Reviewer’s report

Title: Numbers of Mutations To Different Types of Colorectal Cancer

Version: 1 Date: 26 January 2005

Reviewer: Steven Frank

Reviewer’s report:

This paper uses the multistage theory of cancer progression from epidemiology to analyze data on the age of colon cancer onset. The data provide a comparison between different pathways of colon cancer progression, MSI- versus MSI+, and between different genotypes of mismatch repair. These are very interesting data.

The particular method of analysis follows a new approach developed by the authors in a prior paper. With their method, they estimate parameters of cancer progression, with emphasis on the number of stages in progression. The number of stages is roughly the slope of a log-log plot between age-specific incidence and age.

The paper is clear and reasonably concise. This is an acceptable contribution to BMC Cancer.

I am not certain about formal policies for making data available. In this case, I think the raw data on which this study was based should be made available if it is not already accessible on the internet from its prior summary publication (reference 9). Such data are valuable, and the summary analysis in the paper does not provide a full description of the information collected. If the raw data are already accessible, then the paper should provide a URL for direct access.

I have some comments about the analysis. My comments follow from my own recent paper on a related topic (PNAS 102:1071-1075, 2005). The authors certainly should not be compelled to alter their paper in response to the following comments--it is reasonable that different people will choose to analyze the same problem in different ways. What follows is my opinion, which may not be shared by others.

I found the derivation of the Bayesian method of analysis in the earlier paper interesting. However, I feel that such mathematical and statistical refinements are really of interest only to the few of us we care about such details. With regard to data analysis, I wonder whether the old method of plotting the observations on log-log graph paper and using a ruler to estimate slope remains the most reasonable approach for goals of this paper.

The mathematical representation of the multistage model must be considered only a very rough qualitative description of process--valuable but not to be overanalyzed. There are many factors that influence age of onset that are not included in the model, and so claiming the estimated value of k is anything but a very rough guide seems misplaced.

What does seem reasonable is to compare estimates of k between different genotypes or between different pathways of progression, as the authors have done. My own preferred way to do this is to analyze the ratio of the incidence curves in the comparison. This has the advantage that many of the unmeasured factors approximately cancel out as long as one is willing to assume that those factors affect the two classes in the same way. Thus, one only analyzes relative incidence, and no direct estimate is provided for the absolute number of steps.
Last comment: the relative number of steps in FAP versus sporadic colon cancer claimed by Ashley and repeated by Knudson is doubtful. I reanalyzed those data in the PNAS paper mentioned above.

**What next?:** Accept after discretionary revisions

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

**Quality of written English:** Acceptable

**Statistical review:** No

**Declaration of competing interests:**

I have no competing interests.