Reviewer's report

Title: The effect of gold kiwifruit consumed with an iron fortified breakfast cereal meal on iron status in women with low iron stores: A 16 week randomised controlled intervention study

Version: 1 Date: 14 December 2009

Reviewer: Andrew R Willan

Reviewer's report:

1. The authors should state the “sidedness” (i.e. one- or two-sided) for the type I error rate used to determine sample size and for the proposed analyses.
2. The difference used to determine sample size should be the smallest clinically important one, not one expected or observed in other studies.
3. Random allocation should be done using a computer generated list and stratified by serum ferritin level and age, not pair-matched and by the flip of a coin. Using a coin-flip is far too amenable to manipulation, and whether allocation is manipulated or not, it is the perception that matters. There are on-line randomization services which are ideal for small research groups doing one or two trials.
4. Data entry into a spreadsheet does not provide the data integrity that is required. A proper database with controlled data entry should be used, e.g. MS-ACCESS.
5. The investigators propose using statistical tests to compare baseline variables. This is not appropriate. One of the primary interpretations of a p-value is that it is the probability that the observed difference is due to chance. Since the patients are randomized to treatment groups, any difference at baseline must be due to chance. A small difference in a very large trial might be statistically significant, but have no confounding effect on the treatment comparisons with respect to the outcomes. On the other hand, a large difference in a small trial might not be statistically significant, although it could confound the treatment comparisons.
6. Comparing treatment arms with respect to the outcome variables using the change from baseline is inefficient (i.e. low power) compared to using the baseline measure as a covariate. In fact, unless the correlation between the baseline and the follow-up measure is greater than 0.5, comparing the differences is less powerful than comparing the follow-up measure alone. The use of ANCOVA provides the most power use of the baseline measure.

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Acceptable
**Statistical review:** Yes, and I have assessed the statistics in my report.

**Declaration of competing interests:**

I declare that I have no competing interests.