

Statistical issues in clinical trials that involve the double-blind, placebo-controlled food challenge

Vernon M. Chinchilli, PhD,^a Laura Fisher, MD,^b and Timothy J. Craig, DO^b Hershey, Pa

The double-blind, placebo-controlled food challenge is a rigorous tool that has become popular for evaluating adverse reactions to foods. The standard use of the double-blind, placebo-controlled food challenge has been to document food allergies for individual patients, but it recently has been gaining acceptance as a procedure for investigating the effectiveness of therapies to prevent/minimize food-induced anaphylaxis. The purpose of this study was to describe the statistical design and analysis issues for clinical trials that use the double-blind, placebo-controlled food challenge in measuring sensitivity to food allergens. Nonparametric tests for within-group and between-group comparisons are described, as well as a discrete-time survival analysis. The statistical methods are applied to simulated data from a clinical trial that compares control therapy and experimental therapy groups. The results indicate that the experimental therapy is significantly better than control in improving the tolerance to peanut flour in patients with peanut allergy. Although simple nonparametric tests for within-group and between-group comparisons are easy to apply, a discrete-time survival analysis provides the best approach because of its flexibility in accounting for important independent variables (regressors) and longitudinal data. Statistical software packages can be adapted to perform such analyses. (*J Allergy Clin Immunol* 2005;115:592-7.)

Key words: Anaphylaxis, food allergy, statistical design, statistical analysis, nonparametric analysis, discrete-time survival analysis

The double-blind, placebo-controlled food challenge (DBPCFC) is a rigorous tool that has become popular for evaluating adverse reactions to foods. Numerous researchers have attempted to develop standardized protocols for the food preparation component of the DBPCFC and/or methodology for the validation of challenge materials.¹⁻³

The standard use of the DBPCFC has been to document food allergies for individual patients. For example,

Abbreviations used

ANCOVA: Analysis of covariance

DBPCFC: Double-blind, placebo-controlled food challenge

ML: Maximum likelihood

SBFC: Single-blind food challenge

Sampson and Ho⁴ examined the sera of 196 patients with food allergies who had undergone the DBPCFC. The objective of the study was to determine the levels of sera antibodies to egg, milk, peanut, soy, wheat, and fish, and to correlate them with the clinical responses uncovered by the DBPCFC. The authors reported that allergic responses to egg, milk, peanut, and soy accounted for 87% of the confirmed reactions. In another example, Rodriguez et al⁵ used the DBPCFC to determine whether patients' claims about adverse reactions to melon were accurate, and if so, the nature of the clinical manifestations. They observed that the DBPCFC confirmed melon allergy in 36% (19/53) of the patients. Oral allergy syndrome was the most frequently reported syndrome at 26% (14/53); however, 2 patients had life-threatening reactions.

Recently, the DBPCFC has been gaining acceptance as a procedure for investigating the effectiveness of therapies to prevent/minimize food-induced anaphylaxis. For example, Leung et al⁶ incorporated the DBPCFC as a procedure to investigate the effectiveness of anti-IgE therapy in 82 patients with peanut allergy.

The intent of this article is to discuss the statistical issues involved with the use of the DBPCFC in a randomized clinical trial to investigate whether an experimental therapy is more effective than a control therapy in alleviating food-induced anaphylaxis.

METHODS

Statistical design

The logistical difficulties that researchers may encounter with the conduct of the DBPCFC are described elsewhere.¹⁻³ In this section, we describe the statistical issues that researchers need to consider with the conduct of DBPCFC within the context of a clinical trial. With respect to notation, let d_0 denote the initial dose of the food to be invoked in the DBPCFC, and let $d_1, d_2, \dots, d_K$ denote

From ^athe Department of Health Evaluation Sciences and ^bthe Department of Medicine, Penn State College of Medicine.

Disclosure of potential conflict of interest: The authors have no conflicts of interest to disclose.

Received for publication December 6, 2004; revised January 3, 2005; accepted for publication January 4, 2005.

Reprint requests: Vernon M. Chinchilli, PhD, Department of Health Evaluation Sciences, A210, Penn State College of Medicine, 600 Centerview Drive, Suite 2200, Hershey, PA 17033-0855. E-mail: vchinch@psu.edu.

0091-6749/\$30.00

© 2005 American Academy of Allergy, Asthma and Immunology

doi:10.1016/j.jaci.2005.01.008

the increasing doses for the remaining steps of the DBPCFC. If a patient displays clinical signs of allergic response at a particular dose, the patient does not proceed to the next dose, and the DBPCFC is terminated. Clinical signs of allergic response could include nausea, vomiting, diarrhea, rash, and so forth, and each symptom could be graded on an ordinal scale to denote severity (eg, absent, mild, moderate, severe).¹ Although it is possible to construct a total score by summing the scores across all of the clinical signs, this may not be very useful for statistical analysis because the challenge usually is discontinued if any of the clinical signs reach or exceed the mild category.

Ideally, the doses should be selected to form a multiplicative scale of progression rather than an arithmetic scale of progression. For example, doubling doses yield a scheme in which

$$d_1 = 2d_0, d_2 = 2d_1 = 2^2d_0, \\ d_3 = 2d_2 = 2^3d_0 \dots d_K = 2d_{K-1} = 2^Kd_0$$

The practical reasons for considering a multiplicative scheme for the dose progression are that (1) arithmetic schemes may progress too slowly in eliciting clinical signs of an allergic response, and (2) it should be straightforward to manufacture the capsules for the doses. The statistical reason for considering a multiplicative scheme for the dose progression is that a logarithmic transformation of the dose converts to an integer if the initial dose, d_0 , is standardized to 1 unit, eg, when a doubling scheme is invoked.

$$\log_2(d_0) = 0, \log_2(d_1) = 1, \log_2(d_2) = 2, \\ \log_2(d_3) = 3 \dots \log_2(d_K) = K$$

The outcome variable is the lowest dose that elicits clinical signs of allergic response. For mathematical convenience, a patient can be assigned a score of 0, 1, 2...K instead of $d_0, d_1, d_2 \dots d_K$. If a patient does not respond to any of the K doses, then the outcome is right-censored at d_K , and the patient arbitrarily is assigned the dose score of d_{K+1} , the next highest dose that would have been used in the challenge (or, alternatively, the ordinal score of K+1). Usually, the DBPCFC is administered at baseline to assess patient eligibility for the current study. Thus, a patient with a right-censored value of d_{K+1} (or the ordinal score of K+1) at baseline is not eligible to continue through the study, whereas a right-censored value during a postbaseline administration of the DBPCFC is acceptable.

Some researchers may be interested in using the cumulative administered dose as the outcome, instead of the final administered dose. Although this would not alter the results of the statistical analysis within the current study, using the cumulative administered dose does provide the opportunity to compare the results of the current study with results from other studies.

Because the researcher assesses the absence/presence of clinical signs at an administered dose, it is not possible

to interpolate between doses. This is true even if the researcher assigns a severity score to the response (eg, absent, mild, moderate, severe). Thus, the DBPCFC is unlike other types of challenges in which the response is measured on a continuous scale and interpolation plays an important role, such as the methacholine challenge with spirometry performed in patients with asthma. In the latter case, spirometry FEV₁ is measured after each administration of a methacholine concentration, and the methacholine PC₂₀ is determined via interpolation.

The frequency of administering the DBPCFC during the course of a clinical trial is an important issue. Ideally, the DBPCFC should be administered at least twice, at baseline and at study end. If the study is lengthy, other administrations of the DBPCFC can be inserted between the baseline and study end. At the baseline DBPCFC, the patient should exhibit a positive and a negative result, presumably caused by the food and the placebo, respectively, to be eligible for the trial. The DBPCFC at study end is essential to determine whether the patients in the experimental therapy and control groups have improved. In the study by Leung et al⁶ on peanut allergy, the researchers administered the DBPCFC at baseline, but only administered a single-blind food challenge (SBFC) at study end. The SBFC may be sufficient for any postbaseline challenges because the food-specific allergy already has been established for each patient at baseline. In fact, the clinical signs displayed by a particular patient during the baseline DBPCFC could serve as the focus during the postbaseline SBFC. Obviously, using the DBPCFC for postbaseline challenges would enhance the validity of the trial, but the gain may be minimal and not worth the time and effort.

Another feature of the study by Leung et al⁶ is that the researchers modified the dose progression for the SBFC at study end, ie, they used a higher initial dose and higher ending doses than those used at the baseline DBPCFC. It is recommended that the dose progression ($d_0, d_1, d_2 \dots d_K$) remain the same during baseline and postbaseline challenges. The rationale for selecting the dose progression in the baseline challenge should apply to the postbaseline challenges as well. In addition, some of the statistical analyses based on change-from-baseline scores will not be appropriate.

Statistical analysis

Assuming that the trial contains a baseline DBPCFC and a study-end DBPCFC (or SBFC), the primary outcome variable can be constructed as a change-from-baseline score. For convenience, the statistical analysis is presented in terms of the ordinal scores (0, 1, 2...K+1) instead of the dose values ($d_0, d_1, d_2 \dots d_{K+1}$) because the study-end DBPCFC could yield right-censored values. The dose value of d_{K+1} was chosen arbitrarily and indicates that the patient did not respond to any of the administered doses. Using the ordinal score of K+1 instead of d_{K+1} is a better indicator of right-censoring and a better indicator that it could be any dose beyond d_K .

TABLE I. Sample sizes (per treatment arm) for various hazard ratios (1.5, 2.0, 2.5, 3.0) and censoring rates (0%, 10%, 20%, 30%) for a 2-sided, .05 significance level test with 90% statistical power

Hazard ratio	0% Censoring	10% Censoring	20% Censoring	30% Censoring
1.5	128	142	160	182
2.0	44	49	55	63
2.5	25	28	31	36
3.0	18	20	23	26

Because the possible scores at the baseline DBPCFC are 0, 1, 2...K and the possible scores at the study-end DBPCFC are 0, 1, 2...K+1, the change-from-baseline scores are the integers that range from -K through K+1. One reason for using the same dose progression ($d_0, d_1, d_2 \dots d_K$) at baseline and end of study is as follows. Suppose that the dose progression at baseline is ($d_0, d_1, d_2 \dots d_K$) and that the dose progression at the end of the study is ($d_2 \dots d_K, d_{K+1}, d_{K+2}$). Suppose that a patient responds at dose d_1 during the baseline DBPCFC, but the same patient responds at dose d_2 during the study-end DBPCFC. The change-from-baseline score for this patient would be 1, which could be misleading and inaccurate, because the patient might have responded to dose d_1 during the study-end challenge if dose d_1 had been administered.

Let η_E and η_C denote the population median of the change-from-baseline score for the experimental therapy group and the control group, respectively. Also, let n_E and n_C denote the corresponding sample sizes for the 2 groups in the trial. If the experimental therapy (control) has improvement between the baseline and the end of study, then η_E (η_C) would be positive. The sample medians, as well as the sample first and third quartiles, should be calculated for each group. Using these sample percentiles for descriptive statistics, rather than means and SDs, is recommended because the change-from-baseline score is ordinal and not continuous. The sign test and the Wilcoxon signed-rank test can be considered for testing the within-group null hypothesis $H_0: \eta_E = 0$ ($H_0: \eta_C = 0$) versus a 1-sided or 2-sided alternative hypothesis.⁷ The sign statistic is calculated as the proportion of patients within the group whose change-from-baseline score exceeds 0, out of all patients with a nonzero score. This proportion follows a binomial distribution so it is straightforward to perform the hypothesis testing and get a P value. The Wilcoxon signed-rank test is more complex to compute, but it has more statistical power than the sign test in most circumstances. It does require, however, that the data follow a symmetric distribution (normal or nonnormal). If the change-from-baseline data are strongly asymmetric, which can be assessed via a simple histogram, then the sign test may perform better. To compute the Wilcoxon signed-rank test, the absolute values of the change-from-baseline scores are ranked from smallest to largest, and then the test statistic is constructed as the sum of the ranks of the positively valued raw scores. The

permutation distribution of all possible values of the test statistic, under the assumption that the null hypothesis is true, can be tabulated to calculate the 1-sided or 2-sided P value. Most statistical software packages provide canned programs for the sample medians and quartiles and for conducting the sign test and the Wilcoxon signed-rank test.

Ordinarily, the primary comparison of interest will be the between-group test of experimental therapy versus control, as indicated by the null hypothesis $H_0: \eta_E = \eta_C$. Similar to the within-group comparisons, a nonparametric test is recommended because of the ordinal scale of the change-from-baseline scores. Therefore, an appropriate nonparametric test in this situation is the Wilcoxon rank sum test (or its equivalent, the Mann-Whitney test).⁷ The test statistic is constructed by pooling the scores from the 2 groups, ranking them from smallest to largest, and then summing the ranks of the experimental therapy group. Similar to the within-group nonparametric tests, (1) the permutation distribution under the null hypothesis can be tabulated to calculate the 1-sided or 2-sided P value, and (2) statistical software packages usually include the Wilcoxon rank sum test or the Mann-Whitney test.

A more appealing approach for the between-group comparison is a discrete-time survival analysis. Instead of constructing and analyzing the change-from-baseline score, the study-end score becomes the outcome variable in the survival analysis and the baseline score functions as a regressor. As described, the scores are integer-valued and range from 0 through K+1. Because a patient progresses through increasing doses of the food item until an anaphylactic reaction is observed, the scores of 0 through K+1 correspond to a discrete-time survival process in which the last category, K+1, represents right-censoring.

A discrete-time survival analysis is based on a model for the conditional probability of a patient responding to dose d_k , $k = 0, 1, 2 \dots K$, given that the patient has not responded to previously administered doses. This conditional probability is called the *hazard function*, and an appropriate model to describe this process is the discrete-time hazard function

$$\lambda(k; x_1, x_2 \dots x_r) = 1 - \exp\{-\exp(\alpha_k + x_1\beta_1 + x_2\beta_2 + \dots + x_r\beta_r)\}$$

where $k = 0, 1 \dots K$ denotes the dose stage and $x_1, x_2 \dots x_r$ denote a set of r regressors, such as baseline score, gender, age, and so forth. This particular form for the discrete-time hazard function is called the *extreme value hazard function*, and a statistical analysis based on this model requires a logistic regression analysis with the complementary log-log link function.⁸ The extreme value hazard function is the discrete time analogue of the Cox proportional hazard function. Researchers may not be familiar with the extreme value hazard function because they are more familiar with the exponential hazard function used for continuous time survival analysis. The parameters α_0 ,

TABLE II. Simulated data set of trial to compare experimental therapy and control therapy in patients with peanut allergy

Control therapy scores (subject, baseline, study-end)		Experimental therapy scores (subject, baseline, study-end)	
01, 5, 5	26, 1, 1	01, 1, 0	26, 1, 1
02, 0, 0	27, 4, 5	02, 0, 0	27, 5, 6
03, 3, 3	28, 1, 2	03, 4, 4	28, 1, 1
04, 1, 0	29, 0, 0	04, 4, 4	29, 4, 3
05, 5, 5	30, 1, 2	05, 2, 2	30, 4, 6
06, 1, 1	31, 0, 1	06, 5, 4	31, 5, 6
07, 5, 6	32, 0, 1	07, 2, 2	32, 3, 5
08, 2, 1	33, 0, 1	08, 0, 3	33, 3, 5
09, 0, 1	34, 2, 3	09, 5, 5	34, 3, 3
10, 0, 1	35, 4, 3	10, 4, 5	35, 0, 2
11, 0, 0	36, 0, 0	11, 2, 1	36, 4, 6
12, 2, 1	37, 2, 2	12, 5, 6	37, 0, 2
13, 1, 1	38, 4, 5	13, 3, 4	38, 5, 5
14, 3, 2	39, 4, 4	14, 5, 6	39, 1, 1
15, 2, 2	40, 1, 0	15, 3, 4	40, 0, 2
16, 1, 0	41, 0, 0	16, 5, 5	41, 2, 2
17, 5, 4	42, 4, 4	17, 4, 6	42, 3, 2
18, 4, 4	43, 1, 0	18, 0, 0	43, 2, 2
19, 5, 6	44, 1, 0	19, 5, 6	44, 3, 4
20, 3, 4	45, 5, 5	20, 4, 6	45, 1, 3
21, 0, 0	46, 1, 0	21, 4, 6	46, 5, 6
22, 1, 2	47, 2, 2	22, 4, 6	47, 3, 3
23, 2, 2	48, 1, 1	23, 0, 0	48, 4, 3
24, 0, 0	49, 2, 1	24, 0, 1	49, 1, 0
25, 4, 4	50, 2, 2	25, 5, 6	50, 5, 6

$\alpha_1 \dots \alpha_K$ and the parameters $\beta_1, \beta_2 \dots \beta_r$ are estimated via maximum likelihood (ML) estimation. Statistical inference, such as hypothesis testing for comparing the hazard rates of the experimental therapy and control groups, is based on the ML estimates of $\beta_1, \beta_2 \dots \beta_r$.

The documentation for SAS Version 8 (SAS Institute Inc, Cary, NC) provides an example that illustrates how to apply such an analysis with SAS statistical software.⁹ If the DBPCFC (or the SBFC) is administered at interim time points between the baseline and study-end visits, yielding longitudinal data, then the discrete-time survival analysis can be incorporated into the framework of generalized estimating equations.¹⁰ Although it is possible to adapt statistical software for this situation, this step is not pursued in this article.

Sample size and statistical power

Because the discrete-time survival analysis is the discrete analogue of the Cox proportional hazards regression model, sample size for the former can be approximated via a sample size formula based on the latter. One such formula is provided by Schoenfeld.¹¹ This sample size formula requires that the effect size be expressed in terms of the hazard ratio, which is the ratio of the hazard functions for the experimental and control groups. Typical values that researchers use for effect sizes

TABLE III. Descriptive statistics and within-group based on the change-from-baseline scores

	Median (Q1, Q3)	Sign test	Wilcoxon signed-rank test
Control	0 (0, 1)	$P = .84$	$P = .70$
Experimental	1 (0, 2)	$P = .0013$	$P < .0001$

TABLE IV. Results of the discrete-time survival analysis

	β (ML estimate)	SE	P value
Baseline score	-1.109	0.124	<.0001
Therapy group	-1.157	0.263	<.0001

in sample size calculations range from a hazard ratio of 1.5 (small effect size) to 3.0 (moderate effect size). In addition to the effect size, the researcher needs to account for the anticipated censoring rate (averaged across the 2 treatment arms) at the study-end food challenge. For example, if a researcher anticipates that the experimental therapy group will have about 20% censoring and the control group 0% censoring, this is an average of 10% censoring in the study.

Table I lists sample sizes (per treatment arm) for various hazard ratios (1.5, 2.0, 2.5, 3.0) and censoring rates (0%, 10%, 20%, 30%) for a 2-sided, .05 significance level test with 90% statistical power. For a small-to-moderate effect size of a 2.0 hazard ratio with an anticipated 10% overall censoring rate, the researcher should randomize 49 patients to each of the 2 treatment arms.

RESULTS

A data set for a clinical trial for comparing an experimental therapy to a control therapy in patients with peanut allergy was simulated. Six doses of peanut flour were administered during a baseline DBPCFC and study-end DBPCFC: 1, 4, 16, 64, 256, and 1024 mg. A patient was assigned the score of 0, 1, 2, 3, 4, or 5, depending on at which dose the patient displayed clinical signs of food allergy, and was assigned the score of 6 if there were no clinical signs at the highest dose. On the basis of sample size as discussed, 50 patients were randomized to each of the experimental therapy and control groups. The simulated data set appears in Table II.

Within-group descriptive statistics, based on the change-from-baseline scores, appear in Table III. The control therapy group did not display a statistically significant improvement over baseline (sample median = 0; $P = .70$), whereas the experimental therapy group did (sample median = 1; $P < .0001$). Thus, the experimental therapy group improved by a median factor of 4 (4^1) the peanut flour dose that it could tolerate. The Wilcoxon rank sum test indicated a statistically significant difference between the control therapy and experimental

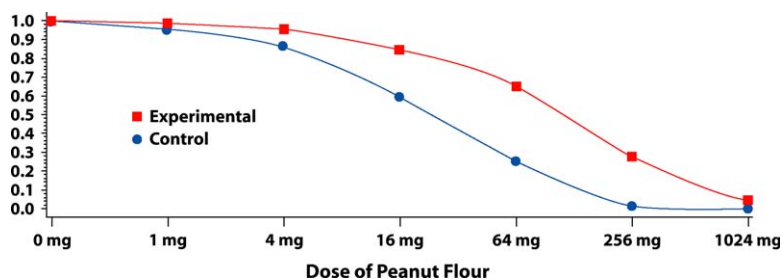


FIG 1. Smoothed Kaplan-Meier survival curves for the simulated data set provided in Table II. The curves have been adjusted for a baseline response of 64 mg. The experimental therapy is superior to the control therapy in improving the tolerance to peanut flour in patients with peanut allergy.

therapy groups with respect to the change-from-baseline scores ($P = .004$).

A discrete-time survival analysis was applied via SAS statistical software.⁹ The logistic regression model with the complementary log-log link function included regressors for the baseline score (x_1) and for therapy group ($x_2 = 0$ for control therapy and 1 for experimental therapy). The parameter estimates, SEs, and P values appear in Table IV. The baseline covariate score and therapy group both yielded highly significant results in the regression analysis ($P < .0001$). The negative sign for the ML estimate of the β coefficient for the baseline score indicates that the hazard decreases (survival increases) as the baseline score increases. Likewise, the negative sign for the ML estimate of the β coefficient for the therapy group indicates that the hazard decreases (survival increases) for the experimental therapy group. A plot of the smoothed Kaplan-Meier survival curves, adjusted for a baseline score of 3 (64 mg), appears for the control and experimental therapy groups in Fig 1. The survival curves vary according to the baseline score, although the relative difference between the experimental and control survival curves is similar for each baseline score. For illustrative purposes, a baseline score of 3 was chosen. In this simulated trial, it is clear that the experimental therapy is superior to the control therapy in improving the tolerance to peanut flour in patients with peanut allergy.

DISCUSSION

The DBPCFD is a rigorous tool that has become popular for evaluating adverse reactions to foods. The standard use of the DBPCFD has been to document food allergies for individual patients, but it recently has been gaining acceptance as a procedure for investigating the effectiveness of therapies to prevent and/or minimize food-induced anaphylaxis.

We describe statistical design issues important to consider when embarking on a comparative trial that invokes the DBPCFC (or the SBFC). We also demonstrate that statistical analysis can be performed by using standard statistical software. Although simple nonparametric tests for within-group and between-group comparisons are easy

to apply, a discrete-time survival analysis provides the best approach because of its handling of right-censored observations and its flexibility in accounting for important independent variables (regressors) and longitudinal data. This situation is comparable to what a researcher encounters when comparing 2 treatment groups with respect to a normally distributed outcome (or approximately normally distributed). A 2-sample t test may suffice, but an analysis of covariance (ANCOVA) that accounts for covariates likely will provide much more statistical power. This advantage was observed in our simulated example because the P value based on the Wilcoxon rank sum test ($P = .004$) was of a lesser magnitude than that based on the discrete-time survival analysis ($P < .0001$).

Leung et al⁶ applied ANOVA with multiple comparisons and trend tests to the change-from-baseline log-transformed dose levels in their study. They could have applied ANCOVA to account for other factors in the study, but did not do so probably because they achieved a high level of statistical significance via the ANOVA trend test ($P < .001$). They did not mention whether they encountered right-censored observations, and if so, how they handled right-censored observations. The discrete-time regression model that we describe is better suited to accommodate right-censored observations than ANCOVA, so the former should be applied if an appreciable rate of right-censoring, say $>10\%$, is expected.

The simulated example we present included only regressors for baseline score and therapy group, and it did not include interim food challenges, but the discrete-time survival analysis can handle such features in a straightforward manner.¹⁰

REFERENCES

1. Bock SA, Sampson HA, Atkins FM, Zeiger RS, Lehrer S, Sachs M, et al. Double-blind, placebo-controlled food challenge (DBPCFC) as an office procedure: a manual. *J Allergy Clin Immunol* 1988;82:986-97.
2. Noe D, Bartemucci L, Mariani N, Cantari D. Practical aspects of preparation of foods for double-blind, placebo-controlled food challenge. *Allergy* 1998;53(suppl 46):75-7.
3. Vlieg-Boerstra BJ, Bijleveld CM, van der Heide S, Beusekamp BJ, Wolt-Plompen SAA, Kukler J, et al. Development and validation of challenge materials for double-blind, placebo-controlled food challenge materials in children. *J Allergy Clin Immunol* 2004;113:341-6.

4. Sampson HA, Ho DG. Relationship between food-specific IgE concentrations and the risk of positive food challenges in children and adolescents. *J Allergy Clin Immunol* 1997;100:444-51.
5. Rodriguez J, Crespo JF, Burks W, Rivas-Plata C, Fernandez-Anaya S, Vives R, et al. Randomized, double-blind, crossover challenge study in 53 subjects reporting adverse reactions to melon (*Cucumis melo*). *J Allergy Clin Immunol* 2000;106:968-72.
6. Leung DYM, Sampson HA, Yunginger JW, Burks AW Jr, Schneider LC, Wortel CH, et al, for the TNX-901 Peanut Allergy Study Group. Effect of anti-IgE therapy in patients with peanut allergy. *N Engl J Med* 2003;348:986-93.
7. Hollander M, Wolfe DA. *Nonparametric statistical methods*, 2nd ed. New York: John Wiley & Sons; 1999. p. 35-140.
8. Agresti A. *Categorical data analysis*, 2nd ed. Hoboken (NJ): John Wiley & Sons; 2002. p. 282-4.
9. SAS Institute. Example 39.11: complementary log-log model for interval-censored survival times. In: *SAS OnlineDoc*, Version 8. Cary (NC): SAS Institute; 1999.
10. Heagerty PJ, Zeger SL. Multivariate continuation ratio models: connections and caveats. *Biometrics* 2000;56:719-32.
11. Schoenfeld DA. Sample-size formula for the proportional hazards regression model. *Biometrics* 1983;39:499-503.