Reviewer's report

**Title:** Tumor location and patient characteristics of colon and rectal adenocarcinomas associating survival and with TNM classes

**Version:** 1  **Date:** 17 June 2010  
**Reviewer:** Owen Franklin Dent

**Reviewer's report:**

**Major Compulsory Revisions**

1. This study is poorly conceptualized, poorly prepared and poorly written and the findings are not novel.

2. The title of the paper is obscure.

3. Regarding the first four lines of the abstract: (a) CRC tends strongly to be a disease of old people and old people tend to die because of what demographers call "normal mortality" – so the reason for the association between age and poor survival in CRC is not unknown, it is quite obvious, (b) the second sentence is unrelated to both the first and the third, (c) the third sentence should indicate the research question and state the aim of the study but does not, rather it says simply that the study provides data.

4. In the first sentence of the introduction it is tautological to say that mortality has declined because survival has improved.

5. Page 3 introduces several disparate ideas but does not develop a clear theme leading towards a statement of the general research question and the aim of the study. Also, as in the abstract, it ignores the fact that CRC is a disease of old people who experience normal mortality.

6. The paragraph at the top of page 4 says briefly what was done but does not say why. There is no clear, explicit statement of the aim of the study.

7. In the second line on page 4 you say that you examined data collected over four years (presumably 2002 to 2006?) but about two-thirds of the way down page 4 you say that you retrieved cancer registry data from 1958 to 2006. What was the point of retrieving data from 1958 to 2001?

8. The description of the T, N and M categories at the top of page 5 does not precisely how they were coded in the analyses in Table 3. Without knowing the coding it is not entirely clear how the odds ratios should be interpreted.

9. At the end of the methods section it remains unclear precisely which group of patients you studied and from what time period. For example, if TNM staging was used then this implies postoperative staging which implies that all patients had a resection whereas the passage beginning in line 5 of the second paragraph of
page 4 implies that patients were included even if their cancer was inoperable. Furthermore, the first line of the results refers to patients diagnosed since 1990; does this include all diagnoses, whether a resection was performed or not, and why 1990 rather than 1958 or 2002?

10. The variables socioeconomic status, medical region and first degree family history of CRC should be defined explicitly.

11. The method followed in logistic regression modelling should be explained.

12. In the second line of the second paragraph on page 5 "sex" would be a more appropriate word than "gender".

13. The first paragraph of the results describes a survival analysis but this is not mentioned in the methods section and the reason for it is not specified. The survival outcome variable should be specified along with censoring and the method of survival analysis. Information should be given about the follow-up protocol.

14. Table 1 should be cited in the first paragraph of the results but is not. The heading to Table 1 refers to cause-specific survival but the method by which CRC-specific survival was analysed is not described. If the method was Kaplan-Meier with deaths due to causes other than CRC censored then this would be incorrect. Because a considerable proportion of deaths among patients with CRC are due to causes other than CRC the survival analysis should take account of competing risks. See Putter H. et al. Tutorial in biostatistics: competing risks and multi-state models. Statistics in Medicine 2007;26:2389-2430 and reference 17 in that paper.

15. In Tables 1 and 2 "ascendens" should be "ascending" and "decendens" should be "descending" and "sigmodeum" should be "sigmoid".

16. It is not clear what the point of the analysis in Table 1 is. Why focus specifically on left versus right side as the outcome of interest (note that site is not adjusted in respect of the other variables) whereas earlier in the paper you emphasized stage as the variable which might explain the association between older age and poorer survival? Why was stage not adjusted? If it is because you did not have stage data for these patients this should be stated explicitly in the methods section. If stage was not available for all of these patients then why didn’t you restrict the analysis to the period for which it was available? The whole question of what this analysis is about and why it was done has not been explained.

17. How do the 17,487 patients in Table 2 relate to the 51,172 patients in paragraph 1 of the results? Why did you not do all analyses on the one group of patients?

18. At the foot of page 5 you say that Table 2 shows the key variables but surely stage is a key variable, yet it is not included whereas family history, which is peripheral to the very large majority of sporadic CRC, is included.
19. The footnote to Table 3 does not say that side was adjusted in the models. Is this an oversight, or if it was not adjusted, why not when all other variables were? You say that sex and family history had no effect in the models, so why have you adjusted for them? Why not reduce the models to adjust for only those variables which had effects, which is the conventional approach?

20. The discussion is confused and confusing. It tries to grapple with too many issues at once without developing a clear progression of ideas. Perhaps the most telling sentence is where you say "The finding of more aggressive tumors among young patients was opposite to the survival data which were better in young patients." yet nothing you have done works towards explaining why this is so. As commented earlier, you have entirely overlooked the issue of normal mortality. Furthermore the sentence "Thus the overall conclusion is that poorer survival of old patients in colon cancer is unlikely to be related to delayed diagnosis." is unjustifiable because none of your analyses concerned diagnostic delay. However the very last sentence of the paper is entirely correct and this is what you should have been taking into account throughout the study. If your data source does not allow you to do this then it is questionable whether any of your analyses are worthwhile.

**Level of interest:** An article of insufficient interest to warrant publication in a scientific/medical journal

**Quality of written English:** Needs some language corrections before being published

**Statistical review:** Yes, and I have assessed the statistics in my report.

**Declaration of competing interests:**

I declare that I have no competing interests.