Author's response to reviews

Title: Hepatitis C prevalence in Denmark - a capture-recapture estimate based on multiple national registers

Authors:

Peer B Christensen (peer.christensen@dadlnet.dk)
Gordon Hay (g.hay@socsci.gla.ac.uk)
Peter Jepsen (PJ@DCE.AU.DK)
Søren A Just (d383813@dadlnet.dk)
Henrik B Krarup (h.krarup@rn.dk)
Nina Weis (ninaweis@dadlnet.dk)
Niels Obel (niels.obel@rh.regionh.dk)
Susan Cowan (sco@ssi.dk)

Version: 2 Date: 25 April 2012

Author's response to reviews: see over
Dear Editor,

We hereby forward the revised version of our manuscript

Hepatitis C prevalence in Denmark—a capture-recapture estimate based on multiple national registers Peer B Christensen, Gordon Hay, Peter Jepsen, Søren A Just, Henrik B Krarup, Nina Weis, Niels Obel and Susan Cowan

The paper has been changed according to our response to the reviewers comments as stated below. Reviewers’ comments are shown in italic followed by AUTHOR REPLY in capital letters. In addition a new language revision has been performed. We would like to thank the reviewers for their effort and many useful suggestions for improvement.

Tables are imbedded in the manuscript as required and the one figure has been redesigned. The manuscript is submitted as Word 2003 files in a final version: MS1148908436639721revision1.doc and a file with all changes visible (hcvdk_BMCinfdisrevision1trackchanges.doc). However the tables have not been modified in this file as this becomes very difficult to read. The revised tables are in the final manuscript (only table 2 changed).

We find that the manuscript has improved considerably by the revision and hope that you will find it suitable for publication.

On behalf of the authors
Sincerely yours
Peer Brehm Christensen

1. Reviewer’s report

Reviewer: ANGELOS HATZAKIS

Reviewer’s report:

1. This is an interesting application of capture-recapture method to estimate the number of HCV infected, within a country. An important assumption is that drug users is the only source of HCV infection in Denmark. This is maybe true after 1991 but before this date a substantial but unknown number of HCV infected come from health care associated infections who survive up to 2007. The authors should discuss this limitation.

AUTHOR REPLY: As mentioned in the discussion we estimate that 15-17% of the Danish HCV population is infected by other routes than drug use, primarily nosocomial infections. However this does not affect the capture-recapture estimate as no assumption of route of infection is made in the source registers. We agree that estimating the diagnostic coverage from the drug treatment register may not be valid for nosocomial infections. However as explained in the discussion a strictly nosocomial infected group (the Danish HCV look back recipients) had a diagnostic coverage of 55% in our source registers practically identical to the drug use based estimate. So we think this does not introduce a major bias in our estimate. This has been clarified in the text.

2. The analysis may violate some of the assumptions of capture-recapture method, but the authors provide independent data that support their conclusions. Finally, authors should explicitly discuss that seroprevalence surveys in the general population and risk groups is the gold standard method for assessing the number of HCV infected within one country.
AUTHOR REPLY: We agree and have added this to the text.

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

**Quality of written English:** Acceptable

**Statistical review:** No, the manuscript does not need to be seen by a statistician.

---

**2. Reviewer’s report**

**Reviewer:** Matthew Hickman

**Reviewer’s report:**

The overall objective of estimating the number of people infected with HCV in a country is important – as often the data are inconsistent and methods can be complicated (e.g. Sweeting and De-Angelis et al). However, I found the methods described in this paper confusing and in parts unconvincing.

1. Under what assumptions can the registers used in the exercise estimate the number of “diagnosed cases not present in the registers”? Is this due to under-reporting to the registers themselves, or are there are other potential sources of diagnosis that are not covered by the registers but may be represented by the overlap? If solely due to under-reporting are there any other studies to corroborate the estimate generated by capture-recapture (CRC).

**AUTHOR REPLY:** There are no other national sources for hepatitis C diagnosis available and we believe underreporting to the four source registers is the most important explanation. This was confirmed in the Danish HCV lookback cohort mentioned in the discussion where only 55% of known HCV cases were identified in our source registers. It was expected that the reporting to hospital and clinical registers may be incomplete, but it may be more difficult for the reader to accept that diagnosed cases were not present in the laboratory register as a hepatitis C diagnosis is based on a positive HCVRNA test. There were two sources of deficits in the laboratory register used in this study: First we only received computer stored results and computer registers were introduced in the last part of the nineties whereas HCV testing was introduced in 1991-1992. Only two laboratories had records of positive results prior to 1996. To adjust for this time dependent variable we stratified for year of first registration in registers. Second: As stated in the methods Section 4, laboratories covering three of the 16 counties with 15% of the Danish population did not participate in the DANVIR register. So it is plausible that not all HCV cases were identified in the laboratory register. As the three other registers covered the whole country we think this was adjusted for in the statistical analysis. We have added this to the discussion. For information to the reviewer: That the laboratory register is indeed complete in a closed population was demonstrated in a survey of an outbreak of acute hepatitis B on the island of Funen, here 97% of all cases were present in the laboratory register (and the 3% missing had been diagnosed during admissions outside the island).

2. Can the authors re-assure readers that CRC is not simply estimating the number of diagnoses excluded i.e. 2380 persons were excluded and an estimated 2231 persons estimated to be diagnosed but missed.

**AUTHOR REPLY:** The estimate was based on individuals alive at the day of register extraction (31.12.2007) and 88% of the 2380 excluded were dead. In our initial analysis we did not exclude dead persons and this resulted in a much larger hidden population would be identified (3500-7000 cases dependent on interaction terms) so the hidden and excluded population was not identical. Inclusion of the dead was rejected early in the analysis as the assumption of independence between
registers was questionable for this group: once a patient had died he could no longer enter the other registers.

3. Under what justification can the authors assume that non-tested drug users had the same prevalence of chronic HCV infection as those tested? Do the non-tested have the same age/gender/duration of injecting/treatment and prison exposure as those tested. Is testing a random event unrelated to likelihood of infection? This assumption is too large to be unsupported by other evidence.

AUTHOR REPLY: Those tested were significantly older (median age 40 years versus 32 among non tested) and had a higher proportion of women (52% versus 40%). Likewise those infected were older (43/38years) but no significant gender difference was found. We do not have access to duration of infection and prison experience in the drug treatment register. The observed age and gender difference would suggest a shorter time of drug use and thus a lower HCV prevalence among non-tested drug users. However several studies have shown that drug users become infected very quickly once they start to inject (in Denmark half become infected within one year). In contrast the vast majority are not tested until they enter a drug treatment facility (on average four years after start of drug use. At the treatment centre all drug users are offered testing for HCV, regardless of duration of drug use. The assumption of the same test rate among infected and uninfected was confirmed by the independent 2007 survey performed at treatment facilities in Funen where the same prevalence of HCV was found as in the tested proportion of the register. We also compared the estimated HCV prevalence in the drug treatment register to a national survey of drug related deaths and found comparable results (35% versus 40%) and this sample had a comparable age and gender distribution to the drug treatment register. So we believe that the difference in prevalence between tested and not tested drug uses was a minor bias. We have added age and gender difference to results and discussion.

4. On page 9 last paragraph there are as series of calculations and assumptions that need to be more fully explained – as I didn’t quite understand them fully. Under what assumptions is 54% of total number of patients infected with HCV diagnosed. This is critical to the whole estimation exercise but is not clear.

AUTHOR REPLY: we have modified this section and tried to make it clearer.

5. The sensitivity analysis also is unclear.

AUTHOR REPLY: We have changed this section and given it a less ambitious title “Excluded populations”

6. The discussion provides some justification for some of the assumptions which is needed in the methods and results.

AUTHOR REPLY: we have moved the look-back data to the results, but find that most of the rest belongs to the discussion section, as the results we refer to were not part of the current study.

7. Information on model fits and overlaps for CRC should be given fully rather than a simple final estimate.

AUTHOR REPLY: The estimates summarized in table two are derived from summing various lower level stratified estimates (i.e. the Male/North estimate is the sum of the three male -40, 40 to 49 and 50+ estimates. We can provide the model fits, including which model fitted and goodness of fit statistics for each of the 60 lower stratified cells however we feel that such a large table may be
excessive. We can also provide the 60 overlap patterns (a 60x16 cell table) but again we feel that such a large table would be excessive and of limited interest to most readers of the journal.

8. It is not clear to me how the method can generate a reliable estimate of the number of people infected with HCV – especially as the data sources do not seem to take account of ex and current injectors.

AUTHOR REPLY: We agree that it would be useful to have information on current and ex-injectors. However as discussed below (reviewer 3) this was not available from any of the registers. We have referred to the British estimate, estimated the size of the ex-drug using population infected with HCV and discussed the consequences of a lower detection rate (and a lower prevalence of hepatitis C) in former drug users. Overall we find our estimate is comparable with the British results previously published.

Level of interest: An article of limited interest
Quality of written English: Needs some language corrections before being published:
Statistical review: No, the manuscript does not need to be seen by a statistician.

3. Reviewer’s report
Reviewer: Ruth King
Reviewer’s report:
Major compulsory revisions
1. Methods, Data sources: DANVIR:
a. why do 4 of the 18 Danish laboratories not contribute to this database? It is also stated that this database is estimated to cover 85% of the Danish population – what is the “missing” 15% - is this regional?

AUTHOR REPLY: The formation of DANVIR was based on voluntary co-operation and four laboratories from three different counties did not respond to our invitation. They represent three different levels of urbanization: rural, suburban and city and served 15% of the population. As the other source registers did include these counties we believe that this was adjusted for in the statistical analysis.

b. “we included all subjects who had a positive HCV-RNA test not followed by a negative HCV-RNA” – In the results section (sensitivity analysis) it states that “We excluded 749 patients from the laboratory register as they were initially HCVRNA positive but later became negative. However, 564 of these were registered with chronic hepatitis C in one or more of the other registers.” Please clarify if these individuals are only removed from the DANVIR register but may still be listed on the other registers (also see below point).

AUTHOR REPLY: The 564 present in other registers were kept in the analysis (see comments to next point).

2. Methods, Data sources above Drug Treatment Register “Except the laboratory register it was not possible to exclude patients who cleared the infection from the source registers” – what is the “clearance” rate of chronic HCV? Could this lead to a potential source of bias (such potential issues are discussed in the Discussion) – it might be useful to add something to the current point of different classification of chronic HCV between the sources, this related issue.
AUTHOR REPLY: The clearance rate for chronic hepatitis C is less than 1% per year. Patients could also become cured by treatment without being eliminated from the registers, but Denmark has a low treatment rate (estimated 2% of the infected population), and as a cure would be defined by HCV RNA testing they would be excluded from the laboratory register. We think that applying an exclusion criterion that is only applicable on one data source is probably more likely to lead to bias. Although this could lead to overestimation of the hidden population by decreasing the overlap, we think this is a minor bias. This issue is mentioned in the discussion.

3. Methods, two step procedure – Step 1
a. “The selection of the best fitting models was based on the Schwartz and Akaike information criterion” – more statistical detail is needed here, including the set of possible log-linear models that are fitted to the data, the definition of “best fitting model” (was this solely on a single information criterion?) and why two information criterion appear to be used (and what the procedure is in the case of discrepancies between the information criterion used for different datasets).

AUTHOR REPLY: In total 113 different models including all possible two-way and three-way interactions were fitted to the overlap data. We primarily used the Akaike information criteria to select the ‘best’ fitting model however when this model produced an estimate that differed markedly from the weighted estimate (averaged across all fitted model using the Schwartz criterion as a weight) then the Schwartz criterion was used to obtain the best fitting model. If there was still a discrepancy or the choice between the Akaike and Schwartz criterion was not clear then the model that produced an estimate closest to the known:estimated ratio found in other strata was selected. This has been specified in the methods section.

b. “stratified by gender, three age groups, five geographical regions and time period” – what are the age groups and geographical regions, and what is the rationale for these stratifications. In addition how do these stratifications compare with the 85% of the Danish population assumed to be covered by the DANVIR source? Note that taking this number of stratifications leads to relatively “sparse” contingency tables (a total of 60 strata leading to an average of only 116 individuals per contingency table with 16 cells per table).

AUTHOR REPLY: The age groups are listed in the table as are the geographical regions. We feel that age and gender and geographical region are relevant covariates when examining hepatitis C and this is confirmed comparing the studies of drug users and patients infected with blood transfusion in Denmark where difference in gender age and geographical distribution was seen. We accept that these stratifications lead to relatively sparse contingency tables however we have tried other levels of stratification and found difficulty fitting models to those data which perhaps suggests geographical or other heterogeneity.

c. “Confidence intervals for the total estimate were derived from boot-strap analysis of 1000 samples” – please clarify whether the bootstrap algorithm used fits only the best fitting model.

AUTHOR REPLY: The bootstrap algorithm only used the best fitting model within the strata (i.e. did not take into account model uncertainty within individual strata).

4. Methods, two step procedure – Step 2
a. “Assuming the same prevalence among the non-tested we calculated the total number of hepatitis C infected in the drug treatment register” – is this assumption reasonable? If individuals are showing some symptoms of HCV are they more likely to be tested for HCV than others who do not
(in other words being tested is NOT independent of having HCV). If this is the case, the proportion of individuals with HCV who are tested will be higher than the proportion of individuals with HCV who are not tested.

AUTHOR REPLY: We accept the skepticism but would like to point out that most patients with hepatitis C are asymptomatic and the vast majority is diagnosed by screening of current and former drug users attending drug treatment facilities. Here the anti-HCV test is offered to all clients at first visit and at a regular basis hereafter. As discussed above, those infected were probably (a little) more likely to be tested. However when we compared the prevalence among the routinely tested population in the drug treatment register with a systematic survey performed at drug treatment facilities in 2007 we found the same prevalence (as mentioned in the discussion).

b. “Assuming the same proportion of diagnosed hepatitis C infection outside the treatment register we calculated the total prevalence of hepatitis C” – again is this assumption reasonable? Is the proportion of diagnosed HCV drug users identified on the drug treatment register likely to be the same as the proportion of diagnosed HCV individuals?

AUTHOR REPLY: We agree this is an important uncertainty in the study. We find it useful to classify those infected with HCV into 3 groups: current injectors, former injectors and never injectors – the latter predominantly infected in health care as mentioned above. We have been able to compare the diagnostic coverage for current injectors (drug related deaths) and nosocomial infections –the Danish HCV look-back and found good agreement with our estimate that 54% of all hepatitis C cases had been diagnosed. However for the former drug users, we had no available independent data source available. This group is likely to have a lower test rate but also a lower HCV prevalence and these two factors counteract each other. We suggest what this could mean to our estimate by deducting the size of the former injector population from the population surveys in Norway and Sweden. In addition we have compared our results to the British estimate where former drug users could be assessed (and had a lower prevalence but contributed more to the HCV population than current users). We have discussed how our assumptions may bias the estimate- that we probably underestimated the true population size.

c. This two-step procedure, in terms of applying a multiplicative factor to the estimated diagnosed number of chronic HCV individuals via the capture-recapture data analysis to obtain an estimate of the total chronic HCV population fails to incorporate any uncertainty with regard to this estimated multiplicative factor.

AUTHOR REPLY: We agree with the comment, and have added the binomial 95% confidence interval to diagnosed fraction based on the observed diagnosed fraction in the drug treatment register (5,136/ 9,463 = 0.5427454 ; 95% CI = 0.5326436- 0.552821). Estimate =9166 (8,973 – 9,877) * 0.5427454 (0.5326436 -0.552821) = 16888 (16231-18543). Multiplying the two CIs is likely to overestimate the true CI, but as we find it narrow (compared to other possible uncertainty and bias in the study) we have not elaborated this further.

5. Results “The regional prevalence ranged from 0.15% in the North region to 0.28% in Copenhagen and the capital region represented 40% of all diagnosed cases – it would be useful to include these regional prevalence estimates (and population sizes for each region) in Table 2, so that it is easier to compare the different regions.

AUTHOR REPLY: We agree, data have been added to the table.

6. Results “Estimate of the undiagnosed population with chronic hepatitis C” – “In the laboratory register among all Danish patients with a positive anti-HCV and a test result for HCV-RNA, 62.2%
(3,999/6,431) were positive” – where does this 3,999 come from? In Table 1 this figure appears to be 3960 (also see point 1 below regarding consistency of numbers).

AUTHOR REPLY: the number 3999 includes all tested for HCVRNA also 39 later excluded (age, no address etc).

7. Results “Estimate of the undiagnosed population with chronic hepatitis C” – “We assumed the same diagnostic coverage (54.3%) among hepatitis C patients outside the drug treatment register” – is this realistic? Will the probability of diagnosis be the same for those identified on the drug treatment register and the rest of the HCV population? (See point 4c regarding treating this estimate as known without error in applying the multiplicative factor for undiagnosed HCV infections).

AUTHOR REPLY: we accept this comment and refer to the answers given to point 4 and reviewer 2. We have tried to clarify this weakness in the study in the discussion.

8. Results “Sensitivity analysis” – “A simple model… gave an estimate of 7,012” – however this appears a little “unfair” as a comparison and in advance it would be predicted that the model will perform poorly, as it assumes independence between the sources (and removes strata). Perhaps a better sensitivity analysis would be to remove only the strata, and consider a range of log-linear models (as before), allowing for interactions between the sources. It is also not surprising that considering only 2 sources typically underestimates the population size. For the use of three sources, it appears independence is again assumed between the sources – why not allow possible interactions? I do not see the point of removing the 749 patients from the simple analyses and rerunning – why not do this sensitivity analysis for the full data analysis (stratified and allowing log-linear interactions).

AUTHOR REPLY: Our ‘sensitivity analyses’ were more about examining the impact of changing the inclusion/exclusion criteria rather than examining the impact of fitting differing statistical models or otherwise altering the statistical approach to the analysis. We agree with the referee that this would be interesting, but feel that this is beyond the scope of this paper, particularly with respect to the readership of the journal. We have changed the title to the less ambitious “Excluded populations”; removed the simple estimate and focused on the effect of not including patients only tested for antiHCV.

9. Results – it might be useful to summarise what could be regarded as the most important interactions between the sources for the different stratified analyses in terms of the interactions being identified in the majority of cases (e.g. “80% of the analyses across the strata identified an interaction between sources A and B; furthermore when the interaction was identified, it was positive in nature”).

AUTHOR REPLY: We have added a short summary of the most frequent interactions across strata but think a detailed description is beyond the scope of this paper. As mentioned in our response to the 2. reviewer’s comment 7, we could provide tables identifying the interactions but feel that this would be excessive.

10. Discussion para 3 starting “The assumption of independence...” – this whole paragraph is unclear.

AUTHOR REPLY: The paragraph was not essential to the discussion and has been deleted.
11. Table 1 and Figure 1 – I cannot “marry” the numbers together provided in these. Using Figure 1, I obtain estimates of the number of individuals identified by each source (in order of Table 1) as 6191, 3225, 4834, 3100 (compared to 3960, 2890, 4484, 3065).

AUTHOR REPLY: Table 1 is correct; the Venn diagram in figure 1 have been corrected, but later omitted to simplify figure 1 as suggested below.

12. Discussion final sentence “Balancing the above stated possible bias and uncertainty we estimate that the total population with chronic hepatitis C is probably in the range 15,000 – 21,000 (0.35%-0.48%).” This seems a little ad-hoc to conclude the discussion with.

AUTHOR REPLY: This sentence has been deleted and instead we have mentioned the range of estimates in the conclusion.

Minor essential revisions
1. Abstract, Results, “We estimated that 46% of the hepatitis C infected individual had not been diagnosed” – wording is poor – rephrase.

AUTHOR REPLY: The sentence has been rephrased.

2. Abstract, Results “16,907 (“ – one of these brackets is not closed. Also “ 18,216), - 0.39%” – use of hyphen is unclear (it could look like a minus sign), perhaps rephrase to something along the lines of “18,216), this corresponds to 0.39%”.

AUTHOR REPLY: This has been corrected as suggested.

3. Abstract, Conclusions “Half of the patients with chronic hepatitis C in Denmark have been identified” – the estimate is 6935/16907 (=41%) – which to me is less considerably than half.

AUTHOR REPLY: This has now been corrected to “Less than half”.

4. Methods, Data sources: DANVIR “177.453” should read “177,453” (i.e. comma instead of a full stop).

AUTHOR REPLY: This has been corrected.

5. Methods, Data sources: Communicable Diseases Register “The register is estimated to cover 35-40% of individuals diagnosed with hepatitis C” – please clarify – is this acute or chronic HCV or simply chronic HCV?

AUTHOR REPLY: The number refer to chronic infections and this has now been corrected

6. Results “Estimate of the undiagnosed population with chronic hepatitis C” –“16,906 (“ – one of these brackets is not closed.

AUTHOR REPLY: This has been corrected

7. Results “Sensitivity analysis” – “HCVRNA” -> “HCV-RNA”; and “age gender” ->“age, gender”.

AUTHOR REPLY: This section has been modified.

8. Discussion end of para 2 “it is likely that 62.2% (3064 patients) of this group had chronic HCV infection” – this is poorly worded statistically – the “probability” that exactly 3064 of these patients had chronic HCV is very small – however, it would be the “best point estimate” for the number of chronic HCV – ideally there should be a confidence interval on this estimate.

AUTHOR REPLY: This has been corrected as suggested.
9. Discussion para 4 “found a 40% hepatitis C prevalence, identical to”… “treatment register were identical to the survey results” – poor use of the word “identical” – perhaps use “very similar” or “consistent with”.
AUTHOR REPLY: This has been corrected as suggested.

10. Table 1 and 2 “Zealand” and “Sealand” – which is correct?
AUTHOR REPLY: Zealand is correct and this has been corrected

Discretionary revisions
1. Methods, Data sources: Drug Treatment Register – it would perhaps be useful to add in some specific numbers, such as the total number on the register, number identified or tests for chronic HCV.
AUTHOR REPLY: The suggestion has been accepted except for the hospital register where only HCV diagnosis were extracted. The total number of all diagnosis for the whole Danish population seems of little relevance.

2. Figure 1 – I found this very difficult to follow – perhaps split into 2 figures and provide some further explanation to the figures?
AUTHOR REPLY: We have omitted the overlay diagram and tried to clarify the rest of the figure.