Reviewer's report

Title: Hemoglobin levels and anemia evaluation during pregnancy in the highlands of Tibet: a hospital-based study

Version: 2 Date: 15 June 2009

Reviewer: Giorgio Bedogni

Reviewer's report:

General comment

I was asked to review this paper as BMC statistical referee. I read version 2 of the manuscript together with the comments made by three reviewers on December 2008.

Major Compulsory Revisions

None

Minor Essential Revisions

P2 L28-30 Does the relationship which “has not been clearly established” refers to pregnant women or to the whole Tibetan population? This is not clear to me based on the present wording of the sentence.

P2 L36 Please, use “association” instead of “influence” as the study design does not allow to model cause-effect relationships.

P3 58 Although there is no doubt that anemia has a negative impact on the immune system, I wonder whether the Authors intend to say that this is the major mechanism by which anemia increases morbidity and mortality.

P3 L70 see P2 L28-30

P4 L98-ff & P5 119-ff Please, explain better how sampling was performed. On which days of the week did you perform sampling? I suppose that this was a kind of systematic sampling, wasn’t it? How many subjects did you sample in a day? Was this number fixed or variable? From your reply to one reviewer, I understand that “random” sampling was performed over 1 year on 1200 pregnant women attending your hospital. Please, report this very important number in the manuscript.

P2 L28-30 You selected 396 women (of whom 16 were excluded from the study). However, you planned to measure erythrocyte constants in only 48% of the study sample (190). How was this sample selected? Moreover, 84 women dropped from the study so that erythrocyte constants were evaluated on 106 women (27% of planned sample size). Please, provide evidence that these 106 women are representative of the whole 396. Otherwise, you cannot be confident that the
conclusions about erythrocyte constants can be generalized to the whole sample.

P6 L154 How the cut-off of 1800 Yuan was chosen? Please, provide a conversion to USA $ among parentheses.

P6 156 (and elsewhere) I suppose that you are referring to the fit of the covariates in the multiple linear and multivariable logistic regression models (not in simple linear and univariable logistic regression models). Is this right?

P6 L162 Please, write “agreement” instead of “consistency”.

P7 L170 and elsewhere I strongly suggest to report data as mean and standard deviation as this allows a better assessment of the variability inside your study sample.

P7 L177 Please, add standard deviations to figure 1. Because there are just 50 non-Tibetans, their “curve” is clearly less defined than those of Tibetans and the comparison must be due with caution. (Provided that there is enough power you could perform a comparison using an appropriate regression model which take into account the “uncertainty” of the curve of non-Tibetans.)

P8 L204-ff Did you take into account the multiple-comparison issue when comparing the three separate kappa statistics?

P8 L212 “For each characteristic...”: this phrase is not clear and I suggest to remove it.

P9 L233 Because you do not provide a reference for your hypothesis, one supposes that this is the first time this hypothesis is done. Is there any data to support your hypothesis (in Tibet or elsewhere)? Biological plausibility is very important in explaining your findings especially because of the potential selection bias.

P9 L249 You measured mean corpuscular volume (MCV) and red cell distribution width (RDW) in a subsample of the women (see P4 L98-ff & P5 119-ff) and this may partly explain the oddity of your finding. Please, comment on this point.

P10 L260-ff How did you collect symptoms? With a standardized form? How did you operationally defined “hypodynamia”?

P6 L168 17% of study women were non-Tibetans. From the data you report for whole Tibet, I learnt that 96% of Tibet residents are of Tibetan origin. Are there more non-Tibetans in Lhasa than in other places of Tibet? Or does your hospital has an high referral rate for non-Tibetans? Please address this point, because ethnicity is an important covariate in your regression models.

Table 1 I strongly suggest to report descriptive data as mean and standard deviations (see P7 L170).

Table 1 Report the operational definition of “Tibetans” (i.e. Han, Hui, Menba... etc.) also under Statistical analysis.
Table 1 Looking at the Table, gestational age appear to have been modeled as 3 trimesters. 1,2 and 3? I wonder how a log-transformation can improve the fit of a three-level equally-spaced variable. I am also not sure that this variable should be modeled as continuous because it takes just 3 values. Please, explain.

Table 1 Parity was model as continuous. However, this is a 3-level variable and the last value (i.e. 2+) covers “more possibilities” than previous values (i.e. 0,1). Have you tested parity as categorical variable before using it as continuous?

Discretionary Revisions

None

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

**Quality of written English:** Acceptable

**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**

I declare that I have no competing interests