

---

## Estimating the mortality rate of hepatitis C using multiple data sources

---

C. WU<sup>1\*</sup>, H.-G. CHANG<sup>1,2</sup>, L.-A. MCNUTT<sup>1</sup> AND P. F. SMITH<sup>1,2</sup>

<sup>1</sup> *Department of Epidemiology, School of Public Health, University at Albany, State University of New York, Rensselaer, NY, USA*

<sup>2</sup> *Division of Epidemiology, New York State Department of Health, Albany, NY, USA*

(Accepted 15 August 2004)

### SUMMARY

The New York State hospital discharge database and the multiple cause-of-death file were used to estimate the mortality rate of hepatitis C in New York State excluding New York City in 1997. The mortality rate with hepatitis C was severely underestimated when each data source was used alone. Applying the capture–recapture method using the hospital discharge database and the multiple cause-of-death file appears to be an efficient method to estimate the mortality rate with hepatitis C.

### INTRODUCTION

Hepatitis C virus (HCV) infection is an important public health issue in the United States. The Third National Health and Nutrition Examination Survey, which was conducted in a representative sample of the civilian non-institutionalized population between 1988 and 1994, estimated that 1·8% of the population was HCV antibody-positive in the United States [1]. This estimated prevalence revealed that approximately 3·9 million people had been infected with HCV in the United States.

Acute HCV infection is often asymptomatic or causes mild clinical manifestations. The acute illness usually lasts 2–12 weeks and rarely causes death by fulminant hepatic failure [2, 3]. Following acute infection, 75–85% of infected people will develop chronic HCV infection. Among persons with chronic HCV infection, 10–20% will develop cirrhosis of the

liver and 1–5% will develop hepatocellular carcinoma (HCC) within 20–30 years. Once cirrhosis has developed, the risk of HCC increases to 1–4% per year [4–7]. In the United States, chronic liver disease accounts for 25 000 deaths annually, and a population-based study suggested that HCV accounts for 40% of chronic liver disease [8]. Thus, it was estimated that HCV causes 8000–10 000 deaths every year in the United States [8]. Hepatitis C will create a substantial health and economic burden over the next 10–20 years [9]. Treatment with interferon in combination with ribavirin for 6–12 months can result in viral clearance in approximately 50% of patients [10, 11]. Treatment may also reduce the risk of HCC [12].

Administrative data, including statewide hospital discharge databases and multiple cause-of-death files, are often used for estimating disease occurrence and mortality. The objectives of this study were to estimate the mortality rate with hepatitis C in New York State excluding New York City (Upstate New York) in 1997 using multiple administrative data sources and to assess the quality of hepatitis C coding in a statewide hospital discharge database.

\* Author for correspondence: Dr C. Wu, Department of Epidemiology, University at Albany, Rensselaer, NY 12144, USA. (Email: ctw09@health.state.ny.us)

## MATERIALS AND METHODS

### Data sources

Two primary data sources were used to identify hepatitis C mortality cases in Upstate New York in 1997; one was the 1997 New York State hospital discharge database, and the other was the 1997 New York State multiple cause-of-death file.

The New York State hospital discharge database includes information on each in-patient admission in a non-federal, non-psychiatric acute care hospital in New York State. Information for each admission included the patient's sex, race, ethnicity, date of birth, county of residence, date of discharge, discharge diagnoses, and vital status at discharge. When the hospital discharge database was used, a hepatitis C mortality case was defined as a death with hepatitis C (ICD-9-CM codes 070.41, 070.44, 070.51, and 070.54) listed as a discharge diagnosis [13].

Each death record in the New York State multiple cause-of-death file includes the deceased's demographic and medical information, including sex, race, date of birth, county of residence, date of death, underlying cause of death, and contributing conditions of death (coded to only 4-digit ICD-9-CM codes). When the multiple cause-of-death file was used, a hepatitis C mortality case was defined as a death with hepatitis C listed as an underlying or contributing condition.

### Medical record review

To evaluate the positive predictive value of hepatitis C coding for mortality cases on the hospital discharge database, we requested and reviewed the medical records of hepatitis C mortality cases identified from the hospital discharge database. The face sheet, admission history and physical examination, discharge summary, and laboratory reports in the medical record for the last hospitalization of those individuals were reviewed to identify hepatitis C mortality cases. The patient's demographic information (i.e. sex, race, date of birth, date of death), disease history, discharge diagnoses, and laboratory data [i.e. hepatitis A, B, and C viruses, and human immunodeficiency virus (HIV) diagnosis test results] were abstracted.

Based on medical record review, a death with hepatitis C was defined as a death in a person having hepatitis C as a discharge diagnosis, having a history of hepatitis C that was recorded in the hospital record, being positive to HCV antibody test (enzyme

Table 1. *Demographic characteristics of hepatitis C mortality cases and mortality rates of hepatitis C, Upstate New York, 1997*

Variables	Hospital discharge database			Cause-of-death file		
	N	%	M*	N	%	M*
Sex						
Male	107	67.7	2.0	55	63.2	1.0
Female	51	32.3	0.9	32	36.8	0.6
Age (years)						
20–39	23	14.6	0.7	5	5.7	0.2
40–49	66	41.8	3.9	34	39.1	2.0
50–59	26	16.5	2.3	16	18.4	1.4
60–69	16	10.1	1.8	9	10.3	1.0
70–79	16	10.1	2.3	13	14.9	1.8
80+	11	7.0	3.0	10	11.5	2.7
Race						
White	96	60.8	—	75	86.2	0.8
Black	31	19.6	—	11	12.6	1.3
Other	16	10.1	—	1	1.1	0.4
Unknown	15	9.5	—	0	0	—
Total	158	100.0	1.4	87	100.0	0.8

\* M is mortality rate, per 100 000 population.

immunoassay alone, or in combination with recombinant immunoblot assay), or being positive to HCV RNA polymerase chain reaction test.

### Statistical analysis

The mortality rate with hepatitis C in Upstate New York in 1997 for the total population, and by sex, race, and age was calculated for each data source. The population estimate of Upstate New York was obtained from New York State Department of Health Vital Statistics data [14]. The capture–recapture method was used to estimate the total number of deaths by matching hepatitis C mortality cases identified from the two different data sources by sex, date of birth and date of death. The Appendix shows the capture–recapture formula used to calculate the estimated total number of deaths and its variance given two data sources [15, 16].

The positive predictive value of the coding of hepatitis C for mortality cases in the hospital discharge database was calculated as the proportion of hepatitis C mortality cases that were confirmed by medical record review.

Table 2. *Estimated total number of deaths with hepatitis C applying the capture–recapture method to the cases from hospital discharge database and multiple cause-of-death file, Upstate New York, 1997*

Variable	N1*	N2*	N3*	N4*	N* (95% CI†)	M‡ (95% CI†)
Total	28	130	59	274	491 (355–627)	4.5 (3.2–5.7)
Sex						
Male	18	89	37	183	327 (214–440)	6.1 (4.0–8.2)
Female	10	41	22	91	164 (88–240)	2.9 (1.6–4.3)

\* N1, number of deaths with hepatitis C on both the hospital discharge database and cause-of-death file; N2, number of cases on the hospital discharge database only; N3, number of cases on the cause-of-death file only; N4, estimated number of cases missed by both data sources; N, estimated number of total cases.

† 95% confidence interval.

‡ M, estimated mortality rate, per 100 000 population.

## RESULTS

### Estimating deaths with hepatitis C using single data sources

The number of deaths with hepatitis C varied by data source (Table 1). From the statewide hospital discharge database, 158 hepatitis C mortality cases in Upstate New York in 1997 were identified. From the multiple cause-of-death file, 87 deaths with hepatitis C were identified; 78 (90%) of them had hepatitis C as an underlying cause of death and 9 (10%) had hepatitis C as a contributing condition. The estimated mortality rate with hepatitis C in Upstate New York in 1997 was 1.4/100 000 population based on the hospital discharge database, and 0.8/100 000 population based on the cause-of-death file.

While different data sources produced different estimates of mortality with hepatitis C, the distributions by sex and age for the different data sources were quite similar. Both data sources revealed that males had higher mortality rates with hepatitis C than females (2.0 vs. 0.9/100 000 population for the hospital discharge database; and 1.0 vs. 0.6/100 000 population for the cause-of-death file). The data from the multiple cause-of-death file revealed that Blacks had higher mortality rates than Whites (1.3 vs. 0.8/100 000 population for the death file). Mortality rates by race were not calculated based on the hospital discharge database because 9% (15/158) of the deaths had race coded as ‘unknown’ (Table 1).

### Estimating the mortality rate with hepatitis C using the capture–recapture method

Comparing the hospital discharge database and multiple cause-of-death file identified 28 cases on both

data sources, 130 cases on the hospital discharge database only, and 59 cases on the multiple cause-of-death file only (Table 2). The capture–recapture method yielded an estimated number of cases missed by both data sources of 274 and an estimated total number of individuals who died with hepatitis C of 491 (95% CI 355–627). Thus, the overall estimated mortality rate with hepatitis C was 4.5/100 000 population (95% CI 3.2–5.7/100 000 population) in 1997. The estimated mortality rate for males and females was 6.1/100 000 population (95% CI 4.0–8.2/100 000 population) and 2.9/100 000 population (95% CI 1.6–4.3/100 000 population) respectively.

### Positive predictive value of hepatitis C coding for mortality cases in the hospital discharge database

The medical records of these 158 hepatitis C mortality cases were requested for review. Medical records were received for 134 out of 158 cases (85%). Among these 134 deaths, 125 were confirmed to be hepatitis C mortality cases through the medical record review. Therefore, the positive predictive value of hepatitis C coding in the hospital discharge database was 93.3% (125/134) when the medical record review was treated as the gold standard.

## DISCUSSION

Estimating mortality rates of diseases of public health importance is often accomplished using multiple cause-of-death files. This study suggested that the mortality rate with hepatitis C might be severely underestimated when using the multiple cause-of-death file alone (0.8/100 000 population in Upstate New York in 1997).

The use of the statewide hospital discharge database alone provided a higher estimate of mortality rate with hepatitis C (1.4/100 000 population), although this estimate was still likely to be an underestimate. Using two data sources and applying the capture–recapture method resulted in a greater estimate of the mortality rate with hepatitis C (4.5/100 000 population); this estimate was approximately six-fold that based on the multiple cause-of-death file alone and approximately three-fold that based on the hospital discharge database alone.

This study found that 93% of hepatitis C mortality cases identified from the hospital discharge database were confirmed by the medical record reviews. However, the estimate of positive predictive value of hepatitis C coding for mortality cases in the hospital discharge database could be an underestimate because only medical records of the patients' last hospitalizations were reviewed. The sensitivity of hepatitis C coding was not estimated because of the difficulty of identifying all hepatitis C mortality cases through medical record review. We were also not able to examine the accuracy of hepatitis C coding in the multiple cause-of-death file because of the difficulty of obtaining medical records based on information from the cause-of-death file.

A few caveats are of note. First, the hospital discharge database tends to be positively correlated with the multiple cause-of-death file because physicians also fill out death certificates when patients die in hospitals. This positive correlation of the two data sources may result in an underestimate of the total number of deaths when applying the capture–recapture method. Second, because only 4-digit ICD-9 codes were available in the multiple cause-of-death file, some viral hepatitis other than hepatitis C could be misclassified as hepatitis C. These codes include 070.42 and 070.52 (hepatitis D without mention of active hepatitis B disease), 070.43 and 070.53 (hepatitis E), and 070.49 and 070.59 (other specified viral hepatitis) [13]. Although a lack of specificity could lead to an overestimation of the number of deaths, we believe that this misclassification should have minimal effect on the estimate because these diseases are rare.

In summary, this study found that using either multiple cause-of-death file or hospital discharge database alone resulted in a severe underestimate of the hepatitis C mortality rate. Applying the capture–recapture method by matching the hospital discharge database to the multiple cause-of-death file appeared useful for monitoring the hepatitis C mortality rate.

## APPENDIX. Capture–recapture method using two data sources

		Source B		
Source A	+	–		Total
+	$N1$	$N2$		$N1 + N2$
–	$N3$	$N4$		$N3 + N4$
Total	$N1 + N3$	$N2 + N4$		$N$

+ Represents presence on one data source; – represents absence on one data source.

$N1$ , number of records on both data sources;  $N2$ , number of records on data source A only;  $N3$ , number of records on data source B only. The maximum-likelihood estimate of the total number of records missed by both data sources is  $N4 = (N2 * N3) / N1$  if these two data sources are independent. Thus, the total number of deaths related to hepatitis C can be calculated by  $N = N1 + N2 + N3 + N4$ . The estimated variance of  $N$  is  $N(1 - P1)(1 - P2) / P1P2$ ; in which  $P1 = (N1 + N2) / N$ , and  $P2 = (N1 + N3) / N$  [15].

## REFERENCES

- Alter MJ, Kruszon-Moran D, Nainan OV, et al. The prevalence of hepatitis C virus infection in the United States, 1988 through 1994. *N Engl J Med* 1999; **341**: 556–562.
- Alter MJ, Margolis HS, Krawczynski K, et al. The natural history of community-acquired hepatitis C in the United States. *N Engl J Med* 1992; **327**: 1899–1905.
- Liang TJ, Jeffers L, Reddy RK, et al. Fulminant or subfulminant non-A, non-B viral hepatitis: the role of hepatitis C and E viruses. *Gastroenterology* 1993; **104**: 556–562.
- Nowicki MJ, Balistreri WF. The hepatitis C virus: identification, epidemiology, and clinical controversies. *J Pediatr Gastroenterol Nutr* 1995; **20**: 248–274.
- Degos F. Natural history of hepatitis C virus infection. *Nephrol Dial Transplant* 1996; **11** (Suppl): 16–18.
- Seeff LB, Buskell-Bales Z, Wright EC, et al. Long-term mortality after transfusion-associated non-A, non-B hepatitis. *N Engl J Med* 1992; **327**: 1906–1911.
- Di Bisceglie AM, Order SE, Klein JL, et al. The role of chronic viral hepatitis in hepatocellular carcinoma in the United States. *Am J Gastroenterol* 1991; **86**: 335–338.
- Centers for Disease Control and Prevention. Recommendations for prevention and control of hepatitis C (HCV) and HCV-related chronic diseases. *MMWR* 1998; **47**: 1–38.
- Wong JB, McQuillan GM, McHutchison JG, Poynard T. Estimating future hepatitis C morbidity, mortality, and costs in the United States. *Am J Public Health* 2000; **90**: 1562–1569.
- Manns MP, McHutchison JG, Gordon SC, et al. Peginterferon alfa-2b plus ribavirin compared with

- interferon alfa-2b plus ribavirin for initial treatment of chronic hepatitis C: a randomised trial. *Lancet* 2001; **358**: 958–965.
11. Fried MW, Shiffman ML, Reddy KR, et al. Peginterferon alfa-2a plus ribavirin for chronic hepatitis C virus infection. *N Engl J Med* 2002; **347**: 975–982.
  12. Nishiguchi S, Kuroki T, Nakatani S, et al. Randomized trial of effects of interferon-alpha on incidence of hepatocellular carcinoma in chronic active hepatitis C with cirrhosis. *Lancet* 1995; **346**: 1051–1055.
  13. International classification of diseases, 9th revision, clinical modification, 4th edn. Los Angeles, CA: Practice Management Information Corporation, 1995.
  14. New York State Department of Health. Vital Statistics of New York State, 1997 ([http://www.health.state.ny.us/nysdoh/vital\\_statistics/1997/toc.htm](http://www.health.state.ny.us/nysdoh/vital_statistics/1997/toc.htm)). Accessed September 2002.
  15. Chandra Sekar C, Deming WE. On a method of estimating birth and death rates and the extent of registration. *J Am Stat Assoc* 1949; **44**: 101–115.
  16. Hook EB, Regal RR. Capture-recapture methods in epidemiology. *Epidemiol Rev* 1995; **17**: 243–264.