LETTER TO THE EDITOR

Analysis of clinical data with breached blindness

by Shein-Chung Chow and Jun Shao, *Statistics in Medicine* 2004; **23**:1185–1193

From: *Harri Hemilä*
Department of Public Health
POB 41, University of Helsinki
Helsinki, FIN-00014, Finland

In their recent paper, Chow and Shao [1] proposed a method for analysing clinical data with breached blindness, claiming that bias caused by knowledge of the identity of the treatment can seriously distort statistical inference on therapeutic effects. Thus, they argued that adjustments to statistical analyses should be made when the integrity of blinding is doubtful. However, if blinding was a fundamentally essential requirement of the validity of studies, it would have dramatic effects on medical research since no meaningful studies could be carried out on cigarette smoking, rare side-effects of drugs, surgery, etc. The validity of Chow and Shao's argument is thus highly important.

First, identification of treatment by subjective observation should not be considered merely a nuisance, because in many cases the unambiguous purpose of physicians is to reduce the subjective symptoms of patients. At the individual level, such effects can be investigated using the '$N = 1$' type of trial in which breaching of blindness often is an important explicit outcome [2, 3].

Second, there is no valid evidence indicating that the so-called placebo effect is large and omnipresent. A recent meta-analysis of trials comparing placebo and no-treatment groups found no placebo effect in studies that measured binary outcomes [4, 5]. In studies measuring continuous outcomes, only those that measured pain found evidence of a placebo effect, but it was quite small. Chow and Shao disregarded these negative empirical findings in arguing that in general any knowledge of treatment may seriously distort statistical inference on the treatment effect.

Chow and Shao briefly discussed two old trials as examples of unreliable results caused by breaching of blindness. The trial by Brownell and Stunkard focused on reducing weight in obese women using an appetite suppressant [6]. Since dosage was adjusted according to reports of side-effects, many study participants could obviously infer their treatment correctly [6]. Chow and Shao speculated that the difference between the trial groups in loss of weight might have been caused by the breached blindness, i.e. by the placebo effect [1]. However, to evaluate the effect of believing, Brownell and Stunkard compared patients who believed they were taking the drug with those who believed they were taking the placebo. These two groups did not differ, which is inconsistent with the placebo effect explanation. Furthermore, pooling the results of eight trials comparing a placebo group to a no-treatment group found no evidence that placebo would affect obesity [5]. These direct empirical comparisons refute Chow and Shao's indirect reasoning that the placebo effect might explain the findings of the Brownell and Stunkard trial [6].

The other trial Chow and Shao cited focused on the effects of vitamin C on the common cold. The Karlowski *et al.* trial lasted for 9 months and used 4 treatment arms [7], not 2 as

stated by Chow and Shao [1]. Each participant received 2 kinds of capsule: prophylactic (each day over the trial) and therapeutic (5 days during a cold). Ascorbic acid (3 g/day) was used in the vitamin C capsules and lactose in the placebo capsules, a different combination being administered to each arm so that $3+3$ g/day was the largest dose. Since lactose is sweet whereas ascorbic acid is acidic, some participants apparently inferred their treatment by taste. After the Karlowski trial was concluded, the participants were asked using a questionnaire which capsules they thought they had been administered. There was strong bias in favour of correct answers in the case of prophylactic capsules ($P < 10^{-6}$), but not in the case of therapeutic capsules ($P = 0.3$). After this puzzling finding, Karlowski carried out a subgroup analysis by dividing participants into those who remained 'blinded' (guessed incorrectly) and those who became 'unblinded' (guessed correctly) during the trial. In this analysis, all the benefit of vitamin C was restricted to the 'unblinded' participants, whereas no differences were observed in the 'blinded' participants. Thus, Karlowski concluded that 'the effects demonstrated might be explained equally well by a break in the double blind' [7]. Because the trial was initiated as double-blind, this was such a spectacular conclusion that the Karlowski trial has been cited as an example of the placebo effect in action by numerous clinical trialists including Chow and Shao [1] and the CONSORT group [8].

However, the data reported are inconsistent with Karlowski's 'placebo explanation.' Some 11 participants answered the type of therapeutic capsule correctly and some 56 the prophylactic capsule [9]. Thus, assuming that the results are explained by the placebo effect, we would expect the prophylactic capsules to be substantially more effective than the therapeutic capsules. However, the prophylactic capsules were 34 per cent less effective than the therapeutic capsules in all study participants (reduction in cold duration by -0.48 and -0.73 days per episode) and 75 per cent less effective in 'unblinded' participants (-0.75 and -3.0 days) [9]. The greater benefit from the therapeutic capsules is inconsistent with Karlowski's 'placebo explanation,' since there is no valid evidence that any participants inferred their therapeutic capsules ($P = 0.3$) and, at most, only some 8 per cent did [9].

Furthermore, Karlowski et al. did not describe how they divided their participants into the 'blinded' and 'unblinded' subgroups. The two subgroups were treated as if they were complementary, yet their sum does not equal all participants. In total, there were 105 common cold episodes (42 per cent of all colds) missing from Karlowski's subgroup analysis [9]. The maximum effect of vitamin C on common cold duration in the 'missing group' was even greater (-1.4 days; 6 vs 0 g/day [9]) than the maximum effect across the whole study population (-1.22 days). Karlowski et al. did not mention the exclusion of the 105 episodes from their subgroup analysis, nor did they offer a rationalization for the greater than average benefit among the participants who were neither 'blinded' nor 'unblinded.' There are a number of further logical inconsistencies with Karlowski's 'placebo explanation' [9]. The re-analysis of the Karlowski et al. trial was commented on but no valid counter-arguments were presented [10, 11].

There has been a long-lasting popular belief that vitamin C is beneficial against the common cold and subjective observation may thus affect the inference of treatment. In fact, the 'inference from subjective observation' concept was directly supported by the parallel report of the Karlowski trial [12]. Among participants who had not tasted their prophylactic placebo capsules, those who had colds during the trial tended to suspect they were being given placebo (15 of 18 participants), whereas those who did not have colds tended to suspect vitamin C (6 of 8 participants) (Fisher-$P = 0.02$). Evidently, a similar inference applies to the duration

of colds, but this was not considered by Karlowski. In placebo-controlled trials vitamin C has reduced the duration and severity of colds up to 20–50 per cent [13, 14] so that some people may correctly infer from subjective observation whether they received vitamin C or placebo. In such a case it is inappropriate to 'adjust' the results for 'breached blindness' because correct identification may be caused by the physiological effects.

Finally, to understand the potentially false conclusions generated by the Chow and Shao method better, let us consider a semi-realistic thought experiment. Let us assume we examine the effect of penicillin on community-acquired pneumonia in 40 patients who are randomized into two groups of identical size. Let half the pneumonia cases be caused by pneumococcus and the other half by mycoplasma so that these are evenly distributed among the treatment groups. Let us also assume that placebo-treated pneumonia (both pneumococcal and mycoplasmal) lasts 10 days (SD 2 days), and penicillin shortens the duration of pneumococcal pneumonia to 4 days (SD 2 days) but has no effect on mycoplasmal pneumonia. Finally, let us assume that all patients guess the type of treatment (for simplicity there are no 'don't know' answers), and all patients with pneumococcal pneumonia who were administered penicillin correctly inferred their treatment from the rapid and dramatic benefit, but all others guessed correctly and incorrectly half and half.

With these semi-realistic assumptions we can calculate the interaction test according to the Chow and Shao method, which leads to $F(1\,\mathrm{df}, 36\,\mathrm{df}) = 8.57$, corresponding to $P = 0.006$. Accordingly, in this thought experiment the Chow and Shao method leads to the conclusion that 'we cannot conclude that the treatment effect [of penicillin on pneumonia] is significant' [1, p. 1190] which is not a reasonable conclusion since we know from bacteriology and decades of clinical experience that penicillin is an effective therapy for pneumococcal pneumonia which covers a large proportion of community-acquired pneumonia [15]. Furthermore, in the presence of interaction, Chow and Shao proposed subgroup analyses of participants with correct and incorrect guesses. In this example, penicillin-treated patients with incorrect guesses consist of 5 patients with mycoplasmal pneumonia, whereas placebo-treated patients with incorrect guesses consist of 5 patients with mycoplasmal and 5 with pneumococcal pneumonia. Thus, while the purpose of randomization is to produce balanced groups, subgroup analysis based on guessing the treatment can lead to grossly unbalanced groups when the treatment does have physiological effects. Obviously, if the proposed method of analysis leads to a false conclusion in such well-established therapy, it is not generally useful. The lack of validity of Chow and Shao's argument for the pivotal role of blinding in medical studies is highly significant for epidemiologists, toxicologists, surgeons, and many others.

REFERENCES

1. Chow SC, Shao J. Analysis of clinical data with breached blindness. *Statistics in Medicine* 2004; **23**: 1185–1193 [comments in: 2005; **24**:819–822].
2. Guyatt G, Sackett D, Taylor DW, Chong J, Roberts R, Pugsley S. Determining optimal therapy: randomized trials in individual patients. *New England Journal of Medicine* 1986; **314**:889–892.
3. Bali L, Callaway E. Vitamin C and migraine: a case report [letter]. *New England Journal of Medicine* 1978; **299**:364.
4. Hrobjartsson A, Gøtzsche PC. Is the placebo powerless? An analysis of clinical trials comparing placebo with no treatment. *New England Journal of Medicine* 2001; **344**:1594–1602.
5. Hrobjartsson A, Gøtzsche PC. Placebo interventions for all clinical conditions. *Cochrane Database of Systematic Reviews* 2004; (3):CD003974.
6. Brownell KD, Stunkard AJ. The double-blind in danger: untoward consequences of informed consent. *American Journal of Psychiatry* 1982; **139**:1487–1489.

7. Karlowski TR, Chalmers TC, Frenkel LD, Kapikian AZ, Lewis TL, Lynch JM. Ascorbic acid for the common cold: a prophylactic and therapeutic trial. *Journal of the American Medical Association* 1975; **231**:1038–1042.
8. Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, Pitkin R, Rennie D, Schulz KF, Simel D, Stroup DF. Improving the quality of reporting of randomized controlled trials: the CONSORT statement. *Journal of the American Medical Association* 1996; **276**:637–639.
9. Hemilä H. Vitamin C, the placebo effect, and the common cold: a case study of how preconceptions influence the analysis of results. *Journal of Clinical Epidemiology* 1996; **49**:1079–1084.
10. Chalmers TC. To the preceding article by H. Hemilä. *Journal of Clinical Epidemiology* 1996; **49**:1085.
11. Hemilä H. To the dissent by Thomas Chalmers. *Journal of Clinical Epidemiology* 1996; **49**:1087.
12. Lewis TL, Karlowski TR, Kapikian AZ, Lynch JM, Shaffer GW, George DA, Chalmers TC. A controlled clinical trial of ascorbic acid for the common cold. *Annals of the New York Academy of Sciences* 1975; **258**: 505–512.
13. Hemilä H. Vitamin C, respiratory infections and the immune system. *Trends in Immunology* 2003; **24**: 579–580.
14. Douglas RM, Hemilä H, D'Souza R, Chalker EB, Treacy B. Vitamin C for preventing and treating the common cold. *Cochrane Database of Systematic Reviews* 2004; (4):CD000980.
15. Donowitz GR, Mandell GL. Acute pneumonia. In *Principles and Practice of Infectious Diseases* (5th edn), Mandell GL, Bennett JE, Dolin R (eds). Churchill Livingstone: Philadelphia, PA, 2000; 717–743.

AUTHORS' REPLY

Analysis of clinincal data with breached blindness, *Statistics in Medicine*
2004; **23**:1185–1193

Our responses to the comments in Hemila's letter to the editor are in the order of the appearance of the comments in the letter.

First, we believe that we did not claim that identification of treatment by subjective observation should be considered as a nuisance.

Second, placebo effect often exists in a clinical trial where subjective assessment is used. Several negative empirical findings cannot prove that breaching blindness does not distort statistical inference.

For the weight reducing example by Brownnell and Stunkard, we are very surprised to see that Hemila did not carefully read our paper before writing his letter. In Section 2 of our paper, we conclude with a high significance level that the blindness is not preserved in this example. But this does not mean the difference between the trial groups in loss of weight is caused by the breached blindness. A further study was given in Section 4 of our paper. When the guessing factor is constructed using guessing correctly or not, the conclusion in our paper is that the treatment effect is significant, regardless of whether the effect of guessing factor is significant or not. By the way, we think that comparing patients who believed they were taking the drug with those who believed they were taking the placebo, as described by Hemila, is not a correct way of assessing the treatment effect.

For the other example from Karlowski *et al.* (prophylactic), all we did in our paper is to find out that the blindness is not preserved. We did not do any further analysis because the original data were not available to us. However, we still receive criticisms from Hemila, because he thinks that showing that the blindness is not preserved is the same as claiming a placebo effect. How can we be inconsistent (or consistent) with Karlowski's analysis if we actually did not do any analysis?