

# Use of the International Classification of Diseases, 9th revision, coding in identifying chronic hepatitis B virus infection in health system data: implications for national surveillance

Reena Mahajan,<sup>1</sup> Anne C Moorman,<sup>1</sup> Stephen J Liu,<sup>1</sup> Lorelee Rupp,<sup>2</sup>  
R Monina Klevens,<sup>1</sup> for the Chronic Hepatitis Cohort Study (CHeCS) investigators\*

<sup>1</sup>Division of Viral Hepatitis, Centers for Disease Control and Prevention, Atlanta, Georgia, USA

<sup>2</sup>Center for Health Policy and Health Services Research, Henry Ford Health System, Detroit, Michigan, USA

## Correspondence to

Dr Reena Mahajan, Division of Viral Hepatitis, National Center for HIV, Viral Hepatitis, STD, and TB Prevention, Centers for Disease Control and Prevention, Mailstop G37, 1600 Clifton Road NE, Atlanta, GA 30333, USA; vif5@cdc.gov

\*For a list of the CHeCS investigators see the end of this paper.

Received 10 December 2012  
Revised 23 January 2013  
Accepted 11 February 2013  
Published Online First  
5 March 2013

## ABSTRACT

**Objective** With increasing use electronic health records (EHR) in the USA, we looked at the predictive values of the International Classification of Diseases, 9th revision (ICD-9) coding system for surveillance of chronic hepatitis B virus (HBV) infection.

**Materials and Methods** The chronic HBV cohort from the Chronic Hepatitis Cohort Study was created based on electronic health records (EHR) of adult patients who accessed services from 2006 to 2008 from four healthcare systems in the USA. Using the gold standard of abstractor review to confirm HBV cases, we calculated the sensitivity, specificity, positive and negative predictive values using one qualifying ICD-9 code versus using two qualifying ICD-9 codes separated by 6 months or greater.

**Results** Of 1 652 055 adult patients, 2202 (0.1%) were confirmed as having chronic HBV. Use of one ICD-9 code had a sensitivity of 83.9%, positive predictive value of 61.0%, and specificity and negative predictive values greater than 99%. Use of two hepatitis B-specific ICD-9 codes resulted in a sensitivity of 58.4% and a positive predictive value of 89.9%.

**Discussion** Use of one or two hepatitis B ICD-9 codes can identify cases with chronic HBV infection with varying sensitivity and positive predictive values.

**Conclusions** As the USA increases the use of EHR, surveillance using ICD-9 codes may be reliable to determine the burden of chronic HBV infection and would be useful to improve reporting by state and local health departments.

## BACKGROUND AND SIGNIFICANCE

In the USA, 800 000–1.4 million people are chronically infected with hepatitis B virus (HBV); these persons are at increased risk of chronic liver disease and its sequelae.<sup>1–2</sup> Current national viral hepatitis surveillance is a mostly passive laboratory-initiated reporting system to state or local health departments. Clinicians and healthcare facilities are also mandated to report chronic HBV, but there is considerable variability by states. As of 2009, 38 state and local health departments report chronic HBV infection in the National Notifiable Disease Surveillance System (NNDSS).<sup>3</sup> National-level surveillance using the National Health and Nutrition Examination Survey (NHANES), estimates chronic HBV infection at 0.28%; however, this nationally representative random survey of approximately 5000 US residents per year, excludes incarcerated

and homeless individuals, which is one of its limitations.<sup>2</sup> Given this limitation, a derived estimate of the national prevalence of chronic HBV infection has shown a range of 0.3–0.5% based on inclusion of the foreign-born, institutionalized, and the US civilian population.<sup>4</sup>

Active surveillance of chronic hepatitis B by health departments can be expensive and labor intensive. As the USA moves towards the use of electronic health records (EHR), use of the International Classification of Diseases, 9th revision (ICD-9) for surveillance has been proposed. Currently, 48% of hospitals and 20% of physicians have implemented EHR.<sup>5</sup> Use of the ICD-9 coding system has been used in healthcare facilities for reimbursement as well as research purposes, but its usefulness has varied depending on the health condition being evaluated and the reimbursement rate.<sup>6–11</sup> With the increasing availability of EHR data, understanding which diseases can use the ICD-9 code for surveillance will be critical.

Health conditions that require input of multiple ICD-9 codes have been shown to have lower sensitivity when compared to conditions that utilize a single diagnosis.<sup>12</sup> For example, use of ICD-9 codes for evaluation of sepsis may better identify the clinical syndrome when multiple codes are included compared to the use of one code.<sup>13</sup> For infectious diseases, such as *Clostridium difficile*, studies suggest that the ICD-9 code has good sensitivity and specificity and may be able to be used for surveillance when compared to the toxin assay.<sup>14–15</sup> Other infectious diseases such as salmonellosis, shigellosis, and pertussis, which have a simple case definition and only require input of one code, have higher positive predictive values when using the ICD-9 codes.<sup>10</sup> Chronic HBV infection may be similar to these infectious diseases, as it does not require multiple ICD-9 codes to identify the infection. Previous data regarding surveillance of hepatitis B have been limited to acute HBV infection or a sample of those with hepatitis C infection or alcoholic liver disease with varying levels of accuracy.<sup>7–16</sup> Although costly, previous researchers have suggested the use of the ICD-9 coding system with other clinical or laboratory data to improve its utility as a surveillance tool.<sup>17–19</sup>

The Centers for Disease Control and Prevention Chronic Hepatitis Cohort Study (CHeCS) is a prospective cohort study with information on over 1.6 million patient records within four large diverse

**To cite:** Mahajan R, Moorman AC, Liu SJ, et al. *J Am Med Inform Assoc* 2013;**20**:441–445.

health systems: Geisinger Health System (GHS), Danville, PA; Henry Ford Health System (HFHS), Detroit, MI; Kaiser Permanente-Northwest, Portland, OR; and Kaiser Permanente-Honolulu, Hawaii. We used data from this study to assess chronic HBV infection. Inclusion criteria required a hepatitis B ICD-9 code or supportive laboratory data with case confirmation by manual abstractor review of the medical record.

Our goal was to measure the performance of the ICD-9 codes for surveillance of chronic HBV infection. We examined the sensitivity, specificity, and positive and negative predictive values for using one or more hepatitis B ICD-9 codes from these four sites and confirmed all cases of chronic HBV infection through abstractor review.

## MATERIALS AND METHODS

### Cohort enrollment

The methods for CHeCS have been summarized in a previous report.<sup>20</sup> Briefly, the initial cohort was created based on electronic and medical billing EHR of patients aged 18 years or older who had a service provided between 1 January 2006 and 31 December 2008 at one of four sites: GHS, Danville, PA; HFHS, Detroit, MI; Kaiser Permanente-Northwest, Portland, OR; and Kaiser Permanente-Honolulu, Hawaii. For this analysis, electronic medical record and billing data were collected for each patient and supplemented with individual chart review. Data collected included patient demographics, medical encounters, and laboratory results.

The catchment area for each of the four health systems, selected for their representation of minority populations, is fairly comprehensive. GHS provides healthcare services to approximately half the residents for the 44 counties in Pennsylvania that it serves.<sup>21</sup> HFHS serves more than one million southeast Michigan residents and is one of the three largest providers of healthcare services in southeastern Michigan. Kaiser Permanente-Honolulu, Hawaii, is the health plan for one-sixth of Hawaiian residents.<sup>22, 23</sup> Finally, Kaiser Permanente-Northwest serves over 476 000 members and represents approximately 17% of the area's population. Each of the four health systems is one of if not the main source of care for its residents.

Algorithms for inclusion in the chronic HBV cohort were developed and applied to the EHR data of patients. The goal was to capture the greatest number of verifiable chronic HBV cases from the raw observational data while excluding those with a single unconfirmed diagnosis or laboratory evidence that might be due to acute disease, error, or lack of necessary work-up. Complete observation time for each patient was determined as the date of the first indication of hepatitis infection in the EHR including retrospective data before 1 January 2006, until the date of either the last health system encounter or 31 December 2008. Electronic data from 2006–8 were reviewed to determine enrollment candidacy, and data from candidate patients were reviewed from their earliest health system encounter to 2008 to determine cohort eligibility.

Patients were included in the chronic HBV cohort based on fulfillment of a combination of laboratory and ICD-9-based criteria. Qualifying hepatitis B ICD-9 codes included: chronic (070.22, 070.23, 070.32, 070.33) or acute/unspecified (070.2, 070.20, 070.21, 070.3, 070.30, 070.31). Qualifying positive laboratory tests included: hepatitis B surface antigen, hepatitis B e-antigen, or hepatitis B DNA. For inclusion in the cohort, patients had to fulfill the following criteria: two positive laboratory tests consistent with HBV infection; a positive laboratory test and an ICD-9 diagnosis code; or two ICD-9 diagnosis

codes obtained at least 6 months apart. Patients in all phases of chronic HBV infection (immune-tolerant, immune-active, and inactive or 'immune carrier') were included.

Each of the four sites utilized five to 12 research abstractors who were either registered nurses or registered health information technologists for abstractor review. All of the abstractors received five classroom hours of instruction regarding the study protocol, utilized a detailed abstraction manual and standardized web-based data collection forms provided by the data coordinating center (DCC), and attended biweekly meetings led by the DCC to discuss quality control issues. Regular quality assurance reports that check for invalid, inconsistent, or missing content were generated and distributed by the DCC, and queries remained outstanding until resolved. The data manager from each site did a 5% double review, in which the data collected independently by two abstractors was compared for accuracy, and discussed findings with the individual abstractor at the biweekly meetings. Interrater reliability statistics were not calculated at each site.

During this process, abstractors flagged charts for case review that lacked sufficient documentation that the patient had been diagnosed with chronic hepatitis B or had documentation that the patient had been diagnosed with acute hepatitis or that chronic hepatitis had been ruled out. Flagged charts were subsequently reviewed by the project coordinator and/or a clinician for case confirmation, using detailed hepatologist-developed criteria provided by the coordinating center. Flagged cases for which chronic viral hepatitis infection could not be confirmed were excluded from this study.

Demographic and clinical data regarding patients who had only one ICD-9 code but did not meet inclusion criteria were not available for analysis.

### Statistical analysis

We examined data by age, gender, race, and site for cases with and without an ICD-9 code. For the purpose of this analysis, true positives are those cases that were included in the chronic hepatitis B cohort after final abstractor review. For all cases in the HBV cohort, we calculated sensitivity, specificity, positive and negative predictive values using one qualifying ICD-9 code versus using two qualifying ICD-9 codes separated by 6 months or greater. We also examined the frequencies of acute and chronic ICD-9 codes for all confirmed cases. The Levene test was used to evaluate homogeneity of variance between the groups with and without an ICD-9 code using SAS software, V9.3.

## RESULTS

Of the 1 652 055 adult patients in the four participating health systems who had one or more services provided between 1 January 2006 and 31 December 2008, 3029 (0.02%) had a qualifying primary or secondary ICD-9 diagnosis. Of the 3029 patients with at least one ICD-9 code, 1847 were true positives and 1182 were false positives, resulting in a positive predictive value equal to 61.0% (1847/3029) (table 1). The specificity, ie, the ability of one ICD-9 to exclude those who were not HBV cases, was over 99.9% (1 648 671/1 649 853), and the ability of one ICD-9 code designation to identify correctly those who did not have HBV, the negative predictive value, was almost 100% (1 648 671/1 649 026).

A total of 2202 (72.7%) met one of the inclusion criteria and were included in the cohort after abstractor review of the medical record and confirmation as true chronic cases. Of the 2202 confirmed cases, 1847 (83.9%) had at least one ICD-9

**Table 1** Measurement of sensitivity, specificity, positive and negative predictive values of using one hepatitis B-specific ICD-9 among persons receiving services from four healthcare systems from 2006 to 2008

	Confirmed HBV case	Not a HBV case	Total
One ICD-9 code	1847	1182	3029
No ICD-9 code	355	1 648 671	1 649 026
Total	2202	1 649 853	1 652 055

Sensitivity 1847/2202 (83.9%).  
 Specificity 1 648 671/1 649 853 (99.9%).  
 Positive predictive value 1847/3029 (61.0%).  
 Negative predictive value 1 648 671/1 649 026 ( $\geq 99.9\%$ ).  
 HBV, hepatitis B virus; ICD-9, International Classification of Diseases, 9th revision.

code, and 355 (16.1%) had no ICD-9 code during the study period, resulting in a sensitivity of 83.9% (1847/2202) (table 1). Approximately 39% of all codes were acute, although after abstractor review, were determined to be true chronic cases.

Of the 2202 confirmed cases, over half were aged 30–50 years and over 30% were Asian. Using the Levene test for equality of variance, there was significant heterogeneity of variance between those with and without an ICD-9 code ( $p < 0.01$ ) among patients who were aged 29 years or younger (table 2). For all confirmed cases, the mean length of time since diagnosis was 4.6 years (range 2 days to 14 years). Sensitivity

**Table 2** Demographic characteristics of HBV cohort cases with and without one ICD-9 code from four healthcare systems, 2006–8

	Cohort cases with $\geq 1$ ICD-9 code N=1847 (%)	Cohort cases with no ICD-9 code N=355 (%)
Age group (years)		
<20	143 (8)*	46 (13)*
20–29	338 (18)*	92 (26)*
30–39	462 (25)	83 (23)
40–49	497 (27)	73 (21)
50–59	276 (15)	41 (12)
60–69	94 (5)	13 (4)
70–79	37 (2)	7 (2)
Gender		
Male	1080 (58)	168 (47)
Female	767 (42)	187 (53)
Race		
White, non-Hispanic	430 (23)	70 (20)
Black, non-Hispanic	178 (10)	59 (17)
Asian, non-Hispanic	718 (39)	119 (34)
Native American	12 (1)	1 (<1)
Hawaiian/Pacific Islander	160 (9)	32 (9)
Unknown	316 (17)	68 (19)
Site		
Kaiser Permanente-Northwest	714 (39)	62 (17)
Kaiser Permanente-Honolulu, Hawaii	511 (28)	142 (40)
Henry Ford Health System, Detroit, Michigan	497 (27)	109 (31)
Geisinger Health System, Danville, Pennsylvania	125 (7)	42 (12)

\* $p < 0.05$ , Levene test for equality of variance.  
 HBV, hepatitis B virus; ICD-9, International Classification of Diseases, 9th revision.

varied slightly when the four geographically diverse healthcare systems were compared (range 75–92%).

The sensitivity, specificity, positive predictive value, and negative predictive value of the use of two ICD-9 codes separated by 6 months to identify HBV cases was also assessed. Of the 1432 individuals with two qualifying ICD-9 codes, 1287 (89.9%) were true positives, and 145 (10.1%) were not considered positive, resulting in a positive predictive value equal to 89.9% (1287/1432). The specificity was over 99.9% (1 649 708/1 649 853), and the negative predictive value was 99.9% (1 649 708/1 650 623). For the 1287 true positives with two hepatitis B-specific ICD-9 codes separated by 6 months, sensitivity was 58.4% (1287/2202) (table 3).

## DISCUSSION

Our data indicate that among these four large integrated healthcare organizations, the use of one hepatitis B-specific ICD-9 code was reliable for predicting the proportion of individuals with and without chronic HBV; sensitivity was 84% and specificity was greater than 99%. However, its ability to categorize correctly those who are labeled as positive (positive predictive value) was only 61%. In comparison, use of two ICD-9 codes separated by 6 months decreased the likelihood of capturing all positives (sensitivity) to 58%, but did increase the likelihood of accurately labeling those who were true positives (positive predictive value) to 90%. For surveillance of chronic HBV infection in the USA, with use of EHR, this represents a less labor-intensive while still reliable measure for health departments to capture the burden of disease compared to laboratory-based disease reporting.

There were a number of limitations to this study. The overall prevalence of chronic hepatitis B in our cohort was 0.1%, which is lower than previous derived estimates of 0.3–0.5%;<sup>4</sup> this difference may reflect the proportion of infected patients who have not yet been tested, diagnosed, and engaged in healthcare. Recent estimates from the CHECS study population suggest that 21% of probable HBV have not been tested and diagnosed, which is a limitation of any surveillance system, whether ICD-9 or laboratory based.<sup>24</sup> In particular, individuals at two of the four sites (Kaiser Permanente-Northwest and Kaiser Permanente-Honolulu, Hawaii) had to be members of those healthcare systems, thus representing individuals with access to care and health insurance.<sup>20</sup> Although it is possible

**Table 3** Measurement of sensitivity, specificity, and predictive values of using two hepatitis B-specific ICD-9 codes separated by 6 months among persons from four healthcare systems from 2006 to 2008

	Confirmed chronic cohort case	Not a chronic cohort case	Total
Two hepatitis B ICD-9 codes separated by 6 months	1287	145	1432
Does not have two hepatitis B ICD-9 codes separated by 6 months	915	1 649 708	1 650 623
Total	2202	1 649 853	1 652 055

Sensitivity 1287/2202 (58.4%).  
 Specificity 1 649 708/1 649 853 ( $\geq 99.9\%$ ).  
 Positive predictive value 1287/1432 (89.9%).  
 Negative predictive value 1 649 708/1 650 623 (99.9%).  
 ICD-9, International Classification of Diseases, 9th revision.

that the 1.6 million individuals who accessed services from these four healthcare systems may not reflect the general US population, they were selected to be geographically diverse, and Asian and Hawaiian/Pacific Islanders were over-sampled as they are a sizeable proportion of chronic cases of HBV infection in the USA.

In addition, there was heterogeneity of variance of younger cohort individuals with and without an ICD-9 code; younger persons with HBV infection may not be considered to be chronically infected and thus may reflect varied provider practice in documentation. In our cohort, 0.1% of individuals in the populations served by these four healthcare organizations ultimately met the inclusion criteria for having chronic HBV. As noted previously, NHANES has been used for surveillance estimates.<sup>2</sup> While NHANES may capture the burden of disease within the population because it accounts for individuals who are not tested, our estimates are useful as they describe how well the ICD-9 code captures individuals with chronic HBV infection who have been tested and accessed healthcare services. Our study is also unique because we have data on the entire cohort, which adds validity to our analysis.

Concern about the use of the ICD-9 coding system for research purposes, particularly with data accuracy and completeness of coding, has been well documented.<sup>11 17 25</sup> Previous research has shown that up to 40% of cases of reported acute hepatitis B may represent chronic infection and that the presence of an ICD-9 code for acute hepatitis B is unlikely to indicate acute infection.<sup>16 26</sup> From our study, we demonstrated that ICD-9 codes for acute hepatitis B can be used for chronic disease even if the reverse is not true,<sup>25 26</sup> because reviewers abstracted the medical records of all potential cases to determine the true chronic HBV cohort cases.

## CONCLUSION

Using EHR, we conducted the gold standard—including the use of laboratory data, ICD-9 codes and abstractor review—to determine true chronic HBV cases.<sup>17–19</sup> The high sensitivity and moderate positive predictive value of the use of one ICD-9 code to identify chronic hepatitis B will be useful for conducting surveillance particularly with increasing use of EHR data. As the USA moves towards EHR, knowing whether ICD-9 codes can be used for the disease of interest, such as chronic HBV, will be essential for health departments and policymakers to know when allocating resources in their data collection efforts.

An area of considerable interest is the use of EHR and the future ICD-10 codes. There will be almost five times the ICD-10 codes with greater specificity and detail useful for public health and future surveillance efforts.<sup>27</sup> With this more complicated coding system, flexibility and training will be required for medical personnel who code the data. Our data would not be likely to change as chronic HBV coding does not have the additional detail or use advanced medical technologies for its diagnosis; however, as a quality improvement measure, it would be valuable to track data before and after implementation of the ICD-10 system. Currently, the USA is the only industrialized country not using the ICD-10 system.<sup>28</sup> As chronic HBV is a global public health issue, transition to the ICD-10 will allow us to generate surveillance data domestically and compare it to information internationally.

**Acknowledgements** The authors would like to thank Fujie Xu, MD, PhD, Division of Viral Hepatitis, Centers for Disease Control and Prevention for her helpful suggestions and editing of this manuscript.

The CHECS Investigators include the following investigators and sites: Scott D Holmberg, Eyasu HTeshale, Philip R Spradling, Anne C Moorman, Division of Viral Hepatitis, National Center for HIV, Viral Hepatitis, STD, and TB Prevention, Centers for Disease Control and Prevention, Atlanta, Georgia; Stuart C Gordon, David R Nerenz, Mei Lu, Lois Lamerato, Lorelee B Rupp, Nonna Akkerman, Nancy Oja-Tebbe, Chad M Cogan, and Dana Larkin, Henry Ford Health System, Detroit, Michigan; Joseph A Boscarino, Zahra S Daar, Joe B Leader, and Robert E Smith, Geisinger Health System, Danville, Pennsylvania; Cynthia C Nakasato, Vinutha Vijayadeva, Kelly E Sylva, John V Parker, and Mark M Schmidt, Kaiser Permanente-Hawaii, Honolulu, Hawaii; Emily M Henkle, Tracy L Dodge, and Erin M Keast, Kaiser Permanente- Northwest, Portland, OR.

**Funding** The Chronic Hepatitis Cohort Study (CHECS) was funded by the CDC Foundation, which received grants from Abbott Laboratories, Genentech, a member of the Roche Group, Janssen Pharmaceutical Companies of Johnson & Johnson and Vertex Pharmaceuticals. Granting corporations do not have access to CHECS data and do not contribute to data analysis or writing of manuscripts.

**Competing interests** None.

**Ethics approval** This study received ethics approval from the United States Department of Health and Human Services.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data sharing statement** The CHECS investigators at CDC and the four participating health systems are granted access to analyze limited datasets extracted from EHR and administrative data from these four participating health systems, as part of data use agreements executed under the CDC Foundation-funded Chronic Hepatitis Cohort Study (CHECS), described in detail here <http://cid.oxfordjournals.org/content/early/2012/09/17/cid.cis815.short?rs=1>. Any further sharing of de-identified datasets would be subject to review and approval by the health system institutional review boards and is not currently part of the study protocol.

## REFERENCES

- Centers for Disease Control and Prevention. Viral hepatitis statistics and surveillance. 2010. <http://www.cdc.gov/hepatitis/Statistics/index.htm> (accessed 20 Jul 2012).
- Wasley A, Kruszon-Moran D, Kuhnert W, et al. The prevalence of hepatitis B virus infection in the United States in the era of vaccination. *J Infect Dis* 2010;202:192–201.
- Centers for Disease Control and Prevention. Viral Hepatitis Surveillance-United States, 2009. <http://www.cdc.gov/hepatitis/Statistics/2009Surveillance/index.htm>. (accessed 22 Feb 2013).
- Centers for Disease Control and Prevention. Recommendations for identification and public health management of persons with chronic hepatitis B virus infection. *MMWR* 2008;57(No. RR-8):1–20.
- Centers for Medicare and Medicaid Services. More than 100 000 healthcare providers paid for using electronic health records [Press Release]. 2012. <http://www.cms.gov> (accessed 5 Dec 2012).
- Jones G, Taright N, Boelle PY, et al. Accuracy of the ICD-10 codes for surveillance of *Clostridium difficile* infections, France. *Emerg Infect Dis* 2012;18:979–81.
- Kramer JR, Davila JA, Miller ED, et al. The validity of viral hepatitis and chronic liver disease diagnoses in Veterans Affairs administrative databases. *Aliment Pharmacol Ther* 2008;27:274–82.
- van de Garde EMW, Oosterheer JJ, Bonten M, et al. International Classification of Diseases codes showed modest sensitivity for detecting community-acquired pneumonia. *J Clin Epidemiol* 2007;60:834–8.
- Movig KLL, Leufkens HGM, Lenderink AW, et al. Validity of hospital discharge International Classification of Diseases (ICD) codes for identifying patients with hyponatremia. *J Clin Epidemiol* 2003;56:530–5.
- Sickbert-Bennett EE, Weber DJ, Poole C, et al. Utility of International Classification of Diseases, Ninth Revision, Clinical Modification codes for communicable diseases surveillance. *Am J Epidemiol* 2010;172:1299–305.
- Jhung MA, Banerjee SN. Administrative coding data and health care-associated infections. *Clin Infect Dis* 2009;49:949–55.
- Schaefer MK, Ellingson K, Conover C, et al. Evaluation of International Classification of Diseases, Ninth Revision, Clinical Modification codes for reporting methicillin-resistant *Staphylococcus aureus* infections at a hospital in Illinois. *Infect Control Hosp Epidemiol* 2010;31:463–8.
- Ollendorf DA, Fendrick AM, Massey K, et al. Is sepsis accurately coded on hospital bills? *Value Health* 2002;5:79–81.
- Dubberke ER, Reske KA, McDonald LC, et al. ICD-9 codes and surveillance for *Clostridium difficile*-associated disease. *Emerg Infect Dis* 2006;12:1576–9.
- Scheurer DB, Hicks LS, Cook EF, et al. Accuracy of ICD-9 coding for *Clostridium difficile* infections: a retrospective cohort. *Epidemiol Infect* 2007;135:1010–13.
- Klompas M, Haney G, Church D, et al. Automated identification of acute hepatitis B using electronic medical record data to facilitate public health surveillance. *PLoS One* 2008;3:e2626.

- 17 Iezzoni L. Assessing quality using administrative data. *Ann Intern Med* 1997; 127:666–74.
- 18 Andrade SE, Gurwitz JH, Chan KA, *et al.* Validation of diagnoses of peptic ulcers and bleeding from administrative databases: a multi-health maintenance organization study. *J Clin Epidemiol* 2002;55:310–13.
- 19 Stevenson KB, Khan Y, Dickman J, *et al.* Administrative coding data, compared with CDC/NHSN criteria, are poor indicators of health care-associated infections. *Am J Infect Control* 2008;36:155–64.
- 20 Moorman AC, Gordon SC, Rupp LB, *et al.* Baseline characteristics and mortality among people in care for chronic viral hepatitis: the Chronic Hepatitis Cohort Study (CHeCS). *Clin Infect Dis* 2013;56:40–50.
- 21 Geisinger Health System. 2012. <http://www.geisinger.org/community/index.html> (accessed 26 Jul 2012).
- 22 US Census Bureau. State and County Quickfacts. 2010. <http://quickfacts.census.gov/qfd/states/15000.html> (accessed 26 Jul 2012).
- 23 Kaiser Permanente. <http://xnet.kp.org/newscenter/aboutkp/fastfacts.html> (accessed 26 Jul 2012).
- 24 Spradling PR, Rupp L, Moorman AC, *et al.* Hepatitis B and C virus infection among 1.2 million persons with access to care: factors associated with testing and infection prevalence. *Clin Infect Dis* 2012;55:1047–55.
- 25 Rosen LM, Liu T, Merchant RC. Efficiency of International Classification of Diseases, Ninth Revision, billing code searches to identify emergency department visits for blood or body fluid exposure through a statewide multicenter database. *Infect Control Hosp Epidemiol* 2012;33:581–8.
- 26 Ramsay ME, Rushdy AA, Harris HE. Surveillance of hepatitis B: an example of a vaccine preventable disease. *Vaccine* 1998;16:S76–80.
- 27 American Medical Association. The Differences Between ICD-9 and ICD-10. 2012. <http://www.ama-assn.org/ama1/pub/upload/mm/399/icd10-icd9-differences-fact-sheet.pdf> (accessed 18 Jan 2013).
- 28 Bowman S. Why ICD-10 is worth the trouble. *J AHIMA* 2008;79:24–9.