# Estimating the prevalence of hepatitis C infection in New York City using surveillance data

S. BALTER[1]*, J. H. STARK[2], J. KENNEDY[2], K. BORNSCHLEGEL[1]
AND K. KONTY[2]

[1] *Division of Disease Control, New York City Department of Health and Mental Hygiene, NY, USA*
[2] *Division of Epidemiology, New York City Department of Health and Mental Hygiene, NY, USA*

## SUMMARY

Hepatitis C virus is the most common chronic blood-borne infection in the USA. Based on results of a serosurvey, national prevalence is estimated to be 1·3% or 3·2 million people. Sub-national estimates are not available for most jurisdictions. Hepatitis C surveillance data was adjusted for death, out-migration, under-diagnosis, and undetectable blood RNA, to estimate prevalence in New York City (NYC). The prevalence of hepatitis C infection in adults aged $\geqslant 20$ years in NYC is 2·37% (range 1·53–4·90%) or 146 500 cases of hepatitis C. This analysis presents a mechanism for generating prevalence estimates using local surveillance data accounting for biases and difficulty in accessing hard to reach populations. As the cohort of patients with hepatitis C age and require additional medical care, local public health officials will need a method to generate prevalence estimates to allocate resources. This approach can serve as a guideline for generating local estimates using surveillance data that is less resource prohibitive.

**Key words**: Estimating, hepatitis C, prevalence of disease, surveillance.

## INTRODUCTION

In the USA, hepatitis C virus is the most common chronic blood-borne infection [1]. Untreated chronic hepatitis C can lead to cirrhosis, hepatocellular carcinoma and death [2]. A recent report indicates that deaths from chronic hepatitis C infection now exceed HIV in the USA [3]. The third National Health and Nutrition Examination Survey (NHANES) conducted from 1999 to 2002 estimated the national prevalence of chronic hepatitis C infection to be 1·3% or 3·2 million persons [4]. In New York City (NYC) a similar study (NYC HANES) estimated a chronic hepatitis C infection prevalence of 1·8% in

persons aged $\geqslant 20$ years ($n = 103000$) [5]. Both surveys only sampled the civilian, non-institutionalized population, potentially excluding populations at high risk of hepatitis C infection, including the homeless, prisoners, and injecting drug users. [6–16]. Chak *et al.* adjusted the NHANES estimate to account for this bias, resulting in an estimated national prevalence of 2% [16]. Even though the adjusted NHANES prevalence estimates were more inclusive of persons at risk for hepatitis C infection, the NHANES sample cannot be used for sub-national estimates.

Conducting additional serosurveys to obtain state- or city-specific prevalence estimates are expensive and labour intensive; this approach is also prone to inaccuracy, because the data needed to adjust for potential measurement error is lacking [17–19]. As a result, there is a need for an approach using local surveillance data to generate prevalence estimates for

* Author for correspondence: S. Balter, M.D., New York City Department of Health and Mental Hygiene, 42–09 28th Street, 6th floor CN 22-A, Queens, NY 11101–4132, USA.
(Email: sbalter@health.nyc.gov)

specific jurisdictions to plan allocation of resources. With the advent of improved treatment regimens for hepatitis C [20–23] it may be possible to address the epidemic not just through prevention, but also through treatment. Thus, prevalence estimates are important for local jurisdictions planning to allocate resources for treatment.

Beginning in January 2000, NYC mandated that providers and laboratories report all positive hepatitis C tests (chronic and acute cases) to the NYC Department of Health and Mental Hygiene (DOHMH). Surveillance data avoids some of the biases found in population surveys. For example, test results are obtained from all settings, including jails and prisons, drug treatment facilities, and needle exchange programmes, ensuring that there is data from all populations that have been tested. In this report, a methodology is described and results are presented estimating the prevalence of hepatitis C infection in adults aged $\geqslant 20$ years using local surveillance data.

## METHODS

### Methodology

The procedure to estimate prevalence follows the following steps. First, duplicate reports on the same individual are extracted from the surveillance data. Second, subjects who died during the study period are eliminated from the case total as they are no longer considered a prevalent case. Third, the probability that a case may have migrated out of NYC is estimated and is applied to the case total because they are no longer in the at-risk population. Fourth, the infection may have resolved or the results may have been false positive in which case the patient would no longer be a prevalent case and the probability that either of these occurred is applied to the case total. Finally, individuals in the NYC population may not have been diagnosed and we therefore estimated the probability of under-diagnosis and applied this estimate to the case total. The case total is divided by the population of adults aged $\geqslant 20$ years in NYC to give the prevalence estimate. Additional details describing each step are listed below.

### Surveillance data

The NYC Health Code requires healthcare providers and laboratories to report hepatitis C cases for NYC

residents to the DOHMH including positive hepatitis C antibody tests [enzyme immunoassay (EIA) with high signal-to-cutoff ratio or recombinant immuno-blot assay (RIBA)] [24] and positive RNA tests. Providers and laboratories report cases to the health department electronically, by fax, and/or by mail [25]. For this study, all persons aged $\geqslant 20$ years reported to the NYC DOHMH with positive hepatitis C tests from 1 January 2000 to 31 December 2010 were included in the initial surveillance dataset.

### Adjusting for duplicates

Because of the possibility for multiple tests on the same individual which would result in inflated case estimates, the dataset was de-duplicated sequentially using two procedures. First, an automated probabilistic de-duplication algorithm without human review using QualityStage® (IBM Corporation, USA) was implemented. Second, another automated algorithm developed in SAS® v. 9.1 (SAS Institute, USA) evaluated key patient identifiers (e.g. name, date of birth, and address) to de-duplicate the dataset. In this algorithm, the dataset was matched against itself using the patient identifiers resulting in three groups: perfect matches, near matches, and those not matched. Cases that did not match on any criteria were considered unique cases; and thus, retained. Perfect matches represented examples when all patient identifiers matched for two or more reports. In this scenario, the most recent report was kept as a unique case and the remaining reports for this case were discarded as duplicates. The most recent report was selected because the case total was further adjusted for out-migration and the most recent report is the best reflection of the probability of living in NYC. Near matches represented multiple reports that matched on one or more patient identifiers but not all. For these matches, the quality and type of identifiers were grouped and scored from 1 to 10 based on strength of match. Characteristics of the groups were reviewed to determine if each group needed manual review or could be accepted or rejected outright. For example, the misspelling of a surname by a single letter but perfect matching on other patient identifiers would result in a high score and would be accepted without further review. For those groups needing review, trained reviewers evaluated a 10% sample from each and accepted the entire group as matches if >75% of the group were estimated to be true matches.

## Adjusting for death

Death certificates reported to the DOHMH Bureau of Vital Statistics up to 31 December 2010 (all deaths available at the time of the estimate) were matched to records from the surveillance database. Similar to the procedure used for de-duplicating the hepatitis C surveillance database, the automated algorithm evaluated key patient identifiers for the surveillance database and the death certificate database. The manual review of matches used thresholds analogous to the de-duplication methods for hepatitis C cases.

## Adjusting for out-migration from NYC

Tax return data from the U.S. Internal Revenue Service (IRS) from 2000 to 2009 was used to estimate the annual proportion of people who had moved out of NYC during this period [26]. This data was used to estimate the probability that a person had relocated out of NYC between the date of his/her last report and 31 December 2010 (Table 1). The IRS estimates migration by matching returns filed from one year to returns filed in the previous year. Changes in residence at the time of filing determined migration to or from a county. For example, a change in geographical residence between 2009 and 2010 was determined by comparing the location of residence where tax returns were filed in 2009 and 2010 for the tax years in 2008 and 2009, respectively. Using this data, the number of residents who migrated out of the five NYC boroughs (counties) was estimated.

In order to determine the probability that a case in the surveillance database remained in NYC in 2010, the annual probabilities that the person had not moved from NYC since the year of last report were multiplied. For instance, the probability of a case last reported to the database in 2000 still living in NYC in 2010 is derived by multiplying the annual probabilities from 2000–2001 to 2009–2010 (0·6527). Multiplying this probability by the number of cases last reported in 2000 ($n = 296$) provided year-specific estimates of the number of cases still living in NYC in 2010 ($n = 193$). Summing the estimated cases for each year of last report provided the total estimated number of cases living in NYC in 2010 (Table 1) [27].

## Adjusting for percent RNA negative

Some patients reported to the NYC DOHMH with a positive hepatitis C antibody test may not currently be infected with hepatitis C, because they may have

Table 1. *Estimated number of adults aged ⩾20 years reported with hepatitis C in NYC between 1 January 2000 and 31 December 2010 who were still in NYC in 2010*

| Year of last report | Number of unique cases | Probability remained in NYC in 2010* | Estimated number remaining in NYC in 2010 |
|---|---|---|---|
| 2000 | 296† | 0·6527 | 193 |
| 2001 | 1853 | 0·6804 | 1261 |
| 2002 | 3268 | 0·7108 | 2323 |
| 2003 | 3895 | 0·7429 | 2894 |
| 2004 | 4666 | 0·7776 | 3628 |
| 2005 | 5599 | 0·8153 | 4565 |
| 2006 | 8498 | 0·8542 | 7259 |
| 2007 | 10842 | 0·8920 | 9671 |
| 2008 | 13718 | 0·9292 | 12747 |
| 2009 | 18668 | 0·9650 | 18015 |
| 2010 | 38479 | 1·0000‡ | 38479 |
| Total | 109782 | | 101035 |

\* The probability of a case remaining in NYC in 2010 was derived by multiplying the annual probability of out-migrating for each year since the case was last reported to the surveillance system. Annual derived probabilities were as follows: 2009–2010 (0·965); 2008–2009 (0·9629); 2007–2008 (0·96); 2006–2007 (0·9576); 2005–2006 (0·9544); 2004–2005 (0·9538); 2003–2004 (0·9554); 2002–2003 (0·9568); 2001–2002 (0·9572); 2000–2001 (0·9593) [26].
† First year of mandatory testing probably resulted in lower ascertainment. Most positive cases were retested and captured by the system in subsequent years.
‡ The probability of living in NYC in 2010 if a person was reported to the database in 2010 was assumed to be 1.

resolved their infection naturally or with treatment, or the antibody report may have been a false positive. Although an RNA test would confirm infection, many patients do not receive follow-up testing and those that have had a negative follow-up test would not be reported. Because the primary objective was to estimate the prevalence of infection, the estimate was adjusted negatively to account for the potential decrease in number of infections. The literature indicates that 25–30% of patients who are antibody positive are RNA negative either because of a resolved infection or a false-positive result [4, 28]. Thus, this range was multiplied to the estimated number of cases after accounting for deaths and out-migration.

## Adjusting for under-diagnosis

Estimates of the number of patients with hepatitis C who are unaware of their infection status ranged

from 25% to 75% [5, 29, 30]. To account for this, estimated number of cases (after adjusting for deaths and out-migration) were adjusted by a range of values at 5% increments (25–75%), resulting in an increase in the estimated number of hepatitis C infections. Electronic laboratory reporting has greatly improved reporting of diagnosed cases; and therefore, the case total was not adjusted for under-reporting.

### Prevalence estimation

To estimate prevalence, the final adjusted numerator was divided by the population of adults aged $\geqslant 20$ years in NYC in 2010 ($n = 6180263$) [31]. A point estimate and range of values are presented. Note, the range provided reflects a sensitivity analysis of % RNA negative and under-diagnosis possible values and is not a statistically derived confidence interval. All analyses were performed using the R statistical program v. 2·14·0 (R Foundation for Statistical Computing, Austria).

## RESULTS

### Adjusting for duplicates

The NYC DOHMH received 535645 reports of hepatitis C in adults aged $\geqslant 20$ years between 1 January 2000 and 31 December 2010 (Fig. 1). De-duplication through the initial automated system using Quality Stage reduced that number to 132113 hepatitis C cases. Additional de-duplication methods using SAS removed 5588 duplicates (4·2%) resulting in a total of 126525 unique persons.

### Death match and out-migration

A total of 16743 (13·2%) patients matched death certificates in the NYC DOHMH death registry with a date of death on or before 31 December 2010 resulting in 109782 persons remaining. To assess out-migration, IRS data was used to estimate that 8747 (8%) persons out-migrated leaving an estimated 101035 persons aged $\geqslant 20$ years with hepatitis C living in NYC as of 2010 (Table 1).

### Resolution and under-diagnosed infections

Because of the uncertainty in the proportion of patients who are RNA negative and those who are aware of their positive status, a range of values were considered. The median value of the range for the

proportion of patients who are RNA negative was 27·5% (25–30%) which resulted in 73250 cases. The median value of the range for the proportion who are aware of their status was 50% (25–75%); thus, the estimate is doubled resulting in 146500 chronic hepatitis C cases corresponding with 2·37% of the adults aged $\geqslant 20$ years in NYC ($n = 6180263$). The minimum and maximum values of these ranges result in 94299 chronic HCV cases (1·53%; proportion with negative RNA status = 30%, proportion aware of hepatitis status = 75%) and 303104 chronic HCV cases (4·90%; proportion with negative RNA status = 25%, proportion aware of hepatitis status = 25%).

## DISCUSSION

The estimated prevalence of hepatitis C in NYC in adults aged $\geqslant 20$ years was 2·37% (146500 cases) in 2010. Because these assumptions increased the uncertainty of this estimate, a range of estimates is provided (94299–303104). This range is substantially influenced by the number of patients who have undiagnosed infections. Some studies have estimated that the number of patients with undiagnosed infections is as high as 75%. In that case, the estimated prevalence would be 4·9% or more than 300000 people.

The surveillance-based prevalence estimate of 2·37% is higher than the prevalence estimated from the 2004 NYC HANES study that included civilian non-institutionalized housed adults aged $\geqslant 20$ years [5]. However, if the surveillance estimate was not adjusted by out-migration, resolution, and under-diagnosis, the number of living cases that are antibody positive in NYC would have been 109782 and almost 37000 cases lower than the adjusted estimate but more similar to the absolute number of antibody-positive cases in the NYC HANES survey ($n = 129000$). Given the known methodological limitations of the NYC HANES survey, adjustment for migration, resolution of infection, and under-diagnosis gives greater confidence to our results.

Estimates based on local data are important for public health practitioners and policy makers to understand the burden of disease in their area and prioritize resource allocation. The largest cohort of people currently living with hepatitis C are those aged 40–59 years [5, 25]. People in this age group were probably infected many years ago and are reaching an age in which chronic hepatitis C infection may result in severe liver disease and hepatic carcinoma. With two drugs newly approved in 2011
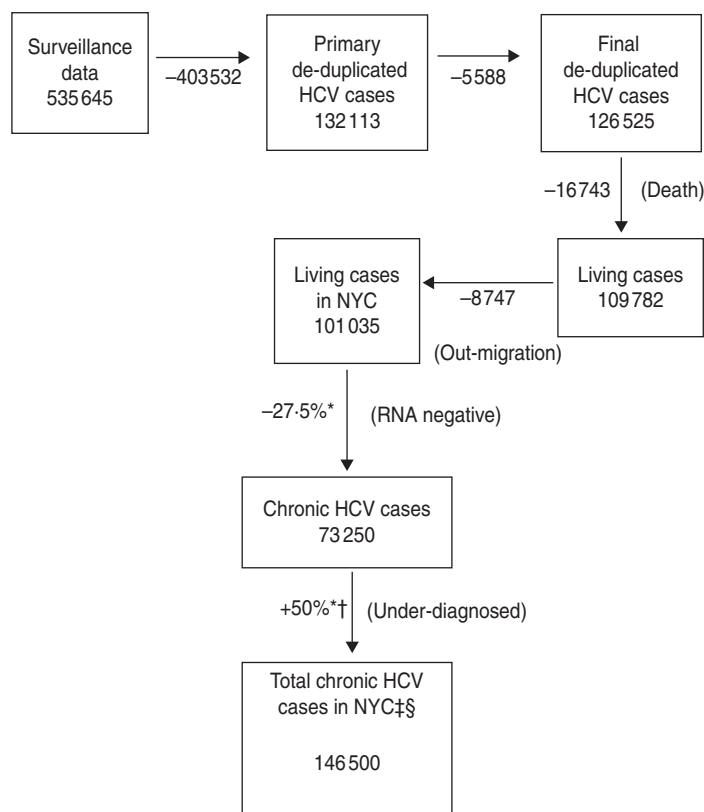
**Fig. 1.** Steps used to estimate the number of adults aged ⩾20 years with hepatitis C infection, NYC, 2010. * Mid-point of range. † 50% unaware of their status results in doubling of the estimate. ‡ Estimated median prevalence based on range of predicted values. § Adults aged ⩾20 years.

for hepatitis C treatment [20, 21], more patients may be referred for treatment increasing the demand for resources for care. Thus, prevalence estimates are important for local jurisdictions planning to allocate resources for treatment. Currently, NYC DOHMH is piloting a project to connect patients with hepatitis C infection to care and may consider expanding the project depending on the results [32]. Additionally, when the Affordable Care Act is implemented, many more patients will have insurance to cover testing and treatment [33]. Thus, while it may be possible for public health programmes to have an impact on the prevalence of hepatitis C, planning for these programmes requires reliable prevalence estimates.

There are several limitations to this analysis. First, because of the uncertainty surrounding the proportion of the population who have never been tested for hepatitis C and are therefore unaware of their diagnosis or do not have a positive RNA test, the true prevalence of chronic hepatitis C in NYC might be higher or lower than what was estimated. To account for this uncertainty, a range of prevalence estimates was

provided using the best available data from the literature [4, 5, 28, 29]. Better estimates for rates of undiagnosed patients would improve our estimate and if they become available the estimate can be adjusted. Under-diagnosis may be greater in some risk groups than others, but without specific data that quantifies these differences it is not possible to adjust the current estimate in a meaningful way. Second, the case estimates listed in Table 1 do not reflect annual incidence but the case with the most recent positive report. Using the most recent positive report enables a more accurate assessment of the number of cases who migrate out of the city. For example, a case reported in 2009 is more informative than a case reported in 2000 because for the 2009 report the case was still living in NYC and thus had a lower probability of out-migration. As a result, the annual case totals presented in this report should neither be viewed as incidence of hepatitis C nor a reflection of the current epidemic. Third, this prevalence estimate excluded the population aged <20 years in order to make this study more comparable with local and national estimates.

Less than 1% of this population is hepatitis C positive; and as a result, excluding this population is not likely to significantly change the absolute total. A total of 1314 cases in adolescents aged <20 years were reported to the surveillance system between 2000 and 2010. When including these adolescents into the study population, the prevalence estimate is 148188 (1·81%). Moreover, despite the predominance of prevalence by certain age groups according to multiple reports, the estimates reported here were not age-stratified. Implementing age-stratification requires having age-specific probabilities for each of the adjustments. Given the limited information currently available for the proportion aware of their status and the proportion that are RNA negative, stratifying the results by age but failing to account for the appropriate probabilities is likely to introduce additional bias into the results.

County-to-county migration data from the IRS only captures migration in those who file yearly federal tax returns and may not accurately represent NYC residents with hepatitis C. Those who are less likely to file federal income tax returns (e.g. the poor, injecting drug users, and undocumented immigrants) may be underrepresented in this data source. Moreover, this estimate did not account for in-migration of patients with hepatitis C who may not have been tested after arrival in NYC. Additionally, our registry only reliably went back as far as 2000. Individuals who were diagnosed prior to 2000 may not be included in the registry; thus, leading to an underestimate of cases. Because most patients who have hepatitis C are tested many times, the number of arrivals with hepatitis C who were never tested in NYC is likely to be small. In particular, intravenous drug users (IDUs) who are disproportionately affected by hepatitis C may be underrepresented in any counts of migration because they may not be filing federal tax returns.

Despite these limitations, this study demonstrates how chronic hepatitis C surveillance data can be used to estimate the prevalence of hepatitis C. Surveillance for hepatitis C can be resource intensive because of the large number of reports and de-duplication challenges. De-duplication programmes used for this surveillance system were not perfect and not every case could be individually reviewed. Cases may have been misclassified as duplicates depending on the amount of available data and how common the patient names were. The sheer volume of hepatitis C data will make de-duplication an ongoing challenge for any public health agency trying to estimate prevalence. Additionally, because of the chronic nature of the disease and under-diagnosis, there is a need to make adjustments for death, out-migration and under-diagnosis in order to estimate local disease prevalence. However, the burden of hepatitis C on the healthcare system is increasing as the largest infected cohort is beginning to reach the need for treatment. Even though mortality from hepatitis C infection now exceeds HIV nationally, surveillance resources for HIV are much greater than for hepatitis C [3]. To our knowledge this is the first time a local jurisdiction has used surveillance data to make local prevalence estimates. While other prevalence estimation techniques exist, they are technical and require significantly more data [34, 35]. As a result, surveillance estimates are important because they can be relatively easy to derive and used for local planning purposes. Using surveillance data also overcomes the lack of exclusion of homeless and incarcerated populations in most population serosurveys because repeat testing captures hard to reach individuals. Prevalence estimates from surveillance data can be made more frequently and kept up to date. Moreover, although it may not be possible to investigate every hepatitis C case reported to the surveillance system, opportunities exist to sample the data for additional in-depth investigation as a mechanism to better characterize the population [36].

## CONCLUSION

About 146500 adults aged ⩾20 years in NYC are living with hepatitis C infection (range 94299–303104 persons). Additional resources will be needed to identify and treat chronic hepatitis C. Local data will be helpful for the health department, policy makers, and healthcare providers in planning for hepatitis C care and treatment programmes.

## ACKNOWLEDGEMENTS

## DECLARATION OF INTEREST

None.

# REFERENCES

1. **CDC.** Hepatitis C Information for Health Professionals, 2012 (http://www.cdc.gov/hepatitis/HCV/index.htm). Accessed 19 March 2012.

2. **Wong T, Lee SS.** Hepatitis C: a review for primary care physicians. *Canadian Medical Association Journal* 2006; **174**: 649–659.

3. **Ly K N, et al.** The increasing burden of mortality from viral hepatitis in the United States between 1999 and 2007. *Annals of Internal Medicine* 2012; **156**: 271–278.

4. **Armstrong GL, et al.** The prevalence of hepatitis C virus infection in the United States, 1999 through 2002. *Annals of Internal Medicine* 2006; **144**: 705–714.

5. **Bornschlegel K, et al.** Prevalence of hepatitis C infection in New York City, 2004. *Journal of Urban Health* 2009; **86**: 909–917.

6. **Riley ED, et al.** Antiretroviral therapy, hepatitis C virus, and AIDS mortality among San Francisco's homeless and marginally housed. *Journal of Acquired Immune Deficiency Syndrome* 2005; **38**: 191–195.

7. **Schwarz KB, et al.** Seroprevalence of HCV infection in homeless Baltimore families. *Journal Health Carefor the Poor and Underserved* 2008; **19**: 580–587.

8. **Stein JA, Nyamathi A.** Correlates of hepatitis C virus infection in homeless men: a latent variable approach. *Drug and Alcohol Dependence* 2004; **75**: 89–95.

9. **Desai RA, Rosenheck RA, Agnello V.** Prevalence of hepatitis C virus infection in a sample of homeless veterans. *Social Psychiatry and Psychiatric Epidemiology* 2003; **38**: 396–401.

10. **Klinkenberg WD, et al.** Prevalence of human immunodeficiency virus, hepatitis B, and hepatitis C among homeless persons with co-occurring severe mental illness and substance use disorders. *Comparative Psychiatry* 2003; **44**: 293–302.

11. **Cheung RC, et al.** Viral hepatitis and other infectious diseases in a homeless population. *Journal of Clinical Gastroenterology* 2002; **34**: 476–480.

12. **Fox RK, et al.** Hepatitis C virus infection among prisoners in the California state correctional system. *Clinical Infectious Disease* 2005; **41**: 177–186.

13. **Solomon L, et al.** Prevalence of HIV, syphilis, hepatitis B, and hepatitis C among entrants to Maryland correctional facilities. *Journal of Urban Health* 2004; **81**: 25–37.

14. **Spaulding A, et al.** Hepatitis C in state correctional facilities. *Preventive Medicine* 1999; **28**: 92–100.

15. **Vlahov D, et al.** Prevalence and incidence of hepatitis C virus infection among male prison inmates in Maryland. *European Journal of Epidemiology* 1993; **9**: 566–569.

16. **Chak E, et al.** Hepatitis C virus infection in USA: an estimate of true prevalence. *Liver International* 2011; **31**: 1090–1101.

17. **Williams IT, et al.** Incidence and transmission patterns of acute hepatitis C in the United States, 1982–2006. *Archives of Internal Medicine*, 2011; **171**: 242–248.

18. **Yalamanchili K, et al.** The prevalence of hepatitis C virus infection in Texas: implications for future health care. *Baylor University Medical Center Proceedings* 2005; **18**: 3–6.

19. **Amon JJ, et al.** Prevalence of hepatitis C virus infection among injection drug users in the United States, 1994–2004. *Clinical Infectious Disease* 2008; **46**: 1852–1858.

20. **Sherman KE, et al.** Response-guided telaprevir combination treatment for hepatitis C virus infection. *New England Journal of Medicine* 2011; **365**: 1014–1024.

21. **Lok AS, et al.** Preliminary study of two antiviral agents for hepatitis C genotype 1. *New England Journal of Medicine* 2012; **366**: 216–224.

22. **Maasoumy B, Manns MP.** Optimal treatment with boceprevir for chronic HCV infection. *Liver International* 2013; **33** (Suppl. 1): 14–22.

23. **Asselah T, Marcellin P.** Interferon free therapy with direct acting antivirals for HCV. *Liver International* 2013; **33** (Suppl. 1): 93–104.

24. **CDC.** Signal to cut-off ratios for commercially available assays, 2012 (http://www.cdc.gov/hepatitis/HCV/LabTesting.htm). Accessed 10 May 2012.

25. **NYC Department of Health and Mental Hygiene.** Hepatitis A, B and C surveillance report (http://www.nyc.gov/html/doh/downloads/pdf/cd/cd-hepabc-surveillance-report-08–09.pdf). Accessed 9 September 2012.

26. **Washington, DC: IRS.** County to county migration data: 2004–2008. (http://www.irs.gov/uac/SOI-Tax-Stats-County-to-County-Migration-Data-Files). Accessed 22 April 2013.

27. **France AM, et al.** Estimating the prevalence of chronic hepatitis B virus infection-new york city, 2008. *Journal of Urban Health* 2012; **89**: 373–383.

28. **Jauncey M, et al.** Clearance of hepatitis C virus after newly acquired infection in injection drug users. *Journal of Infectious Disease* 2004; **190**: 1270–1274.

29. **Colvin HM, Mitchell AE (eds).** *Hepatitis and Liver Cancer: a National Stratgey for Prevention and Control of Hepatitis B and C.* Washington, DC: Institute of Medicine of the National Academies National Academies Press, 2010.

30. **Denniston MM, et al.** Awareness of infection, knowledge of hepatitis C, and medical follow-up among individuals testing positive for hepatitis C: NHANES 2001–08. *Hepatology* 2011; **55**: 1652–1661.

31. **U.S. Census Bureau.** Census 2010 (http://factfinder2.census.gov/legacy/aff_sunset.html?_lang&en). Accessed 29 November 2011.

32. **NYC Hepatitis B and Hepatitis C community site.** (http://nychepbc.org/checkhepc/). Accessed 19 November 2012.

33. **U.S. Department of Health & Human Services.** The health care law and you (http://www.healthcare.gov/law/index.html), 2012. Accessed 19 November 2012.

34. **De Angelis D,** *et al.* An evidence synthesis approach to estimating hepatitis C prevalence in England and Wales. *Statistical Methods in Medical Research*, 2009; **18**: 361–379.

35. **Deuffic-Burban S, Mathurin P, Valleron AJ.** Modelling the past, current and future HCV burden in France: detailed analysis and perspectives. *Statistical Methods in Medical Research* 2009; **18**: 233–252.

36. **CDC.** Surveillance for chronic hepatitis B virus infection – New York City, June 2008-November 2009. *Morbidity and Mortality Weekly Report* 2012; **61**: 6–9.