

ENDGAMES

STATISTICAL QUESTION

What is a superiority trial?

Philip Sedgwick *reader in medical statistics and medical education*

Centre for Medical and Healthcare Education, St George's, University of London, London, UK

Researchers investigated the efficacy of agnus castus fruit (*Vitex agnus castus* L extract Ze 440) in relieving symptoms of the premenstrual syndrome. A randomised, double blind, placebo controlled, parallel group, superiority trial was performed. The intervention was agnus castus (dry extract tablets), with one tablet daily for three consecutive cycles.¹

Participants were 170 women with diagnosed premenstrual syndrome recruited from general medicine community clinics, with a mean age of 36 years, mean cycle length 28 days, and a mean duration of menses 4.5 days. In total 86 women were randomised to intervention and 84 to placebo. The primary outcome variable was the combined scores of self assessment of six symptoms: irritability, mood alteration, anger, headache, other menstrual symptoms (including bloating), and breast fullness. Participants used a visual analogue scale to rate each symptom. The scale, validated for the assessment of the premenstrual syndrome, was of length 10 cm, ranging from 0 (no symptoms) to 10 (unbearable). The primary outcome was recorded at baseline and at the end of the third menstrual cycle, and the change in total score was recorded for each woman.

The mean reduction in the primary outcome was significantly greater in the intervention group (mean reduction 128.5 mm) than in the placebo group (78.1 mm) (mean difference 50.5 mm (95% confidence interval 23.5 mm to 77.5 mm; $P < 0.001$)). The researchers concluded that dry extract of agnus castus fruit was an effective treatment for the relief of symptoms of the premenstrual syndrome.

Which of the following statements, if any, are true?

- The statistical null hypothesis states that intervention is superior to placebo in reducing symptoms of the premenstrual syndrome.
- It can be concluded that placebo is ineffective at reducing symptoms of the premenstrual syndrome.
- The research hypothesis states that intervention is superior to placebo at reducing symptoms of the premenstrual syndrome.

Answer

Statement *c* is true, while *a* and *b* are false.

The purpose of the above trial was to investigate the efficacy of agnus castus fruit in relieving symptoms of the premenstrual syndrome. The trial was designed as a superiority trial, with the aim of establishing whether the intervention was superior to placebo in effectiveness or whether placebo was more effective than intervention. The primary outcome was the combined scores of the women's self assessment of six symptoms of the premenstrual syndrome. As a parallel group trial, women received the treatment allocated at baseline for the entire duration of the study period. Treatment groups were compared in the mean change in the primary outcome variable from baseline to the end of the third cycle. Statistical hypothesis testing was used to test for superiority, with a null and alternative hypothesis as described in a previous question.²

The null hypothesis incorporates the traditional starting position of equipoise: it states that in the population of women from where the sample was obtained there was no difference between the intervention and placebo groups in the mean change in the primary outcome variable from baseline to the end of the third menstrual cycle (*a* is false). The objective was to establish whether the trial results supported the null hypothesis or provided evidence of a difference between the intervention and placebo, as specified by the alternative hypothesis.

The alternative hypothesis states that in the population of women from where the sample was obtained there was a difference between the intervention and placebo groups in the mean change in the primary outcome from baseline to end of the third menstrual cycle. No direction is specified, and the alternative hypothesis is termed two sided: the mean change in the primary outcome in the intervention group could be less or greater than that in the placebo group. The researchers stated that superiority of either group over the other would have been shown if the mean difference between the groups in the primary outcome was at least 12 mm. This difference, called the smallest effect of clinical interest, was predetermined. The sample size was calculated from the smallest effect of clinical interest so as to ensure that the trial had adequate statistical power to detect this difference, if it existed.^{3 4}

The mean change in the primary outcome was 128.5 mm in the intervention group and 78.1 mm in the placebo group. The

difference of 50.5 mm (95% confidence interval 23.5 to 77.5) was significant ($P < 0.001$); therefore the null hypothesis was rejected in favour of the alternative, and it was concluded that there was a difference between treatment groups. The inference was that intervention was superior to placebo in reducing symptoms of the premenstrual syndrome. However, it cannot be concluded that placebo was ineffective at reducing symptoms of the premenstrual syndrome (b is false). Placebo is often thought of as equivalent to no treatment. Nevertheless, the placebo group showed a considerable mean reduction in self assessed symptoms of the premenstrual syndrome. The response by the placebo group is a complex one that would, in part, be described by the placebo effect.⁵

It is important to distinguish between the research hypothesis and the statistical hypotheses. The researchers will have stated their research hypothesis before the study started, making predictions about the study results. The research hypothesis was that intervention was superior to placebo (c is true); this would have been based on anecdotal evidence or perhaps on a pilot or exploratory study that did not have the statistical power to show a clinically important difference between treatment groups. It was this prior expectation that agnus castus would be beneficial in relieving symptoms of the premenstrual syndrome that provided the basis for undertaking the trial, it being essential to obtain evidence that the intervention was effective by undertaking a placebo controlled trial.

As well as the research hypothesis being stated before the trial began, so also were the statistical hypotheses. However, in contrast to the research hypothesis, the statistical null hypothesis starts at the position of equipoise, with the alternative not

specifying a direction. The aim of the trial was to establish whether the data provided sufficient evidence to reject the null hypothesis in favour of the alternative and therefore to support the research hypothesis. Despite any anecdotal evidence or previous data, the alternative hypothesis is two sided. This is because it is possible that the results of the trial may not be as expected and that, for example, placebo may be superior to intervention in effectiveness.

Traditionally, clinical trials have been performed as superiority trials—that is, they have tried to establish whether a new drug or therapy was more effective than the standard treatment or placebo. With the continuing development of medical drugs, therapies, and devices, it is becoming increasingly difficult to develop new ones that can be shown to be sufficiently more effective than the standard regimen. This has led to the development of other types of trial, including non-inferiority and equivalence trials. Such trials will be discussed in future questions.

Competing interests: None declared.

- 1 Schellenberg R for the study group. Treatment for the premenstrual syndrome with agnus castus fruit extract: prospective, randomised, placebo controlled study. *BMJ* 2001;322:134-7.
- 2 Sedgwick P. Statistical hypothesis testing. *BMJ* 2010;340:c2059.
- 3 Sedgwick P. Sample size calculations I. *BMJ* 2010;340:c3104.
- 4 Sedgwick P. Sample size calculations II. *BMJ* 2010;340:c3299.
- 5 Sedgwick P. The placebo effect. *BMJ* 2011;343:d7665.

Cite this as: *BMJ* 2013;347:f5420

© BMJ Publishing Group Ltd 2013