Reviewer's report

Title: Hospital factors and patient characteristics in the treatment of colorectal cancer: a population based study.

Version: 2 Date: 16 April 2012

Reviewer: Dianne O'Connell

Reviewer's report:

Major Compulsory Revisions

Abstract
1. The abstract should include a section on methods.

Results and Discussion
2. Para 3. Table 1 refers to the use of preoperative radiotherapy (RT) for patients with rectal cancer who underwent an elective intervention. However the use of preoperative RT is not indicated for all rectal cancer cases (only stage II/III) according to the Background. Therefore the proportion of patients receiving preoperative RT may be underestimated, and some of the differences in patients who did and did not receive this may be confounded by cancer stage. If stage, other than whether or not metastases are present, cannot be obtained in these data, then this is a major limitation of the study.

Methods

Study population
3. A main concern is the use of the Hospital Discharge Record system for identifying colorectal cancer cases. Is this truly population-based? Does this system really include all hospitals in the region? Are there private hospitals that are not included?
4. Were the colorectal cancer cases histopathologically confirmed?
5. Information from the Piedmont Cancer Registry was used in a validation study. Why was it not linked to the clinical information in the HDR for this study?

Statistical analysis
6. Para 1. The usual indicator of quality of care is 30-day mortality (ie the proportion who died up to 30 days after surgery). This requires information on vital status possibly after discharge which would be obtained by active follow-up of patients or passive follow-up via linkage with death records. The in-hospital mortality analysed in this study is a time-dependent variable with censoring when the patient is discharged alive. Therefore it should be analysed using statistical methods for time-to event data (often referred to as “survival analysis”).

Validation
7. Para 2. The positive predictive value of the identification of colorectal cancer cases in the HDR compared with the Piedmont Cancer Registry was 89.9%. This
means that 10% of cases included in the analysis of the HDR were not found in the Cancer Registry. Who are these cases? Also what is the false negative rate? That is, how many cases in the Cancer Registry were not identified in the HDR?

8. Para 3. The “high resolution sample” used to validate the information in the HDR is restricted in age to 50-69 years. Why did this occur? The cohort included in the wider study is all ages with 49% aged over 70. Given that the data may be less accurate for older patients this casts some doubt on the validity of the HDR data for this cohort.

Minor Essential Revisions

Abstract

9. Para 1 of results. N=24,187 patients were included in the analysis, not 25,302. Also there was no association between being less educated and post-operative mortality.

10. Para 2 of results. This is a repeat of the last sentence in the previous paragraph.

Abstract and Background (para 4)

11. The authors indicate in the first sentence that the focus of the study is on “non-clinical” factors that can lead to disparities in the management and outcome of care. However the next sentence indicates that “social, clinical and hospital determinants” were investigated. This is inconsistent.

Results and Discussion

12. Para 2. The numbers of patients with colon and rectal cancers in Figure 1 add to 22,431 not 22,289. Did some patients have tumours in both their colon and rectum? If so, how were these classified in Table 3 where 15,256 colon and 7,033 rectal cancer patients (which add to 22,289) are reported on?

13. Para 3, last sentence. The odds ratio does not vary with distance, rather the odds of receiving neoadjuvant RT vary. Was a test for trend conducted? This would be required to draw a conclusion of decreasing odds with increasing distance.

14. Para 4. Is the difference in odds of AP resection for patients with and without an emergency admission statistically significant? The upper 95% confidence limit is 1.01. What is the corresponding p-value?

15. Para 4, last sentence. This is poorly expressed. It should say something like: “There was no independent association between disease stage, Charlson index or year of admission and having an AP resection for rectal cancer.”

16. Para 5. As odds ratios are calculated, not relative risks, it is incorrect to say that mortality is “two to three times higher ...” based on the magnitude of the odds ratio. It is the odds of dying that are two to three times higher.

17. Para 5. For education level, the only odds ratio that is significantly different from 1 is for the “unknown” group, not the less educated groups.

18. Para 6, second sentence. This is speculation but you could (and should)
examine this in the data. Did older patients have more comorbidities (a cutpoint of 1 or more may be too insensitive) or live further from RT facilities? Also, this could be confounded by older patients being diagnosed with later stage disease.

19. Para 6, last sentence. Did older patients have “a higher burden of comorbidities”? This should be examined in the data, not speculated on.

20. Para 7. Again, the likelihood of having preoperative RT may be confounded by cancer stage.

21. Para 9. The association between education level and mortality was confined to the group with “unknown” education level (Table 3).

22. Para 10. Do not know what “comparing row percentages” means in this context. Please describe the proportions that are being compared. It is unclear what is meant by “non-clinical” factors. Are the authors distinguishing between individual patient characteristics and health system variables?

Conclusion

23. Para 2. The discussion of the limitations of the study should be moved to the discussion, not included in the conclusions.

24. Para 3. This conclusion cannot be drawn from this study. The conclusions should refer to what was found in the study about the management of people with colorectal cancer.

Methods

25. The methods section should be moved to between the Introduction and Results sections.

Patient characteristics

26. Para 1. More information is required about the patient records in the HDR system. What is actually meant by the “data set [is] de-identified at source”? How are different patient episodes linked? Do individual patients have an encrypted or some other form of identification so multiple records can be linked?

27. Para 5. What is the basis for the cutpoints for annual hospital caseloads? Are they based on the distribution in the data (e.g. tertiles) or are they based on categories used in the literature previously?

Statistical analysis

28. Para 1. The first sentence describing the outcomes of interest should be moved to a new section “Outcomes” following the section on patient characteristics. Also “rates” of preoperative RT etc were not analysed. It is the proportion of patients receiving RT etc that are analysed and reported. Suggested wording:

“... the proportions of rectal cancer patients who received preoperative RT or abdominoperineal (AP) resection, and the proportion who died post-operatively in hospital.”

Tables

29. Overall p-values for each factor should be included in the tables. These will
indicate whether overall, there is a significant association between each variable and the outcome, adjusting for the other variables in the model.

30. Table 1. Hospital volume is not included in the logistic model due to its strong association with presence of an RT service. As a sensitivity analysis, the model should be refitted including hospital volume and omitting RT service and the results reported in the text.

31. Table 3. The APR-DRG risk-of-mortality score also includes age and age is included in the logistic model. Is there multicollinearity?

32. Table 3. Title should indicate that these are colorectal cancer patients who had curative surgery. Also the number of patients with colon cancer (15,256) differs from that in Figure 1 (where, as previously indicated, some patients are double counted). This requires explanation.

33. Table 4. The labels on the rows and columns are misleading. The patients are the same in the two “samples”. Should they refer to the data sources “HDR” for the rows and “Medical records” for the columns.

34. Table 4. There are inconsistencies in the numbers of patients. There were 390 rectal cancer patients but data for only 277 patients are shown for type of resection. If patients were classified as not having a resection in either data source then this category needs to be included.

35. Why are only 575 patients included in the comparison of presence of metastasis? If this is not recorded in either data source then a category for “unknown” is required.

Minor issues to be corrected (but not for publication)

Results and Discussion

36. I suggest splitting the results and discussion into two sections. The discussion would start at the current para 6 of the combined section (“In our population ....”).

37. Para 3. Suggest minor changes to wording:

“The probability of undergoing RT was also reduced in females versus males .....for patients with at least one comorbidity in the Charlson Index (OR 0.73...”

“Furthermore ....had a higher probability of receiving ...”

38. Para 6. Suggest changes to wording of first sentence:

“In our population, older people with rectal cancer were less frequently treated with preoperative RT, were more likely to undergo AP ...”

Study population

39. Para 2, first sentence. Change to:

“ ... or malignant neoplasm of the rectum or rectosigmoid junction ...”

Patient characteristics

40. Para 1. Move the first sentence to the end of the first paragraph in the section on study population. Start second sentence with “The HDR includes ...”
41. Para 2, last sentence. Suggested word change:
“Cases were classified as having concomitant obstruction ..., perforation ..., or an emergency admission (OPE) or not.”

42. Para 4. The labels for the APR-DRG classification system shown in Table 3 (low, medium, high, extreme) need to be mentioned.

43. Para 6. Suggested changing to wording of first sentence:
“For each patient, accessibility to preoperative RT was measured as the distance between their residence and the nearest city with a RT service, by car and under normal traffic conditions [43].”

44. Para 6. The second and third categories are not mutually exclusive. Should they be: same city or less than 15 minutes; 15 to less than 30 minutes; 30 minutes or more?

Ethical issues
45. What does CPO stand for?

46. I believe that the authors are arguing that this study was considered to be a clinical audit to monitor the quality of treatment, which is considered to be an essential activity rather than research and therefore was exempt from ethical review.

Abbreviations
47. Delete “surgical” from the definition of DRG.

Tables
48. Delete all vertical lines in tables (bottom of Table 1 and right-hand edge of Table 2).
49. Table 1. Label for middle category of distance to RT should be 15’ to <30’
50. Table 2. What does the second superscript in the title refer to? Also the relevant superscript should be added to the line for OPE in the body of the table.
51. Why are the numbers of rectal cancer patients in Table 1 (5437) and Table 2 without OPE (5382) different?

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

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.