LSHTM Research Online

https://researchonline.lshtm.ac.uk

Diagnoses of, and deaths from, severe liver disease due to hepatitis C in England between 2000 and 2005 estimated using multiple data sources

A. G. MANN*, M. E. RAMSAY, L. J. BRANT, M. A. BALOGUN, A. COSTELLA
AND H. E. HARRIS, on behalf of the HPA Sentinel Surveillance Scheme†

Immunisation Department, Health Protection Agency Centre for Infections, London, UK

SUMMARY

Matching individuals reported to a sentinel surveillance scheme for hepatitis C between 2000 and 2005 to individuals with a hospital episode for hepatitis C-related liver disease in the same hospitals, we estimated that the number of cases of hepatitis C-related end-stage liver disease in these English hospitals was 42% (597/419) higher than Hospital Episode Statistics (HES) would indicate. Further, matching records of hepatitis C-related deaths in HES to death certificates, we estimated that, between 2000 and 2005, the true number of deaths from hepatitis C-related end-stage liver disease was between 185% (353/124) and 257% (378/106) higher than the number recorded in routine mortality statistics. We provide estimates of under-recording that can be used to modify existing models of disease burden due to hepatitis C and provide a simple approach to improve the monitoring of trends in severe hepatitis C-related morbidity over time.

Key words: England, hepatitis C, hepatocellular carcinoma, liver cirrhosis.

INTRODUCTION

In England and Wales, the age-standardized mortality rates for liver cirrhosis increased by 75% between 1992 and 2002 [1]. The component of this burden that can be attributed to hepatitis C, however, is unclear [2]. Modelling is underway to improve our understanding of the current and future impact of chronic liver disease due to hepatitis C in England [3]. Use of routinely collected data, such as data on hospital admissions collated in Hospital Episode Statistics (HES), allows us to estimate and monitor the number of individuals in England infected with hepatitis C virus who experience sequelae of their infection. As the number of cases of hepatitis C-related liver disease will be strongly influenced by increased case-finding and improved investigation of infected individuals, we chose to monitor cases of end-stage liver disease (ESLD) that would be expected to spontaneously present with signs or symptoms of disease. The number of such cases identified in routine data sources is likely to underestimate the true number of cases attributed to hepatitis C because of failure to diagnose hepatitis C as the cause and because of inadequate recording and coding of all diagnoses. We aimed to estimate the under-recording of diagnoses and deaths due to hepatitis C-related ESLD by HES and the Office for National Statistics (ONS).

* Author for correspondence: Ms. A. G. Mann, Infectious Disease Epidemiology Unit, Department of Epidemiology & Population Health, London School of Hygiene & Tropical Medicine, Keppel Street, London WC1E 7HT, UK.
(Email: andrea.mann@lshtm.ac.uk)
† Members of the HPA Sentinel Surveillance Scheme are listed in the Appendix.

METHODS

Data sources and case definitions

Sentinel surveillance of ESLD

In 2000, the Health Protection Agency Centre for Infections (HPA CfI) established sentinel surveillance of ESLD due to viral hepatitis to supplement routinely collected data [4]. One nominated physician with a special interest in liver disease from each centre reports the number of hepatitis C-infected individuals (both in-patients and out-patients) presenting for the first time with a symptom of ESLD. Information on individuals presenting for the first time to each hospital with ascites, bleeding varices, hepatic encephalopathy, or primary hepatocellular carcinoma due to hepatitis C are reported monthly by the nominated physician to the HPA CfI. Evidence of hepatitis C virus infection is defined as a positive test for antibody to hepatitis C virus or for hepatitis C virus RNA. The case definition was chosen in collaboration with experienced physicians to identify individuals with severe enough complications to present spontaneously to a hospital with clinical manifestations of liver disease. This definition, based solely on clinically apparent signs or symptoms, was considered unlikely to be influenced by increased case-finding for hepatitis C, accompanied by active investigation (such as liver biopsy) of such cases.

Participating centres have included a range of hospitals, from specialist liver transplant centres to district general hospitals, in a range of regions. To encourage complete reporting clinicians are requested to report to us each month, even if no cases fitting the case definition have been seen. Information extracted from the database included hospital, month and year of birth, sex, diagnosis, and date of presentation.

Mortality data from death certificates

Information was available from the death certificates of patients who died between 2000 and 2005 with any mention of hepatitis C in the text or presence of an ICD10 code for hepatitis C (B17.1 or B18.2) in any of the 13 fields supplied by the ONS [5]. These data are provided on request to the HPA CfI for use as part of national surveillance. Deaths due to hepatitis C-related ESLD were defined as those for which there was mention of severe liver disease or hepatocellular carcinoma in the text or in the ICD10 codes (R18, K72.9, K72.0, K72.1, K70.4, I85.0, I85.9, I98.2, C22.0) on the certificate. The code for acute hepatitis

was included after review of the death certificates for all deaths with ICD codes found that none specified an acute infection in the text entries. This suggests that where chronic infection is not specified, the code for acute infection is used as a default by ONS. Place of death was available for all records and each death was categorized as having occurred in hospital or out of hospital.

HES

Information on individuals diagnosed with, and who died from, severe liver disease due to hepatitis C was available from HES recorded between 1 April 1997 and 31 March 2006 [6]. HES contain separate records for each episode of hospital care that an individual receives within a given admission to a National Health Service (NHS) hospital in England. Each HES record contains clinical, demographic, administrative, and geographical information about an individual patient. Data are supplied to the HPA in an anonymized form, therfore it is not possible to validate the accuracy of clinical or other coding. A person was defined as having been diagnosed in HES with ESLD if one or more of their episodes included an ICD10 code for ascites, bleeding varices, or hepatic encephalopathy (R18, K72.9, K72.0, K72.1, K70.4, I85.0, I85.9, I98.2) or hepatocellular carcinoma (C22.0) in any of the 14 diagnosis fields. These episodes were then matched using the unique identifier field (HESID) to all episodes coded to hepatitis C (ICD10 codes B17.1 or B18.2) within the same financial year to identify those likely to be diagnosed with hepatitis C-related ESLD. The ICD10 code for acute hepatitis C was included because of our experience with ONS data.

To identify individuals newly presenting with hepatitis C-related ESLD, the individuals diagnosed as above were de-duplicated using HESID, to identify the first date of admission for each individual. To avoid including individuals whose first admission for hepatitis C-related severe liver disease occurred prior to April 1997, only individuals with first admission dates between 2000 and 2005 were analysed. It was therefore assumed that it was unlikely that individuals diagnosed with ESLD disease prior to April 1997 would then be admitted after January 2000 without any intervening care. To identify individuals where hepatitis C-related ESLD contributed to death, individuals diagnosed with hepatitis C-related ESLD were also matched to all episodes with a death

Table 1. *The capture–recapture method*

	Dataset A		
	Present	Absent	Total
Dataset B			
Present	a	b	
Absent	c	$d = b*c/a$	
Total			$e = a+b+c+d$

recorded in the same financial year and with dates of death between 2000 and 2005.

Statistical methods

Pooling data over the years 2000–2005, records for individuals who first presented with ESLD related to hepatitis C in one of the sentinel centres were matched between the sentinel surveillance and HES on month and year of birth, sex, and hospital provider code. For centres that did not participate in the sentinel surveillance scheme for the whole period, cases were matched to individuals in HES only for the period of participation.

Deaths with hepatitis C-related ESLD recorded in HES and registered by the ONS were matched on month and year of birth, date of death, and sex. Individuals without date of birth and sex were necessarily excluded from the matching exercise. Matching records were checked by hand for additional information (such as place of death, date of admission, ethnic status) to corroborate the match.

A capture–recapture analysis was undertaken to estimate the total number of diagnoses of hepatitis C-related severe liver disease in participating centres. A second capture–recapture analysis was conducted to estimate the total number of deaths due to severe liver disease and hepatitis C in English hospitals. Estimates of the total number individuals diagnosed with, and who died from, severe liver disease due to hepatitis C (e) were calculated using the formula given in Table 1.

Goodness-of-fit-based confidence intervals were calculated using the method of Regal & Hook [7]. All analyses were done using Stata statistical software (version 8.2, Stata Corp, College Station, TX, USA) and Microsoft Excel (Microsoft, Redmond, WA, USA).

A sensitivity analysis was performed to calculate the total number of individuals diagnosed, or dying, with hepatitis C-related severe liver disease assuming that we failed to detect 5% of true matches (false negatives) and assuming 5% of the matches were not, in fact, the same individual (false positives).

RESULTS

Sentinel surveillance and HES

Between 2000 and 2005, nine English centres participated for between 1 and 6 years, returning a total of 377/468 (80%) of the expected monthly returns. A total of 247 cases due to hepatitis C were reported to sentinel surveillance from these collaborating centres, compared to 419 recorded in HES for the same centres (Table 2). After excluding 11 cases with missing date of birth and five duplicates from HES, 166 cases were matched between both datasets (leaving 80 ESLD and 237 HES cases that did not match), making the total estimated cases in that period to be 597 [95% confidence interval (CI) 562–642]. Sensitivity analysis assuming 5% false-positive and 5% false-negative matches gives a total estimated number of cases of 629 (95% CI 587–679) and 570 (95% CI 540–608), respectively.

The cases reported via ESLD that matched to cases in HES were similar to those that did not match in terms of sex, diagnosis and age. The proportion of cases in males was 128/166 (77%) and 60/80 (75%), and with a diagnosis of hepatocellular cancer was 55/166 (33%) and 31/80 (39%) in matching and non-matching cases, respectively. The mean age of matching cases was 53 years compared to 55 years in those that did not match.

Mortality: HES and ONS

A total of 935 deaths were reported to ONS in the period, one was excluded from the matching because of missing date of birth. Of the remaining 934, 740 had a hospital as place of death. In the same period 1362 deaths were reported in HES, 16 were excluded from matching due to missing dates of birth and two were found to be duplicates, leaving 1344 available for matching. A total of 417 cases matched between the two datasets, leaving 927 deaths reported only to HES and 323 only to ONS (Table 3).

The ONS deaths that matched to deaths in HES were similar to those that did not match in terms of sex, diagnosis and age. The proportion of cases in males was 319/417 (76%) and 222/323 (69%), and with a diagnosis of hepatocellular cancer was 150/417 (36%) and 121/323 (38%) in matching and

Table 2. *Individuals with hepatitis C-related severe liver disease reported to HPA sentinel surveillance and/or recorded in HES, 2000–2005**

	Trust 1		Trust 2		Trust 3		Trust 4		Trust 5		Trust 6		Trust 7		Trust 8		Trust 9		Total	
	ESLD	HES	ESLD	HES	ESLD	HES	ESLD	HES	ESLD	HES	ESLD	HES	ESLD	HES	ESLD	HES	ESLD	HES	ESLD	HES
2000	14	23	—	—	7	4	1	2	—	—	0	1	5	34	—	—	1	4	28	68
2001	25	33	1	1	9	7	1	2	—	—	0	0	—	—	—	—	0	11	36	54
2002	36	37	2	1	3	4	—	—	—	—	0	2	—	—	0	2	6	9	47	55
2003	26	34	5	5	2	9	—	—	—	—	0	1	—	—	1	1	4	7	38	57
2004	20	27	2	4	4	5	—	—	22	16	0	2	—	—	0	0	3	11	51	65
2005	17	48	7	9	8	12	—	—	9	21	0	5	—	—	1	4	5	21	47	120
Total	138	202	17	20	33	41	2	4	31	37	0	11	5	34	2	7	19	63	247	419

HPA, Health Protection Agency; HES, Hospital Episode Statistics; ESLD, end-stage liver disease.
* Blank cells represent years when this centre was not participating in the HPA surveillance.

non-matching deaths, respectively. The mean age of both groups was 56 years.

DISCUSSION

The yearly number of new cases of hepatitis C virus infection in England is estimated to have increased from about nil in 1960 to around 12 000 in 1995 [3]. To inform health-care planning for hepatitis C and to make the case for providing resources for the treatment of hepatitis C virus-infected individuals it is vital to know the current and future burden of infection. Routinely collected data, including HES and death certificates, can be used to monitor the number of diagnoses of, and deaths from, hepatitis C. However, as shown in the present study, these sources of data underestimate the burden of disease. Using capture–recapture techniques, we suggest that the total number of individuals diagnosed with hepatitis C-related ESLD in England is around 42 % higher than the number reported in HES and the total number of individuals who die with this condition in hospital is more than 200 % higher than the number recorded by ONS. Any estimates of the future burden of hepatitis C and severe liver disease using routine data sources needs to adjust for under-reporting and any analysis of trends need to account for possible changes in under-reporting over time.

Capture–recapture techniques have been used for a number of diseases to estimate the level of 'missing' data in routine sources [8–11]. There are four main assumptions underlying this technique: that the underlying population from which the individuals are drawn remains unchanged during the study period, that the lists of individuals being compared are independent, that each member of the population has the same probability of being on a given list, and that individuals are matched correctly [12, 13]. It is also important that the case definition for each list has a high specificity (i.e. a low false positivity rate), otherwise capture–recapture methods will overestimate the burden of disease [9].

The underlying populations for this study were the catchment populations of the sentinel centres and residents of England. Although there will be some changes in both of these populations, given the severe and uncommon nature of the disease studied, changes within the period are unlikely to be significant enough to influence this analysis. We will have missed some true matches as cases may have been reported to the other data source outside of the study period and we

Table 3. *The estimated total number of individuals in England dying with hepatitis C-related severe liver disease, 2000–2005*

	ONS deaths		HES deaths	Capture–recapture analysis					
	Total	In hospital		ONS only	HES only	HES and ONS	Total estimated deaths in hospital (95% CI)	5% false positive (95% CI)	5% false negative (95% CI)
2000	152	124	168	65	109	59	353 (310–416)	372 (323–443)	329 (293–385)
2001	139	114	186	50	122	64	331 (295–383)	349 (307–406)	310 (281–356)
2002	133	106	203	49	146	57	378 (331, 446)	397 (346–476)	348 (309–403)
2003	157	124	228	59	163	65	435 (383–509)	458 (399–538)	401 (356–458)
2004	169	131	252	47	168	84	393 (359–441)	414 (373–467)	369 (342–412)
2005	185	141	307	53	219	88	492 (447–554)	518 (465–585)	457 (419–506)
2000–2005	934	740	1344	323	927	417	2385 (2268–2519)	2511 (2380–2663)	2219 (2122–2333)

HES, Hospital Episode Statistics; ONS, Office for National Statistics; CI, confidence interval.

may have matched incorrectly due to the limited identifiers available in HES. However, assuming that ±5% cases should have been matched did not materially alter our conclusions. To minimize the possibility of including cases of ESLD which are not truly hepatitis C-related we chose a clinically specific case-definition and required the hospitalized cases to be recorded with a diagnosis of hepatitis C in the same financial year. However, we were not able to validate diagnostic coding in HES and it is possible that in some cases there was another cause for the severe liver disease (such as alcohol) and that hepatitis C virus infection was not thought to have contributed to the condition. We assumed, however, that in anyone diagnosed with hepatitis C and symptoms of ESLD in the same financial year, hepatitis C was either causal or contributing to morbidity, for example as a co-factor with alcohol.

We cannot exclude the possibility that a person's probability of being on one list is dependent on their probability of being on the second list. A diagnosed patient not ascertained in HES (e.g. because the results of their hepatitis C test were not included in their hospital records) may also be unlikely to have been reported via sentinel surveillance. However, participating clinicians with an interest in viral hepatitis are likely to actively seek hepatitis test results for a patient seen with a life-threatening liver-related condition and so we feel that this is unlikely to be a major problem in the sentinel surveillance. A clinician who completes a death certificate, however, may use only the information recorded in the notes and so it is likely that there is some dependency between death certificates and hospital episodes. Another potential limitation is the possibility of heterogeneity, where

particular members of the population may be less likely to appear on a given list. Certain groups of individuals may be less likely to have an episode coded as hepatitis C-related when seeking hospital care, perhaps because of concerns about stigma or insurance risks. However, comparison of those matched with those reported only to sentinel surveillance or recorded by the ONS suggested that the two groups were largely similar.

These limitations may have produced an over- or underestimate of the total burden of hepatitis C-related disease. It is plausible that failure to match because of missing or incorrect identifiers has occurred, or that certain groups are less likely to appear on certain lists, leading to an overestimate of total cases or deaths. In contrast, possible dependency between the data sources used may have led to an underestimate of total burden. Importantly, both datasets exclude individuals where the diagnosis, of either hepatitis C or severe liver disease, has never been made. This problem should have been minimized by the choice of a case definition of severe disease, which would normally warrant intensive investigation, including hepatitis C testing. However, individuals who present and die quickly before investigations can be performed may not be diagnosed and will therefore not be included in either dataset. This may be particularly true for individuals where other more common aetiological factors, such as alcohol, are present. The number of individuals with less severe manifestation of hepatitis C infection, such as early cirrhosis, is likely to be underestimated to a much greater extent.

Despite the limitations we have discussed, we believe that our results provide an estimate of

under-ascertainment that may be useful for modelling. They also provide a mechanism for monitoring trends and for validating predictions based on previous models. With this new information we will be better able to inform health-care planning and the targeting of scarce resources within the NHS.

APPENDIX. HPA Sentinel Surveillance Scheme

T. Woodall, G. Alexander (*Addenbrooke's Hospital, Cambridge, UK*); J. Collier, R. Chapman (*John Radcliffe Hospital, Oxford, UK*); M. Farrant (*Royal United Hospital, Bath, UK*); N. Tatman, N. Naoumov (*University College Hospital, London, UK*); A. Clements, M. Cramp (*Derriford Hospital, Plymouth, UK*); A. Brind (*Stoke City General Hospital, Stoke, UK*); S. Solaiman, G. Foster (*Royal London Hospital, London, UK*); S. Ryder (*Queens Medical Centre, Nottingham, UK*); J. Green (*Llandough Hospital, Cardiff, UK*).

ACKNOWLEDGEMENTS

DECLARATION OF INTEREST

None.

REFERENCES

1. **Bosetti C, et al.** Worldwide mortality from cirrhosis: an update to 2002. *Journal of Hepatology* 2007; **46**: 827–839.
2. **Department of Health.** Hepatitis C strategy for England. London: Department of Health, 2002.
3. **Sweeting MJ, et al.** The burden of hepatitis C in England. *Journal of Viral Hepatitis* 2007; **14**: 570–576.
4. **Health Protection Agency.** Hepatitis C in England – The First Health Protection Agency Annual Report. London: Health Protection Agency; 2005 December 2005. Report No.: 0 901144 78 9. (http://www.hpa.org.uk/webw/HPAweb&HPAwebStandard/HPAweb_C/1196942156476?p=1158945066450. Accessed 12 December 2005.
5. **Office for National Statistics.** Mortality (http://www.statistics.gov.uk/CCI/Nscl.asp?ID=8371&Pos=2&ColRank=1&Rank=198). Accessed 9 May 2008.
6. **HESonline.** Hospital Episode Statistics (http://www.hesonline.nhs.uk/). Accessed 9 May 2008.
7. **Regal RR, Hook EB.** Goodness-of-fit based confidence intervals for estimates of the size of a closed population. *Statistics in Medicine* 1984; **3**: 287–291.
8. **Wu C, et al.** Estimating the mortality rate of hepatitis C using multiple data sources. *Epidemiology and Infection* 2005; **133**: 121–125.
9. **Trotter C, et al.** Ascertainment of meningococcal disease in Europe. *Eurosurveillance* 2005; **10**(12): 247–250.
10. **Klein S, Bosman A.** Completeness of malaria notification in the Netherlands 1995–2003 assessed by capture-recapture method. *Eurosurveillance* 2005; **10**(10): 244–246.
11. **Hickman M, et al.** Estimating the prevalence of problem drug use in inner London: a discussion of three capture–recapture studies. *Addiction* 1999; **94**: 1653–1662.
12. **Seber GA, Huakau JT, Simmons D.** Capture-recapture, epidemiology, and list mismatches: two lists. *Biometrics* 2000; **56**: 1227–1232.
13. **Hook EB, Regal RR.** Capture–recapture methods in epidemiology: methods and limitations. *Epidemiologic Review* 1995; **17**: 243–264.