Reviewer's report

Title: Completeness of Registration of HIV and Hepatitis B and C Coinfection in The Danish National Hospital Registry, 1995-2004

Version: 1 Date: 8 October 2007

Reviewer: David B. Preen

Reviewer's report:

The proposed work initially aims to examine the completeness and accuracy of recording of HIV infection in the Danish National Hospital Registry (DNHR) from 1995-2004, by comparison with data obtained for the Danish HIV Cohort Study. Additionally, the study proposes to identify the level of recording of hepatitis B and C for people with HIV in the DNHR, as well as establish temporal associations with recording of such information.

The study focuses on an interesting methodological area which, will likely be of interest to those relying on similar data sources for research purposes. Further, given the continued rapid developments in information technology and willingness of researchers to utilise such data, the results perhaps have implications for future investigations.

The manuscript was generally well structured. However, while the primary aspects of the planned research were well presented, some methodological issues possibly exist that require clarification or further consideration. Furthermore, the manuscript was a little too brief in many instances. As such, there were some sections of the paper (specifically the Methods and Discussion) which could have benefited from the provision of a greater level of detail or additional commentary. Specific comments relating to the primary areas of concern, categorised according to the journal’s revision criteria, are as follows:

Major compulsory revisions

Abstract – While generally well-structured, the abstract is somewhat short/brief. Specifically, a large portion of the methodology and results (eg, Cox regression, hazard ratios etc) are not mentioned. Further, while sensitivity results are displayed, this analysis is not mentioned in the Methods section. This is similar for the temporal associations reported in the results but not referred to in the methods section of the abstract. In addition, the 1½ line conclusion requires more detail.

Methods – One aspect that was not clear from the paper is whether the DNHR contains only primary admitting diagnosis/procedure information or whether it also includes secondary diagnoses and co-morbidities (as with many other hospital discharge registries). This requires clarification as it has major implications for the study. If only primary diagnosis is recorded, this would likely be the major underlying explanation for the low sensitivity observed in this data
set for hepatitis B and C by the study. Further, if comorbidity information is collected, how many conditions are coded in the registry data? Also, if coded is it likely that only conditions relevant (or perhaps similar) to the principal diagnosis are recorded in the medical notes? Information on these aspects of the DNHR data is essential for the study methodology. In addition, a greater level of commentary is required on this issue in the Discussion section.

Discussion – The Discussion section was too brief and perhaps a little superficial in some places. Few comparisons with previous research findings were included. Further, when comparisons were made few specific details were provided. For instance, with comparison of the current findings to previous work on acute myocardial infarction etc, some specific results from these other studies would have been useful. In addition, given the limitations, and the jurisdictional-specific focus of the current study, some discussion of potential recommendations for further/future research was required.

Minor essential revisions

Methods – While briefly discussed in the Discussion section of the paper, some information on the number and proportion of patients with HIV who do not seek medical treatment for their condition is warranted. While, I am assuming that it would be a relatively small number, some published precedent or more detailed information regarding this issue would be useful in the Methods section.

Given that many readers may not be familiar with the DNHR, a greater level of background information on this data source should be included in the Methodology. Specifically, what information is collected (eg, only primary diagnosis or also comorbidities)? Also, is the Registry routinely audited, or has any validation work ever been carried out indicating the accuracy and completeness of the data for other diagnoses?

While acknowledged by the authors, the major limitation for the study was that they were “not able to compute predictive value of a HIV diagnosis in the DNHR since [they] did not have permission from the Danish Data Protection Agency to identify HIV cases recorded only in the DNHR”. In addition to restricting analysis of positive and negative predictive value, this limitation also restricts evaluation of specificity of the data maintained within the DNHR. Consequently, it represents a major issue for the study and considerably decreases the quantity and quality of the findings that could be generated. However, given the implications of this limitation, relatively little discussion is devoted to this issue in the Discussion section of the manuscript and further commentary is warranted.

From the description presented, I am assuming that the study is comparing DNHR data to that recorded in the DHCS for only those HIV patients actually admitted to hospital, or attending a captured outpatient service, during the study period. However, this is not clearly stated in the manuscript. Some comment to this effect is warranted, as otherwise it could be perceived that comparisons were made for all patients captured in the DHCS regardless of whether they were ‘hospitalised’. This would be a completely different research question, as it would be expected that there would be an under-ascertainment simply due to those
patients not requiring secondary health care services for their condition.

In the first paragraph of the Methods section, age and race are not mentioned as terms to be included in the Cox regression analysis. However, these factors are then presented in the text of the Results section.

Results – Throughout the Results and Discussion Sections (as well as in Table 2) comments are made relating to the “risk” or “relative risk” of various outcomes. However, these estimations are derived from Cox regression analyses which provide proportional hazard estimates (ie, hazard ratios). While they may approximate relative risks, they are technically two different measures. Consequently, these instances in the text should be reworded to remove mention of relative risk.

For the comment that “of note, some patients with [a] HBV or HCV diagnosis recorded in the DNHR did not have a corresponding record in [the] DHCS”, some indication of the number or proportion of patients who fell into such a category would be useful here. Further, were sufficient data noted to allow at least a quasi-analysis of specificity or predictive value for these variables? If so, then such an inclusion would strengthen the article. If not, perhaps a brief statement to this affect would be a worthwhile addition.

Discretionary revisions
Title – The title is perhaps a little misleading as to the true focus of the study. As it currently stands, the title does not make it clear that the study will investigate hepatitis C and B in only those patients previously identified with HIV. I would suggest altering the title to improve clarity of your study focus and target population.

Background – This section, while perhaps a little brief, was generally well-structured and created a clear rationale for the study question to be investigated.

Methods – It would be worthwhile providing some additional brief description for the “other variables” contained in the DHCS. For example, was there information available on demographic, socioeconomic, clinical etc data?

Results – Table 3 is mentioned in the text of the Results section prior to Table 2. Tables should be numbered in the order in which they appear in the text.

Note that age, CD4 count and log viral load results presented in Table 1, do not really fit under the “total number of patients” column heading. In addition, units should be included for age, CD4 count and log viral load results.

Discussion – While I agree that “given the nature of HIV infection” some of the study results may be able to be generalised to other countries with similarly organised health care systems, more detail on the ability to extrapolate findings to other specific countries (eg, other European nations, the US, Canada, Australia etc) would be useful.
Also, given the ever emerging capacity to analyse large administrative health data sets, some commentary on the implications of the use of data from such registries is warranted. In addition, some discussion of previous researchers who have already attempted to investigate hepatitis C or B from such hospital registries would be an important inclusion for the Discussion section, as it is likely that such researchers have made erroneous conclusions from their research.

In-text references – There are a number of instances where factual statements are made without supporting citations to back up the assertions. Examples of such instances where supporting references are required as follows:

- Introduction, paragraph 1, sentence 2
- Introduction, paragraph 2, sentence 3

Recommendation:
Based on the review of the article, it is my opinion that the article is of sufficient quality to warranted publication, subject to the authors attending to the above-mentioned issues.

Verdict: “Accept after minor essential revisions”

Level of interest
As indicated in the above review comments, this paper while methodologically focussed, of limited clinical scope and jurisdiction specific, would still likely be of interest to investigators relying on similar data sources for research purposes. Further, given the continued willingness of researchers to utilise such administrative data sources, the results perhaps have implications for future researchers.

Verdict “An article whose findings are important to those with closely related research interests”

Quality of written English
One more minor issue was that a reasonable number of grammatical errors or formatting inconsistencies were noted throughout the manuscript. Examples of such grammatical and formatting issues are outlined below.

- Title – Should “The Danish …” be capitalised?
- Background section – On the first line of text, the abbreviations AIDS and HIV are not defined, likely as they were previously defined in the abstract. However, in contrast abbreviations such as HBV, HCV, DNHR and DHCS are re-defined in the introduction. Further, some abbreviated terms are additionally defined on multiple occasions throughout the text (eg, DNHS and DHCS defined in abstract, introduction and methods). It is recommended that the formatting structure be kept consistent for all such terms, as per the journal’s formatting guidelines
- Minor point, but in the first paragraph of the Results section two percentages are rounded to the nearest whole number (ie, 93% and 96%), while another is
expressed more precisely (ie, 98.7%). I would suggest expressing all percentages consistently, according to the journal’s formatting guidelines.

· Discussion – Reference is commonly made to “an HIV treatment centre”, whereas this should be rewritten as ‘a HIV treatment centre”. Similar for “an HBV or an HCV diagnosis”, “an HIV hospital treatment centre” etc. Also, it is grammatically correct to refer to the DNHR or the DHCS.

· Spelling/grammatical error (ie, “twese”) on third line of last paragraph in the Results section. Also, in the same paragraph, is this meant to be \( p<0.05 \) and \( p<0.01 \), as \( p=0.05 \) is not technically statistically significant?

· In table 2, the lower bound confidence interval of 0.863 should be rounded to two decimal places to be consistent with the rest of the table.

· Also for Table 2, a closed-bracket is required for the confidence interval under the “crude analysis” heading for “Centre 3”.

Verdict “Needs some (minor) language corrections before being published”

Declaration of competing interests

I declare that I have no conflicts of interest or competing interests (financial or non-financial) with either the submitted paper or contributing authors.