Reviewer's report

Title: A cross sectional study of Schistosoma japonicum prevalence in 50 villages of Samar Province, the Philippines.

Version: 1 Date: 16 December 2005

Reviewer: Robert Spear

Reviewer's report:

While the authors set out to "describe the village-to-village variation" in infection parameters in their study sites, their analysis excludes all but one village-level variable (irrigation status), this despite them having collected other village environmental and agricultural data. For example, the village selection criteria used by the authors provide 6 village variables seemingly relevant to a village level analysis of risk factors, including fractions of irrigated agricultural area, absolute areas of rice cultivation, presence or absence of developed canals, pumps and dams, etc. A more fundamental question, however, concerns what the authors presume to be the influence of irrigation (MAJ). Is it on the exposure of villagers to the parasite or on the snail? Certainly, a missing element is a discussion of the intermediate snail host including the extent and nature of its habitat. It seems reasonable to expect that there has been progress in the understanding of the snail and its habitat since the extensive work of Pesigan and collaborators in the Philippines in the 1950s and this would certainly seem to have a place in a paper discussing village level risk factors. Further, some discussion of the agriculture in these villages would have been welcome (MAJ). The implication is that exposure occurs in rice culture, but clothes washing and other activities most probably are also involved and account for the gender differences in infection discussed.

With respect to the infection data itself, the authors have taken great pains to collect a first-class infection-related dataset, and are right to emphasize the ordinal nature of epg data, generally. That said, however, the continuous epg data that they have collected may be alone among datasets for its quality (owing to the multiple stool samples, multiple smears) and the rigor in its collection (owing to the careful analysis of the characteristics of participants vs. those that refused). One is curious as to the distributions of these data, prior to aggregation into ordinal intensity classes (MAJ). Of particular interest is the nature of the right tail of these data, which harbor most of the parasites in the population. What are the dominant attributes of these individuals, reported even qualitatively? How does this size of this group compare with an oft-mentioned 80/20 population? Moreover, the intensity classes selected seem unstandardized in the literature -- are these conventional cutoffs (MIN)?

The authors use a clever, indirect way to take account of the many false negatives from Kato-Katz – they fit a random effects, ordered logistic regression model and then get a posterior probability estimate for each subject, for each level of infection, based on their covariates and observed status. They allow for random effects at both the individual and village-level. This way, they can borrow information across the whole group, as well as the individual data, when estimating the probability of being in each infection category. However, they should:

1. State random effects model explicitly, including assumptions (such the latent random effects come from a normal distribution) (MAJ)?
2. Justify the use of this Bayesian Hierarchical model, since there are no informative priors on the parameters anyway. Why not just a straightforward random effects approach (multi-level model) and then use the empirical Bayes estimates for individual and village random effects (MAJ)?
3. Take out the word TRUE everywhere. It only represents the truth if 1) the model is right (which, of course it isn’t) and 2) the parameters are known exactly (they’re not, they’re estimated) (MIN).
4. Take out the part on post-hoc adjusting for misclassification (MIN). Why assume that here – can't one assume the eggs are measured without error? The model already accounts for the fact that the true mean egg count (for an infinite stool sample) is not known. Overall, the authors present a good case for the importance of village as a risk factor for the intensity of schistosomiasis infection, but give little insight into what village-level factors might be responsible for this outcome. The unclear effect of the single irrigation variable used in the analysis is not surprising given that the irrigated status implies >20% of the village hectares are irrigated whereas non-irrigated implies <10%, that is, potentially a quite modest difference. One would hope that in their 12 month follow-up more environmental variables will be collected to allow a more thorough exploration of the determinants of village-level risk. The individual risk factors discussed seem reasonable and generally unremarkable although the lack of any water contact data makes a good part of the discussion necessarily speculative.

What next?: Unable to decide on acceptance or rejection until the authors have responded to the major compulsory revisions

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

Quality of written English: Acceptable

Statistical review: Yes

Declaration of competing interests:

I declare that I have no competing interests.