the incremental food challenge until either objective symptoms occur or the subjective symptoms are repeated at subsequent steps of the challenge (4, 8). The considerable differences that have been reported in the prevalence of reactions to placebo foods, ranging from 2.8 to 29.3% (6, 7, 9), illustrate that the interpretation of signs and symptoms during a DBPCFC is subject to potential observer biases. These include the perceived pretest probability of food allergy, which partly depends on the timing of unblinding the key of the DBPCFC (which day was placebo and which was the verum day). A clinician may either interpret signs and symptoms of each challenge day before or after opening the challenge key. The importance of this sequence has not been addressed in published DBPCFC guidelines (2, 8, 10, 11) or, to our knowledge, in clinical studies. The aim of this study was to assess the impact of timing of opening the DBPCFC key on its interpretation and outcome.

Methods

This was a randomized controlled trial comparing two strategies in interpreting results of DBPCFCs in children, the 'key first' and 'symptoms first' strategies, respectively. In the 'key first' strategy, clinicians first opened the DBPCFC key (so they knew which was the verum and which the placebo day), then reviewed, and interpreted symptoms; in the 'symptoms first' strategy, clinicians first evaluated symptoms on each challenge day before they opened the DBPCFC key.

Participants

Study participants were pediatric allergists, general pediatricians, pediatric specialist registrars, and pediatric nurse specialists with at least 2-yr experience in interpreting DBPCFCs. We approached clinicians in 15 hospitals in the Netherlands (five tertiary care university children's hospitals and 10 secondary care general hospitals). Participants were randomized to either the key first or the symptoms first strategy, using block randomization per participating hospital. Randomization sequence was performed using a computer-generated random number list. Participants were randomized to strategy group before being sent study materials.

DBPCFC results

We used data from 19 previously performed DBPCFCs in our hospital (15 in boys; age ranging from 4 months to 12 yr, median age 2.8 yr; 12 challenges to cow's milk, six to peanut, and one to hen's egg), which had proven difficult to interpret by the attending physician, as published previously (6). All these DBPCFCs were performed following a standardized protocol, as previously described in detail (6). Briefly, challenges with verum and placebo were performed in random order on two half-days, at least 1 wk apart. Test foods were offered in six incremental steps until a cumulative dose had been ingested that would elicit a reaction in those allergic to the food (12). Every symptom was recorded by specifically trained pediatric nursing staff on a standardized case record form. Challenges were stopped if the patient had objective allergic symptoms or repeated subjective symptoms, as previously described (6). During both challenge days, patients and their parents, nursing staff and attending physicians remained blinded for the test key. All symptoms were therefore recorded in a double-blind fashion, also for the purpose of this analysis.

Methods

Structured case report forms with data on the symptoms recorded (in a double-blind fashion) on both challenge days of each of the 19 DBPCFCs (example in Appendix A) were sent to study participants, along with instructions consistent with the strategy group the participant had been randomized to (key first or symptom first). The key for each challenge was included in the study package in sealed opaque envelopes. The key first strategy participants were instructed to open the key envelope first, then to review and interpret the symptoms on both

challenge days, to interpret each challenge day as positive or negative, and finally, to draw a final conclusion on the challenge result (positive or negative). The symptom first strategy participants were instructed to review and interpret the symptoms on both challenge days first (including a conclusion for each test day), then open the key, and then draw a final conclusion on the overall challenge result. They recorded their interpretation of each challenge test day and their final conclusion on a registration form which they returned to the investigators. Participants were strictly instructed to adhere to their allocated challenge interpretation strategy, and not to discuss their findings and interpretations with other participants until all had returned their written reports to the investigators.

Ethics

Because this study involved only anonymized data from patients, and study participants were healthy volunteers (clinicians), this study was exempt from formal ethical review under Dutch law. Consequently, the Isala Hospital's ethical review board approved of the study after an abbreviated procedure. All participants provided written informed consent.

Statistical analysis

We used the non-parametric sign test to compare the likelihood of positive or negative challenge results between both challenge interpretation strategies. Results were also analyzed after taking inconclusive challenge test results into account in three ways: by deleting inconclusive test results, by counting them as a positive challenge, or by counting them as a negative challenge. Results were analyzed with SPSS (SPSS Inc, Chicago IL USA) for Windows, version 20.

A power analysis performed before the study showed that 25 participants in each strategy group would be needed to achieve 90% power at α of 0.05, assuming prevalence of positive tests to be 20% (6).

Results

Participants

There were 51 participants from 15 hospitals: 15 general pediatricians, 14 pediatric allergists, 10 pediatric nurse specialists, one dietician, and 11 pediatric specialist registrars. Twenty-five participants were randomized to the key first and 26 to the symptom first strategy, respectively. Characteristics of the two study groups are presented in Table 1. There were more pediatricians in the symptoms first group than in the key first group.

DBPCFC interpretation in two strategy groups

Results of the interpretation of DBPCFCs in both strategy groups are presented in Table 2. A number of participants reported that they were unable to interpret some of the DBPCFCs as positive or negative and therefore rendered an

Table 1 Characteristics of the two study groups

	Symptoms first (n = 26)	Key first (n = 25)
Male gender	7 (27%)	8 (32%)
Mean (SD) age	45 (9)	40 (10)
Pediatrician	19 (73%)	10 (40%)
Mean (SD) years of experience	12.8 (8.7)	11.3 (8.1)

interpretation of 'inconclusive'. Only one challenge test result was considered positive by all participants. There were considerable differences in the interpretation of DBPCFC test results between clinicians. This was not due to differences in tests being evaluated as 'inconclusive' between groups: the median percentage of inconclusive interpretations of DBPCFCs was 7.7 (IQR 0–15.4) and 4.0 (0–16) in the symptoms first and key first groups, respectively (sign test, $p = 0.791$).

The median percentage of positive interpretations of DBPCFCs was 50.0 (interquartile range [IQR] 23.1–76.9) in the symptoms first group and 44.0 (IQR 12.0–68.0) in the key first group (sign test, $p = 0.031$). This difference remained significant for each of the three methods to take inconclusive test results into account (all p values <0.035). The likelihood of interpreting a DBPCFC as positive was independent of the participant's profession (pediatrician or other), age, gender, or years of experience (all p values >0.3). There were no differences between the two strategy groups in the interpretation of DBPCFC results performed in children younger than 3 yr ($p = 1.00$), but in all 9 DBPCFCs performed in children 3 yr of age and older, participants in the symptoms first group were more likely to consider test results as positive than those in the key first group (sign test, $p = 0.004$).

Discussion

This study shows that the likelihood of interpreting a DBPCFC as positive is influenced by the sequence of interpreting the symptoms during both challenge days and obtaining knowledge of the test key. Participants in the symptoms first strategy group were significantly more likely to interpret a DBPCFC as positive (in 74% of tests) than participants in the key first strategy group (21%, $p = 0.031$). This was independent of the participant's profession (pediatrician or other), age, gender, and years of experience. These results strongly suggest that the sequence of reviewing the symptoms during challenge days and unveiling the challenge key should be standardized in DBPCFC protocols and guidelines.

Because this is the first study examining the effect of the sequence of symptom interpretation and challenge key knowledge on the DBPCFC interpretation, our study results cannot be compared to earlier work. Studies from social psychology may help to explain why knowledge of the key before interpreting the challenge symptoms may have a significant impact on the interpretation of the challenge test result. The integration and evaluation of work-related activities and experience leads to a foundation upon which future clinical

Table 2 Interpretation of symptoms during 19 DBPCFCs by 51 clinicians in two different strategy groups: symptoms first and key first (see text). Participants in the symptoms first strategy group were more likely to consider a DBPCFC as positive (in 14 tests, 73.7%) than subjects in the key first strategy group (four tests, 21.1%)

	Symptoms first (n = 26)		Key first (n = 25)	
	Test judged as positive (%)	Test considered inconclusive (%)	Test judged as positive (%)	Test considered inconclusive (%)
Test 1	73.1	11.5	52.0	12.0
Test 2	53.8	15.4	52.0	4.0
Test 3	38.5	7.7	16.0	32.0
Test 4	7.7	0	0	0
Test 5	11.5	0	4.0	4.0
Test 6	26.9	3.8	12.0	12.0
Test 7	84.6	0	76.0	0
Test 8	96.2	0	88	0
Test 9	80.8	11.5	68.0	12.0
Test 10	76.9	15.4	72.0	4.0
Test 11	3.8	7.7	12.0	0
Test 12	61.5	23.1	56.0	16.0
Test 13	50.0	3.8	12.0	20.0
Test 14	23.1	15.4	24.0	4.0
Test 15	30.8	34.6	44.0	20.0
Test 16	57.7	11.5	44.0	12.0
Test 17	30.8	0	36.0	0
Test 18	100	0	100	0
Test 19	23.1	11.5	12	16.0

impressions, judgements, and decisions are based (13). When encountering a clinical scenario, activated personal knowledge and experience influence the decisions the physician makes. This involves two key mechanisms: framing and priming (13, 14). Framing relates to the way in which the presentation of data interprets their interpretation and impact. In an experiment involving patients, medical students, and physicians, subjects were asked to imagine that they had lung cancer and to choose between the two therapies on the basis of life-expectancy data after treatment. The data on the effects of treatment were presented either as 90% short-term survival after surgery or as 10% death after surgery. Significantly, more subjects chose surgery when presented with survival data (85.5%) than when presented with mortality data (65.5%) (15). Priming relates to an implicit memory effect in which exposure to a stimulus influences the response to a later, non-related stimulus (14). In a classic experiment, psychology students primed with words associated with elderly people (such as Florida, wrinkles, and gray) showed a significantly reduced walking speed as compared to students primed with neutral words (16). In our study population, all participants can be expected to have considerable foundation by their previous experience with DBPCFCs. Because most food challenges are negative (3, 6, 17), participants are likely to be primed toward a negative challenge test results. Opening the key before

interpreting the symptoms observed during the challenge (key first strategy) may have augmented this priming toward a negative challenge result. If, by contrast, clinical data are framed as a neutral set of observations without knowledge of the challenge key (symptom first strategy), clinicians may be more likely to interpret the test as positive. It appears that this is primarily the case in children aged 3 yr and over. These children are more likely to express subjective symptoms than younger children (6), which may be related to the anxiety associated with food allergy testing (18).

The main strength of this study is that it is the first to examine differences between observers in the interpretation of DBPCFC test results and to study the effect of the sequence of opening the challenge key and reviewing the symptoms on the interpretation of the test result. The relatively large number of study participants from different pediatric health-care settings increases the generalizability of findings. The main weaknesses include the relative small number of challenge test results presented to the participants, and the fact that these were all challenges in which the clinical symptoms were more or less ambiguous. It is likely that the observer bias effect of opening the challenge key before reviewing and interpreting the symptoms is less important in

children with a clear-cut DBPCFC result. In addition, the distribution of pediatricians among the two strategy groups was unequal (Table 1). It is unlikely that this had a major influence on our results, however, because pediatricians were equally likely to interpret DBPCFCs as positive as the other clinicians.

In conclusion, this study shows that clinicians differ in their interpretation of DBPCFC results when symptoms are ambiguous. This highlights a limitation of the DBPCFC in the diagnosis of food allergy, because even this procedure, which is universally considered to be the gold standard of food allergy diagnosis (2, 11), is subject to the biases of human interpretation of data (6). Opening the key of a DBPCFC before reviewing and interpreting the symptoms occurring on both test days significantly reduces the likelihood of the challenge being interpreted as positive. Guidelines for performing DBPCFCs should standardize the moment of opening the challenge key.

Acknowledgements

The authors wish to thank all the participating clinicians for their invaluable contribution to this article by interpreting the DBPCFC test results.

References

- Burks AW, Tang M, Sicherer S, et al. ICON: food allergy. *J Allergy Clin Immunol* 2012; **129**: 906–20.
- Boyce JA, Assa'ad A, Burks AW, et al. Guidelines for the Diagnosis and Management of Food Allergy in the United States: summary of the NIAID-Sponsored Expert Panel Report. *J Allergy Clin Immunol* 2010; **126**: 1105–18.
- Chafen JJ, Newberry SJ, Riedl MA, et al. Diagnosing and managing common food allergies: a systematic review. *JAMA* 2010; **303**: 1848–56.
- Niggemann B, Beyer K. Pitfalls in double-blind, placebo-controlled oral food challenges. *Allergy* 2007; **62**: 729–32.
- Gellerstedt M, Magnusson J, Grajo U, Ahlstedt S, Bengtsson U. Interpretation of subjective symptoms in double-blind placebo-controlled food challenges - interobserver reliability. *Allergy* 2004; **59**: 354–56.
- Oole-Groen CJ, Brand PL. Double-blind food challenges in children in general paediatric practice: useful and safe, but not without pitfalls. *Allergol Immunopathol (Madr)* 2014; **42**: 269–74.
- Vlieg-Boerstra BJ, van der Heide S, Bijleveld CM, Kukler J, Duiverman EJ, Dubois AE. Placebo reactions in double-blind, placebo-controlled food challenges in children. *Allergy* 2007; **62**: 905–12.
- Niggemann B. When is an oral food challenge positive? *Allergy* 2010; **65**: 2–6.
- Ahrens B, Niggemann B, Wahn U, Beyer K. Positive reactions to placebo in children undergoing double-blind, placebo-controlled food challenge. *Clin Exp Allergy* 2014; **44**: 572–78.
- Bock SA, Sampson HA, Atkins FM, et al. Double-blind, placebo-controlled food challenge (DBPCFC) as an office procedure: a manual. *J Allergy Clin Immunol* 1988; **82**: 986–97.
- Sampson HA, van Gerth WR, Bindslev-Jensen C, et al. Standardizing double-blind, placebo-controlled oral food challenges: American Academy of Allergy, Asthma & Immunology-European Academy of Allergy and Clinical Immunology PRACTALL consensus report. *J Allergy Clin Immunol* 2012; **130**: 1260–74.
- Vlieg-Boerstra BJ, Bijleveld CM, van der Heide S, et al. Development and validation of challenge materials for double-blind, placebo-controlled food challenges in children. *J Allergy Clin Immunol* 2004; **113**: 341–46.
- Teunissen PW, Stapel DA, Scheele F, et al. The influence of context on residents' evaluations: effects of priming on clinical judgment and affect. *Adv Health Sci Educ Theory Pract* 2009; **14**: 23–41.
- Tversky A, Kahneman D. The framing of decisions and the psychology of choice. *Science* 1981; **211**: 453–58.
- McNeil BJ, Pauker SG, Sox HC Jr, Tversky A. On the elicitation of preferences for alternative therapies. *N Engl J Med* 1982; **306**: 1259–62.
- Bargh JA, Chen M, Burrows L. Automaticity of social behavior: direct effects of trait construct and stereotype activation on action. *J Pers Soc Psychol* 1996; **71**: 230–44.
- Fleischer DM, Bock SA, Spears GC, et al. Oral food challenges in children with a diagnosis of food allergy. *J Pediatr* 2011; **158**: 578–83.
- Le TM, Zijlstra WT, van Opstal EY, et al. Food avoidance in children with adverse food reactions: influence of anxiety and clinical parameters. *Pediatr Allergy Immunol* 2013; **24**: 650–55.

Appendix A

Example of case record form used to present symptoms on both DBPCFC days to study participants

History

Thirteen-month-old boy, who has had symptoms of erythema, urticaria, angioedema, colic, rhinitis, wheeze, and dyspnea while using cow's milk formula after weaning from breast-feeding during the third and fourth month of life, and who has been using extensively hydrolyzed formula since then. The patient has been prescribed fluticasone and salbutamol for recurrent wheeze, and emollient and topical corticosteroid ointments for atopic dermatitis.

Challenge

DBPCFC of cow's milk in the same extensively hydrolyzed formula as the child is using at home.

Symptoms day 1

Challenge performed 4 days after MMR immunization. Patient has a runny nose. Mild eczema (six areas, six-sign

[SASSAD] score 11). Mild vomiting (one mouthful) after 1st step. Similar vomiting after 4th step, along with loose stools and crying and fussing. Erythema of left elbow fold after 6th and final step, no increase in eczema score (SASSAD). Parents and attending nurse concluded that this was an inconclusive test day.

Late symptoms after day 1

Increase in eczema, with increased fussing and crying, and loose stools during the first 2 days after the challenge. Increase in viral cold symptoms, mild wheeze.

Symptoms day 2 (7 days after day 1)

No runny nose. Mild eczema, SASSAD score 11. After 4th step, parents reported that their child's skin felt warmer, no visible erythema. Repeated vomiting after step 6. SASSAD score slightly increased after challenge (score: 14). Parents and attending nurse concluded that this was an inconclusive test day.

Late symptoms after day 2

Increase in eczema, some fussing and crying during the first 3 days after the challenge.