

Statistical model for assessing the proportion of subjects with subjective sensitisations in adverse reactions to foods

The outcome from a Double-Blind Placebo-Controlled Food Challenge is often of a subjective nature and cannot be measured directly. Reactions to placebo challenges are frequently observed, implying that some of the responses in the study are in fact 'false responses'. In order to adjust for these false responses, previous studies have used various methods, including removing subjects from the analysis who reacted to the placebo. Simply ignoring the false responses can lead to misleading estimates for the true proportion of sensitised individuals. This paper outlines two models which can account for these false responses. In the single challenge study, a simple model is developed which enables the estimation of the rate of false responses in the study, as well as the true proportion of sensitised subjects. This model is very easy to apply in practice. For a repeated challenge study, a more complicated model is developed which again enables the estimation of the rate of false responses and the true proportion of sensitised subjects.

**David Briggs¹, Laurence Aspinal¹,
Andrea Dickens¹,
Carsten Bindslev-Jensen²**

¹Unilever Research, Colworth House, Sharnbrook, Bedfordshire, UK, ²Department of Dermatology, Allergy Centre, Odense University Hospital, Denmark

Key words: statistical significance; subject sensitisation; false responses.

David Briggs
Unilever Research
Colworth House
Sharnbrook
Bedfordshire MK44 1LQ
UK

The diagnosis of an adverse reaction to a food must rely on the outcome of Double-Blind, Placebo-Controlled Food Challenge (DBPCFC) (1). At present, however, no statistical model has been developed and validated to ensure proper validation of the outcome of the challenges.

Especially in patients with only subjective symptoms upon challenge, a model needs to be developed, since in this group of patients who often present vague and/or controversial complaints, reactions to placebo challenges are expected to be frequent. One method to overcome this possible risk is repeated challenges, which has been used with success previously (2).

Furthermore, it is critical to decide the fate of the patients reacting to placebo, since these patients will definitely alter the validity of the challenges. In some centres, patients reacting to placebo are excluded from analysis but, as previously demonstrated (3), this will hamper validation of the procedure.

The key challenge when analysing subjective sensitisation is to develop a model which can account for false positive responses. Studies of this type are susceptible to a substantial proportion of false positives, although estimates of the false positive rate are difficult to obtain. An estimate of the proportion of "truly" sensitive subjects can be heavily biased by the false positive

responses. Two models are outlined below which enable us to estimate the proportion of "truly" sensitive subjects, and the proportion of false positive responses, for two different study types.

One active, one placebo challenge

An underpinning assumption of this model is that there will be no false negatives, i.e. if a subject is truly sensitive to a treatment, then they will give a positive response. The model is based upon the following parameterisation:

p = the probability that a subject is truly sensitive to the active. It is assumed that nobody is truly sensitive to the placebo.

Q = the proportion of subjects who have a tendency to give false responses.

z = the probability that a subject who has a tendency to give false responses actually reports that they are sensitive, regardless of the true response.

The responses can be classified in three ways:

- 1) Qz The proportion of subjects who have a tendency to give false responses and who report sensitivity.

- 2) $Q(1-z)$ The proportion of subjects who have a tendency to give false responses, but do not report sensitivity.
- 3) $(1-Q)$ The proportion of subjects who do not have a tendency to give false responses.

The model shows very neatly that the proportion of subjects who are truly *not* sensitive to the active is the number of subjects who respond with a negative on the active, divided by the number of subjects who respond with a negative on the placebo. This result is extremely easy to apply in practice and leads to the estimate of the parameter p .

If our data is presented as in Table 1, then the estimate of the false positive rate (Qz) is

$$\hat{Qz} = \frac{c+d}{a+b+c+d}$$

and the estimate of the proportion of truly sensitive subjects (p) is

$$\hat{p} = \frac{b-c}{a+b}$$

Significance test for p

We might also wish to test whether or not our estimate of p is statistically significantly different from zero. If $p=0$, Table 1 has the property of being symmetrical about the main diagonal. Any test for symmetry is therefore essentially testing the null hypothesis that $p=0$. A simple binomial comparison of the top right hand cell count with the bottom left hand cell count tests the null hypothesis that $p=0$ (5). The resulting test statistic is approximately a Standard Normal Deviate and can be compared at the 95% level of the Standard Normal Distribution.

Model implications

Fig. 1 illustrates how the statistical significance of p changes as the sample size and proportion of false positives change. The height of the surface shows the P value for the statistical significance of p . The other axes show the sample size and the proportion of false positives.

Table 1. Table showing the structure of the data across the four categories.

Placebo	Active		
	Negative	Positive	
Negative	a	b	$a+b$
Positive	c	d	$c+d$
	$a+c$	$b+d$	$a+b+c+d$

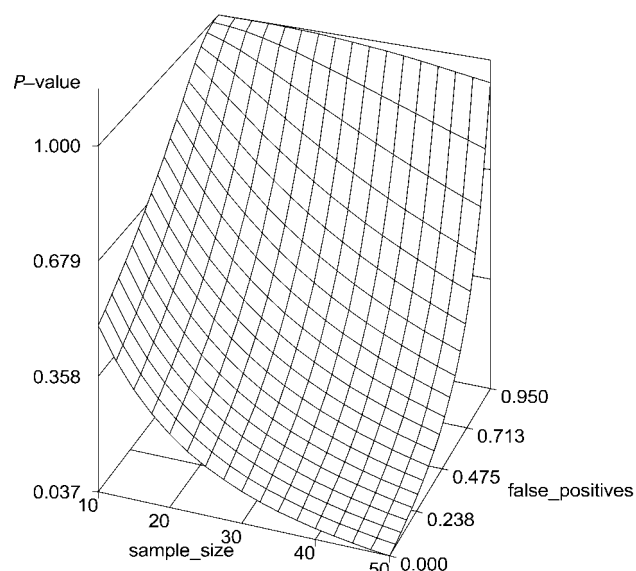


Figure 1. Graph showing the relationship between the P , sample size, and proportion of false positives (for $P=0.1$, $z=1$).

- For a given proportion of false positives, as the sample size increases the P value decreases, as given by standard power calculations.
- For a given sample size, as the proportion of false positives increases, the P value increases. As the proportion of false positives increases, there will be fewer off-diagonal entries in the data table. The data table will become more symmetrical, thus reducing the ability of our test to show that p is significantly different from zero.

An analysis which simply ignores the positive responses on the placebo results in overestimating p . This is because some of the positive results on the active will be “false positives”. These false positives are accounted for in the model described here.

FREQUENCY BLOCK CHART

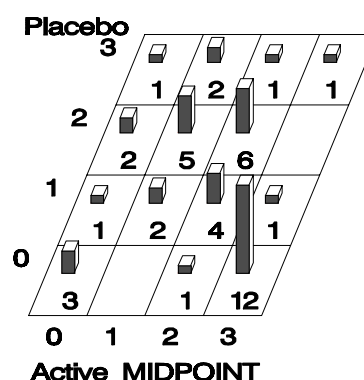


Figure 2. Block chart showing the number of positive responses on the placebo and the number on the active.

Extension of the model to three active, three placebo challenges

In this situation, each subject in the study is exposed to 3 repeats on the active and 3 repeats on the placebo. In order to account for this, the previous model has been extended to analyse data of the type displayed in Fig. 2 (data from (4), where 42 adults with subjective symptoms only were challenged repeatedly. A similar parameterisation has been used, with the inclusion of an extra parameter π . This parameter allows for subjects not always giving positive responses, even if they are truly sensitive to the active.

The following parameters are defined in our model:
 Q = the proportion of subjects who have a tendency to give a false positive response.

z = the probability that a subject who has a tendency to give false positive responses, actually reports sensitivity, regardless of the actual response.

p = the true proportion of subjects who are sensitive to the active.

π = the probability that a subject who is truly sensitive to the active actually experiences sensitivity.

In Fig. 2, the number of positives on the active treatment is on the horizontal axis, the number of positives on the placebo is on the vertical axis, and the height of the block represents the number of subjects in each of the response categories. This data is an example of the type of data that might be collected in a repeated challenge sensitivity study.

In this model there is no simple algebraic solution to estimate the true proportion of sensitive subjects in the population. The method that we suggest using is based upon maximising the likelihood of a multinomial distribution. This results in maximum likelihood estimates and confidence intervals for each of the parameters of interest. The method is coded in the statistical package SAS, and can easily be implemented by anybody who has access to this software package.

After running the model on the above data, the parameter estimates and 95% confidence intervals are as shown in Table 2.

From these maximum likelihood estimates we can see that:

Table 2. Parameter estimates from the repeated challenge data.

Parameter	Estimate	Confidence interval
Q	0.73	0.54, 0.90
z	0.50	0.40, 0.59
p	0.30	0.15, 0.55
π	1.00	0.76, 1.00

- the estimate of the proportion of subjects sensitive to the active is 0.3. This is significantly different from zero at the 95% confidence level.
- the estimate of the proportion of “false reactors” is 0.73.
- for those subjects who are sensitive to the active, the estimate of the probability that they will also experience sensitivity is 1. This implies (in this example) that a subject will always experience sensitivity if they are truly sensitive.

The probability that a “false reactor” will actually report sensitivity regardless of the actual response is 0.5. This implies that 50% of the time a “false reactor” will give a false positive response.

Extensions/refinements to the model

The model currently assumes that there is no time effect present – it is assumed that responses do not increase over time. It is possible that as a subject goes through a study, they could become more likely to give a positive response, and it would be beneficial to estimate this effect. A more complicated model could potentially include this effect and estimate its impact. In the above repeated challenge model, this would equate to π and z being functions of time.

Conclusions

A model for statistical analysis of DBPCFC is presented. The model enables the estimation of the statistical significance of the parameters, both in very clear cases (e.g. patients reacting with severe systemic reactions to active challenge) and in more complicated cases, where patients claim to suffer from more controversial and subjective symptoms.

References

1. BRUINJEEL-KOOMEN C, ORTOLANI C, AAS K, et al. Adverse reactions to food. Position Paper. *Allergy* 1995;**50**: 623–635.
2. HANSEN TK, BINDSLEV-JENSEN C. Allergy to codfish in adults. *Allergy* 1992;**47**: 610–617.
3. BINDSLEV-JENSEN C. Standardisation of double-blind, placebo-controlled food challenges. *Allergy* 2001;**56** (suppl 67): 73–75.
4. BINDSLEV-JENSEN C, MADSEN F. Investigations on adults with suspected food intolerance. In preparation.
5. ARMITAGE P, BERRY G. Statistical methods in medical research. Oxford: Blackwell Scientific, 1987.