Reviewer's report

Title: Strong association between non alcoholic fatty liver disease (NAFLD) and low 25(OH) vitamin D levels in an adult population with normal serum liver enzymes.

Version: 3 Date: 13 April 2011

Reviewer: William Barlow

Reviewer's report:

Statistical review of “Strong association between non alcoholic fatty liver disease (NAFLD) and low 25(OH) vitamin D levels in an adult population with normal serum liver enzymes”

Major compulsory issues:

The paper was scientifically interesting, but I found the presentation of results to be somewhat confusing. I believe the goal is to suggest a possible path between vitamin D levels and subsequent NAFLD. Of course, in observational data one can only show an association, rather than cause. Nonetheless, the most appropriate analysis is logistic regression on the primary outcome NAFLD using predictors such as vitamin-D levels, age, etc. This is what is actually done in Table 3, showing a strong association of vitamin-D with NAFLD after adjustment for other variables. However, those results are not described in the abstract nor correctly referenced in the methods section. Instead, the abstract presents the mean vitamin D levels by NAFLD status (OK but not what is being tested in the analysis), rather than the odds ratio for a single unit change in vitamin-D. The methods section described this analysis incorrectly as multiple linear regression, rather than logistic regression on a dichotomous outcome. Similarly, there was an analysis of NAFLD scoring using linear regression (Table 2) assuming it was continuous but it seems to only have three ordered values. Ordinal regression would be better here. Finally, there are some analyses relating clinical predictors to each other. Table 4 has FLI as the outcome, but it is just estimated from a logistic regression model based on triglycerides, ggt, and waist circumference so seems removed from the main question unless FLI is commonly used in this area as a measure. Similarly, Table 5 showed the lower and upper quartiles of vitamin D by other factors. I actually prefer to see all 4 quartile values with a trend test, but recognize that the approach used here is commonly done. Finally, Table 6 is back to a logistic regression on NAFLD. It would be better to put all the primary analyses (NAFLD as dichotomous outcome) together and then make it clear the other analyses are secondary. Table 1 is needed first as a description of the population. It was somewhat unclear when log-values of continuous variables were used, versus untransformed, values.

I doubt that any scientific conclusions would change after redoing the statistical section. The methods need to be in agreement with what was actually done.
There needs to be a primary question which I believe is the association of vitamin-D and other variables with the dichotomous outcome NAFLD. The other analyses can be added afterward if they add something that is not known.

Minor compulsory issues: Please spell out NAFLD in the abstract the first time. There are a few typographical errors that need to be fixed.

**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.