Published in final edited form as:

Pharmacoepidemiol Drug Saf. 2009 September ; 18(9): 807–814. doi:10.1002/pds.1784.

Validity of The Health Improvement Network (THIN) for Epidemiologic Studies of Hepatitis C Virus Infection

Vincent Lo Re III, MD, MSCE^{1,2,3}, Kevin Haynes, PharmD, MSCE^{1,3}, Kimberly A. Forde, MD, MHS^{1,2,3}, A. Russell Localio, PhD^{1,3}, Rita Schinnar, MPA^{1,3}, and James D. Lewis, MD, MSCE^{1,2,3}

¹ Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania School of Medicine, Philadelphia, PA, USA

² Department of Medicine, University of Pennsylvania School of Medicine, Philadelphia, PA, USA

³ Department of Biostatistics and Epidemiology, University of Pennsylvania School of Medicine, Philadelphia, PA, USA

Abstract

Purpose: Before using computerized databases to study hepatitis C virus (HCV) epidemiology, the validity of the diagnosis must be assessed. We determined the accuracy of HCV diagnostic codes within The Health Improvement Network (THIN), an electronic database containing medical record data from general medical practices in the United Kingdom.

Methods: Patients with initial diagnostic codes for HCV infection and nonspecific viral hepatitis between 2000 and 2007 in the THIN database were identified. Questionnaires were mailed to general practitioners caring for a random sample of 150 of these patients (75 with an HCV code; 75 with a nonspecific viral hepatitis code) to collect information on HCV and other hepatitis diagnoses. We determined the positive predictive value of the database's HCV diagnostic codes and its ability to identify the date of a new HCV diagnosis.

Results: Usable surveys were returned for 146 (97%) patients. Among 74 patients with an HCV code and questionnaire data, HCV was confirmed in 64 (positive predictive value, 86%; 95% CI, 77% – 93%). In 40 (63%), the first recorded diagnosis in THIN was within 30 days of the date reported in the questionnaire (median difference, 11 days; interquartile range, 0 – 362 days). Among 72 patients with a nonspecific viral hepatitis code, 16 (22%) had HCV, but manual review of the database's electronic records correctly identified 12/16 (75%).

Address Correspondence to: V. Lo Re III, M.D., M.S.C.E. Center for Clinical Epidemiology and Biostatistics University of Pennsylvania School of Medicine 809 Blockley Hall, 423 Guardian Drive Philadelphia, PA 19104-6021 E-mail: vincentl@mail.med.upenn.edu Tel: 215-573-5964; Fax: 215-349-5111.

Potential conflicts of interest: Drs. Lewis has received research support for THIN and served as an unpaid consultant to THIN as a member of their Advisory Group. Dr. Haynes has also served as a consultant to THIN. All other authors report no potential conflicts of interest related to this manuscript.

KEY POINTS

- The hepatitis C virus (HCV)-specific diagnostic codes within The Health Improvement Network (THIN) are highly predictive of HCV infection.
- Manual review of THIN's records can help identify the cause of hepatitis among viral hepatitis NOS patients and reduce misclassification of HCV infection as viral hepatitis NOS.
- Limited data on HCV-related laboratory results and medications is a limitation that can potentially be overcome by obtaining supplemental information from general practitioners participating in THIN.

Conclusions: In THIN, the HCV-specific diagnostic codes are highly predictive of HCV infection. After manual review, few patients with a nonspecific viral hepatitis code were misclassified as having HCV infection.

Keywords

epidemiology; hepatitis C virus; HCV; validity; database

INTRODUCTION

Researchers have been interested in the epidemiology of hepatitis C virus (HCV) infection since its discovery in 1989 as the cause of non-A, non-B viral hepatitis.¹ Epidemiologic research on HCV infection has focused on prevalence and incidence rates,^{2,3} natural history studies,⁴⁻⁸ identification of risk factors for infection,⁹ and clinical outcomes of chronic infection.^{7,8,10,11} However, existing studies have had methodological limitations, including small sample sizes, uncontrolled confounding, selection bias, and recall bias for medication exposures, thereby preventing a full understanding of HCV epidemiology. As a result, there has been a need for methods by which large numbers of HCV-infected patients can be identified, correctly classified, and followed over extended periods of time.

One possible data source that can fill this need is The Health Improvement Network (THIN), a primary care medical records database in the United Kingdom (UK).¹² THIN has the potential to provide accurate and complete data on HCV-related diagnoses, outcomes, therapies, and relevant sociodemographic factors, and could be a valuable resource for epidemiologic research in HCV infection. However, before THIN can be used for epidemiologic research, the validity of the diagnosis of HCV infection in the database and the availability of data on HCV-related laboratory tests, anti-HCV medications, relevant confounding variables, and liver-related outcomes must be determined. As such, the goals of this study were to assess the accuracy of THIN's HCV diagnostic codes compared to GP confirmation of a documented diagnosis in the medical record and to evaluate the availability of information relevant to the study of HCV epidemiology in the database.

METHODS

Data Source

THIN is a database of electronic medical records from over 1,500 general practitioners (GPs) in over 380 UK practices.¹² Approximately 98% of the UK population is registered with a GP, who is responsible for almost the entirety of patients' medical care.¹³ GPs participating in THIN are trained to record information using the Vision general practice system (In Practice Systems; London, UK), and this electronic database serves as the primary medical record for participating practices. Data recorded in THIN include demographic information, medical diagnoses (including those resulting from referrals to specialists), prescriptions written by GPs, laboratory results, lifestyle characteristics, measurements taken during medical practice, and free text comments. Diagnoses are recorded using the READ diagnostic code scheme and prescriptions are recorded using codes from the UK Prescription Pricing Authority.^{14,15} GPs may maintain paper files with laboratory data, hospital discharge summaries, consultant letters, and additional patient-specific information, which can be obtained by requesting copies of paper files and/or through surveys of GPs without breach of confidentiality.

Prospective data collection in THIN began in September 2002; however, for numerous practices, electronic records were used as early as 1987 and are also included in THIN. The database currently has approximately 6.0 million patients, of which 2.8 million are

registered with THIN practices and can be followed prospectively. The remaining patients have historical data but have either left THIN practices or died. There are currently approximately 55 million patient-years of follow-up, or nearly 10 years per patient.

Study Design and Patients

We conducted a cross-sectional study among patients identified with HCV diagnostic codes between 1 January 2000 and 31 December 2007. Since clinicians might code HCV infection using nonspecific codes for viral hepatitis (referred to hereafter as viral hepatitis not otherwise specified [NOS]), we also examined patients identified with at least one diagnostic code for viral hepatitis NOS during this same time period to estimate the degree of misclassification of HCV infection as viral hepatitis NOS. For inclusion, patients had to be seen at a practice willing to participate in THIN validation studies and had to have an initial HCV or viral hepatitis NOS diagnosis recorded six months after registration with the practice to define incident diagnoses.¹⁶ Patients in the database were identified as HCV-infected if they had: 1) a diagnostic code for HCV infection (Table 1), or 2) a diagnostic code for viral hepatitis NOS (Table 2) and “hepatitis C” or “C” noted in the database's free text comments field.

Survey of General Practitioners

The primary study outcome was documented HCV infection, defined as a positive HCV antibody, positive recombinant immunoblot assay (RIBA), and/or quantifiable HCV RNA. To ascertain this outcome, we surveyed GPs caring for randomly-selected patients with diagnostic codes for HCV and viral hepatitis NOS, using a mailed questionnaire. The questionnaire was sent to practitioners by THIN's Additional Information Services, a company licensed to contact GPs for research purposes up to three times over a 3-month period.

The questionnaire asked the GP to confirm whether the patient had ever been diagnosed with HCV infection, and if so, to record the date of diagnosis and how it was made (e.g., positive HCV antibody, RIBA, and/or RNA). The GP was asked to provide copies of all laboratory results and/or consultant letters relevant to the diagnosis. If the patient was never diagnosed with HCV, the GP was asked if there was a diagnosis of hepatitis due to another cause (i.e., other viral infection [e.g., hepatitis A, B, D, or E]; non-viral cause of hepatitis [e.g., medications, alcohol]; cause never determined) or if hepatitis had not been diagnosed. Finally, information regarding conditions suggestive of hepatic decompensation (i.e., ascites, variceal hemorrhage, spontaneous bacterial peritonitis, hepatic encephalopathy, and/or hepatocellular carcinoma) was requested.

Availability of HCV-Related Data in the THIN Database

Among all subjects in THIN who ever had a diagnostic code recorded for HCV infection or viral hepatitis NOS, age, sex, body mass index (BMI), socioeconomic status, and alcohol consumption (based on GP assessment) were collected from the initial date at which the diagnostic code was recorded. Socioeconomic status was assessed with the Townsend index score (in quintiles, with higher quintiles representing worsening socioeconomic status), which was calculated using data from the 2001 UK census on house and car ownership, overcrowding of accommodation, and employment status.¹⁷

From the date when a diagnostic code for HCV infection or viral hepatitis NOS was first recorded through follow-up, we evaluated the availability of the following data in the THIN database: 1) HCV-related laboratory results (i.e., HCV antibody, RIBA, RNA, genotype, and hepatic aminotransferases), 2) prescriptions for anti-HCV medications (i.e., standard interferon alfa, pegylated interferon alfa, ribavirin), and 3) diagnostic codes for hepatic

decompensation (i.e., ascites, variceal hemorrhage, spontaneous bacterial peritonitis, hepatic encephalopathy, hepatocellular carcinoma).

Statistical Analyses

We first determined the positive predictive value of THIN's HCV diagnostic codes. We constructed contingency tables comparing the presence of any THIN HCV diagnostic code with the actual presence or absence of HCV infection as determined by confirmation from the GP via the mailed questionnaire. Since the prevalence of HCV infection in the UK is estimated to be <1% [18], we expected the negative predictive value of THIN's HCV codes to exceed 99% and so chose to focus on positive predictive value. Since our analyses only included subjects with HCV diagnostic codes, we could not determine the percentage of true cases of HCV infection in THIN and could not calculate sensitivity, specificity, or the kappa statistic. To determine the ability of THIN to identify incident HCV diagnoses, we compared the date of the first recorded HCV diagnosis in the database with the initial HCV diagnosis date reported by the GP in the questionnaire.

The point prevalence of HCV infection and viral hepatitis NOS within THIN was estimated for 31 December 2007 by dividing the number of prevalent cases of HCV on this date by the total population size of THIN at this time. Prevalent cases included any patient identified with an HCV or nonspecific viral hepatitis diagnostic code prior to 31 December 2007 and who were still receiving care from THIN GPs.

We next determined how often a viral hepatitis NOS code in the random sample was HCV infection. Given the lack of specificity of these codes, two investigators (V.L.R. and K.F.) manually reviewed the electronic records (i.e., diagnostic codes, antiviral medications, laboratory tests and results) of the randomly-selected viral hepatitis NOS patients to determine if a cause of liver disease was recorded in the database. This review was done without knowledge of information provided by GPs for individual patients. The review process identified an additional HCV diagnostic code ("hepatitis C status"), and patients with this code who had a free text comment for a positive result were subsequently considered HCV-infected for the determination of available HCV-related data and calculation of HCV prevalence.

We estimated that a simple random sample of 75 patients would allow determination of the positive predictive value of THIN's HCV diagnostic codes with a maximum 95% confidence interval (CI) of ± 0.12 , assuming a positive predictive value of 80%. All analyses were performed using STATA 10.0 (Stata Corp., College Station, TX).

Ethics Board Approval

The study protocol was approved by the University of Pennsylvania's Institutional Review Board and the UK's National Health System Multi-Centre Research Ethics Committee.

RESULTS

Validity of HCV-Specific Diagnostic Codes

We received responses from the GPs for 146 (97%) of the 150 patients (74/75 [99%] patients with an HCV diagnostic code and 72/75 [96%] patients with a viral hepatitis NOS code). Of the 150 questionnaires initially mailed to the GPs, 116 (77% response) were received after the initial mailing. After a second mailing requesting completion of the questionnaire, an additional 12 questionnaires were received (85% response). After a third mailing, 18 more questionnaires were returned (97% response).

Among the 74 patients with an HCV code and questionnaire data, a diagnosis of HCV infection was confirmed in 64, corresponding to a positive predictive value of 86% (95% CI, 77% – 93%). The diagnosis was confirmed by HCV antibody and/or RIBA in 59 (92%) and HCV RNA in 40 (63%). The median difference in the initial HCV diagnosis date reported by the GP and the first recorded HCV diagnosis in the THIN database was 11 days (interquartile range, 0 – 372 days). A total of 40 (63%) had the first recorded diagnosis in THIN within 30 days of the date reported by the GP. The GPs indicated that 6 (9%) of the 64 HCV patients received antiviral therapy with interferon plus ribavirin, but receipt of HCV therapy was recorded in the THIN database in only one patient. For the 10 subjects with an HCV diagnostic code but no GP-confirmed HCV infection, 7 had non-specific viral hepatitis that was incorrectly coded as HCV infection. The remaining 3 subjects were identified as never having had HCV infection by their GP, and manual review of the THIN database did not identify an alternative diagnosis.

Among the 146 patients with questionnaire data, 4 (3%) had diagnostic codes for hepatic decompensation recorded (3 patients with an HCV code; 1 with a viral hepatitis NOS code). The GPs confirmed that hepatic decompensation occurred in all 4 (variceal hemorrhage [2], spontaneous bacterial peritonitis [1], hepatic encephalopathy [1]).

Availability of HCV-Related Data in the THIN Database

The characteristics at the date of diagnosis of all patients with HCV or viral hepatitis NOS codes recorded between 1 January 2000 and 31 December 2007 are listed in Table 3. Table 4 shows data on HCV-related laboratory tests, anti-HCV medications, and liver-related outcomes for these patients that were available from the date when a diagnostic code for HCV infection or viral hepatitis NOS was first recorded through follow-up. The 976 HCV patients were followed up for a total of 2,812 person-years (median, 2.5 years per patient; interquartile range [IQR], 1.1 – 4.5 years), and the 978 patients with viral hepatitis NOS were followed for a total of 3,164 person-years (median, 3.0 years per patient; IQR, 1.3 – 4.9 years). Among the 976 HCV patients, only 328 (34%) had a code for an HCV laboratory test (i.e., HCV antibody, RIBA, and/or RNA), and only 69/328 (21%) had the result available in the database. In addition, 659 (68%) HCV patients had an aminotransferase result recorded during follow-up. The proportion of patients who received a prescription for an anti-HCV medication was low (Table 4). Fifty-five HCV patients (5.6%; 20 per 1,000 person-years) and 41 viral hepatitis NOS patients (4.2%; 13 per 1,000 person-years) had diagnostic codes for hepatic decompensation recorded during follow-up.

Prevalence of HCV Infection in the THIN Database

On 31 December 2007, there were 707 cases of HCV infection and 692 cases of viral hepatitis NOS identified in the THIN database among a total population size of 2,664,008. Given patient deaths and transfers from GP practices, the number of cases of HCV infection and nonspecific viral hepatitis was lower for these analyses. Thus, the prevalence of HCV infection was 26.5 (95% CI, 25 – 29) per 100,000 population, and the prevalence of viral hepatitis NOS was 26.0 (95% CI, 24 – 28) per 100,000 population.

Causes of Hepatitis for Patients with Nonspecific Viral Hepatitis Codes

Among the 72 patients with a diagnostic code for viral hepatitis NOS, the GP reported that 16 (22%) had HCV infection, 12 (17%) had another viral hepatitis infection (hepatitis A [2], hepatitis B [6], CMV [1], EBV [3]), 28 (39%) had non-viral hepatitis (autoimmune [8], alcohol [4], nonalcoholic steatohepatitis [3], drug-induced [4], gallstones [7], sarcoidosis [1], pregnancy [1]), 10 (14%) never had the cause of hepatitis identified, and 6 (8%) never had hepatitis diagnosed.

Manual review of the THIN database identified a cause of hepatitis in 55 (76%) of the 72 viral hepatitis NOS patients. Among the 16 viral hepatitis NOS patients who the GP reported had HCV infection, 12 (75%) were found to have a diagnostic code for “hepatitis C status” in the database with no free text comment that their status was negative. For the remaining 56 patients with a viral hepatitis NOS code, 43 (77%) had a cause of liver disease identified by manual review. Thus, after excluding the 12 patients with the “hepatitis C status” code, 43 of 60 (72%) patients with a code for viral hepatitis NOS had an identifiable cause of liver disease in the database and only 4 of the 60 (7%) had HCV infection.

DISCUSSION

Our study demonstrated that the THIN database contains a reliably diagnosed cohort of HCV-infected patients. THIN's HCV diagnostic codes had a positive predictive value exceeding 85%, suggesting that the majority of patients coded as having HCV infection indeed have the disease. Moreover, THIN contains information on important confounding variables in HCV epidemiologic studies such as alcohol and socioeconomic status, although additional data are needed to determine the validity of these variables.¹⁹ Thus, THIN could be used to conduct studies of HCV epidemiology.

However, our study identified important limitations in utilizing THIN for HCV epidemiologic research. First, 22% of randomly-selected viral hepatitis NOS patients had HCV infection. Manual review of these patients' electronic records enabled us to identify the cause for hepatitis in 76%. This review revealed an additional HCV diagnostic code (“hepatitis C status”) that, when reported as positive in the free text comments, improved our ability to identify HCV-infected patients. Thus, using manual review, only 7% of viral hepatitis NOS patients actually had HCV infection. Knowledge of this misclassification rate can be used in future THIN HCV studies when designing inclusion criteria, determining the potential efficiency of implementing surveys to identify additional HCV patients, and interpreting results.

Second, the THIN database misclassified the date of incident HCV diagnosis in approximately 40% of HCV-infected patients. This misclassification may be because HCV infection is a chronic disease that may be present for years prior to diagnosis. Moreover, GPs may not always accurately enter into the electronic record diagnoses that occurred prior to registration with the practice or prior to implementation of the electronic record. We examined the database's free text comments to identify additional information on the date of incident HCV diagnosis, but GPs did not record this date in that field. It should be noted that in most cases, incident HCV diagnoses recorded in the database may be months or years removed from the actual date of HCV infection.

Another limitation of the THIN database for HCV epidemiologic research was the lack of data on HCV antibody, RNA, and genotype. Hepatic aminotransferase results, which are markers of liver inflammation, were also not available for all HCV-infected patients. The absence of such data limits the ability to identify spontaneous HCV clearance, determine chronic infection, examine data on HCV genotypes, and evaluate hepatotoxicity of medications. This information could be obtained through GP surveys.

Very few HCV patients had prescriptions for anti-HCV medications recorded in the THIN database. Among the randomly-selected sample of HCV-infected patients whose GP's were surveyed, receipt of interferon-based treatment was recorded in the THIN database in only one patient among the six reported by their GP to have received such therapy. In the UK, liver specialists typically prescribe antiviral therapy for chronic HCV patients, and this information might not be recorded in GP's electronic records. This limits the usefulness of

THIN for pharmacoepidemiologic studies in which knowledge of exposure to anti-HCV therapy is important.

This study provided a unique opportunity to compare the prevalence of HCV infection in THIN with that previously reported in the UK. The estimated prevalence of HCV infection in England and Wales has ranged from 0.1% to 1%.^{18,20,21} We observed the prevalence of HCV infection in THIN to be 0.03%, which is considerably lower than these estimates. This finding is due in part to the misclassification of HCV infection as viral hepatitis NOS described above. However, this is counterbalanced by the knowledge that 14% of patients with an HCV code did not have documented HCV infection. Differences between our results and that of prior studies could be related to differences in the methods of ascertaining HCV prevalence. In prior studies, prevalence of HCV was assessed by antibody testing of sera from residual specimens submitted to laboratories for routine diagnostic examination. Furthermore, since HCV infection is often asymptomatic, there are likely many patients in THIN who have not yet been diagnosed. Finally, the characteristics of THIN patients could differ from other populations that have been studied, such as having a slightly higher socioeconomic status.

The cumulative incidence of hepatic decompensation in this study was comparable to that reported in prior cohort studies examining long-term outcomes of HCV infection. These studies reported incidences of hepatic decompensation ranging from 2% to 12%.⁴⁻⁸ The cumulative incidence of hepatic decompensation among the THIN HCV patients followed in this study fell within this range. Future studies should confirm further the validity of THIN's diagnostic codes for hepatic decompensation.

There are several potential limitations to our study. First, in any survey-based study, one must consider the potential for non-response bias. Our response rate was high, so this form of bias should have little impact on our results. Second, the questionnaire was administered to GPs in practices that are willing to participate in THIN research studies, and therefore results might not be generalizable to non-participating practices. Likewise, one must consider whether the patients randomly sampled for validation were representative of the full population. We observed that those sampled were similar to the full population in terms of Townsend scores, suggesting this to be the case (data not shown). Finally, we assumed that GP's responses to the questionnaire reflected the truth. It is possible that errors might have occurred in completing the questionnaire. However, we did request copies of the correspondences between the GPs and specialists related to the HCV diagnosis to ensure misclassification of the outcome was minimized.

In conclusion, THIN can be an important resource for studying HCV epidemiology. The positive predictive value of THIN's HCV diagnostic codes was high. Manual review of THIN's records can help identify the cause of hepatitis among viral hepatitis NOS patients. The lack of data on HCV-related laboratory results and medications are limitations that can potentially be overcome by obtaining supplemental information from GPs.

Acknowledgments

Financial support: National Institutes of Health research grant K01 AI070001 [V.L.R.], grant number UL1RR024134 from the National Center for Research Resources [K.H., R.S., J.D.L.], and Centers for Education and Research on Therapeutics cooperative agreement U18HS106946 from the Agency for Healthcare Research and Quality [V.L.R., K.H., J.D.L.]. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Agency for Healthcare Research and Quality, the National Center for Research Resources, or the National Institutes of Health.

REFERENCES

1. Alter HJ, Purcell RH, Shih JW, Melpolder JC, Houghton M, Choo QL, Kuo G. Detection of antibody to hepatitis C virus in prospectively followed transfusion recipients with acute and chronic non-A, non-B hepatitis. *N Engl J Med.* 1989; 321(22):1494–500. [PubMed: 2509915]
2. Armstrong GL, Wasley A, Simard EP, McQuillan GM, Kuhnert WL, Alter MJ. The prevalence of hepatitis C virus infection in the United States, 1999 through 2002. *Ann Intern Med.* 2006; 144(10):705–14. [PubMed: 16702586]
3. Alter MJ, Kruszon-Moran D, Nainan OV, McQuillan GM, Gao F, Moyer LA, Kaslow RA, Margolis HS. The prevalence of hepatitis C virus infection in the United States, 1988 through 1994. *N Engl J Med.* 1999; 341(8):556–62. [PubMed: 10451460]
4. Di Bisceglie AM, Goodman ZD, Ishak KG, Hoofnagle JH, Melpolder JJ, Alter HJ. Long-term clinical and histopathological follow-up of chronic posttransfusion hepatitis. *Hepatology.* 1991; 14(6):969–74. [PubMed: 1959884]
5. Makris M, Preston FE, Rosendaal FR, Underwood JC, Rice KM, Triger DR. The natural history of chronic hepatitis C in haemophiliacs. *Br J Haematol.* 1996; 94(4):746–52. [PubMed: 8826904]
6. Ragni MV, Belle SH. Impact of human immunodeficiency virus infection on progression to end-stage liver disease in individuals with hemophilia and hepatitis C virus infection. *J Infect Dis.* 2001; 183(7):1112–5. [PubMed: 11237838]
7. Telfer P, Sabin C, Devereux H, Scott F, Dusheiko G, Lee C. The progression of HCV-associated liver disease in a cohort of haemophilic patients. *Br J Haematol.* 1994; 87(3):555–61. [PubMed: 7993796]
8. Thomas DL, Astemborski J, Rai RM, Anania FA, Schaeffer M, Galai N, Nolt K, Nelson KE, Strathdee SA, Johnson L, Laeyendecker O, Boitnott J, Wilson LE, Vlahov D. The natural history of hepatitis C virus infection: host, viral, and environmental factors. *JAMA.* 2000; 284(4):450–6. [PubMed: 10904508]
9. Thomas DL, Vlahov D, Solomon L, Cohn S, Taylor E, Garfein R, Nelson KE. Correlates of hepatitis C virus infections among injection drug users. *Medicine (Baltimore).* 1995; 74(4):212–20. [PubMed: 7623656]
10. El-Serag HB. Hepatocellular carcinoma and hepatitis C in the United States. *Hepatology.* 2002; 36(5 Suppl 1):S74–83. [PubMed: 12407579]
11. Kenny-Walsh E. Clinical outcomes after hepatitis C infection from contaminated anti-D immune globulin. Irish Hepatology Research Group. *N Engl J Med.* 1999; 340(16):1228–33. [PubMed: 10210705]
12. Lewis JD, Schinnar R, Bilker WB, Wang X, Strom BL. Validation studies of the health improvement network (THIN) database for pharmacoepidemiology research. *Pharmacoepidemiol Drug Saf.* 2007; 16(4):393–401. [PubMed: 17066486]
13. Lis Y, Mann RD. The VAMP Research multi-purpose database in the U.K. *J Clin Epidemiol.* 1995; 48(3):431–43. [PubMed: 7897464]
14. Chisholm J. The Read clinical classification. *BMJ.* 1990; 300(6732):1092. [PubMed: 2344534]
15. Garcia Rodriguez LA, Perez Gutthann S. Use of the UK General Practice Research Database for pharmacoepidemiology. *Br J Clin Pharmacol.* 1998; 45(5):419–25. [PubMed: 9643612]
16. Lewis JD, Brensinger C, Bilker WB, Strom BL. Validity and completeness of the General Practice Research Database for studies of inflammatory bowel disease. *Pharmacoepidemiol Drug Saf.* 2002; 11(3):211–8. [PubMed: 12051120]
17. Adams J, Ryan V, White M. How accurate are Townsend Deprivation Scores as predictors of self-reported health? A comparison with individual level data. *J Public Health (Oxf).* 2005; 27(1):101–6. [PubMed: 15564276]
18. Balogun MA, Ramsay ME, Hesketh LM, Andrews N, Osborne KP, Gay NJ, Morgan-Capner P. The prevalence of hepatitis C in England and Wales. *J Infect.* 2002; 45(4):219–26. [PubMed: 12423608]
19. Lewis JD, Brensinger C. Agreement between GPRD smoking data: a survey of general practitioners and a population-based survey. *Pharmacoepidemiol Drug Saf.* 2004; 13(7):437–41. [PubMed: 15269927]

20. Booth JC, Brown JL, Thomas HC. The management of chronic hepatitis C virus infection. *Gut*. 1995; 37(4):449–54. [PubMed: 7489926]
21. Mutimer DJ, Harrison RF, O'Donnell KB, Shaw J, Martin BA, Atrah H, Ala FA, Skidmore S, Hubscher SG, Neuberger JM, et al. Hepatitis C virus infection in the asymptomatic British blood donor. *J Viral Hepat*. 1995; 2(1):47–53. [PubMed: 7493294]

Table 1

READ diagnostic codes used to identify patients with HCV infection in the THIN database.

READ Diagnostic Code	Disease Classification
A704000	Viral hepatitis C with coma
A705000	Viral hepatitis C without hepatic coma
A707200	Chronic viral hepatitis C
A70z000	Hepatitis C
ZV02C00	Hepatitis C carrier
2J11.00	Hepatitis C immune
43X3.00	Hepatitis C antibody test positive
65Q7.00	Viral hepatitis carrier with GP comment of 'hepatitis C'
A705400	Hepatitis non-A, non-B
2J1..00*	Hepatitis C status

* This HCV diagnostic code was added following manual review. When present with a free text comment for a positive finding, this code identified additional cases of HCV infection. This code was not included in the analysis of the positive predictive value of the HCV diagnostic codes.

Table 2

READ diagnostic codes used to identify patients with viral hepatitis not otherwise specified in the THIN database.

READ Diagnostic Code	Disease Classification
A70.00	Viral hepatitis
A704.00	Other specified viral hepatitis with coma
A704z00	Other specified viral hepatitis with hepatic coma NOS
A705.00	Other specified viral hepatitis without coma
A705z00	Other specified viral hepatitis without mention of coma NOS
A706.00	Unspecified viral hepatitis with coma
A707.00	Chronic viral hepatitis
A707X00	Chronic viral hepatitis, unspecified
A70z.00	Unspecified viral hepatitis
AyuB.00	Viral hepatitis
AyuB000	Other specified acute viral hepatitis
AyuB100	Other chronic viral hepatitis
AyuB200	Chronic viral hepatitis, unspecified
AyuB300	Unspecified viral hepatitis with coma
AyuB400	Unspecified viral hepatitis without coma
J614.00	Chronic hepatitis
J614000	Chronic persistent hepatitis
J614100	Chronic active hepatitis
J614200	Chronic aggressive hepatitis
J614y00	Chronic hepatitis unspecified
J614z00	Chronic hepatitis NOS
J631.00	Hepatitis in viral diseases EC
J631z00	Hepatitis in viral diseases EC NOS
J633.00	Hepatitis unspecified
J633z00	Hepatitis unspecified NOS

Table 3

Patient characteristics at the initial date a diagnostic code for hepatitis C virus infection or viral hepatitis not otherwise specified was recorded in the THIN database during the period 1 January 2000 through 31 December 2007.

Characteristic	Hepatitis C (n=976)	Viral Hepatitis NOS (n=978)
Median age (yrs, interquartile range)	41 (33-48)	47 (35-59)
Sex (no., %)		
Male	625 (64%)	504 (52%)
Female	351 (36%)	474 (48%)
Body mass index (no., %)		
Underweight (<18.4 kg/m ²)	35 (4%)	23 (2%)
Ideal (18.5 kg/m ² – 24.9 kg/m ²)	399 (41%)	353 (36%)
Overweight (25.0 kg/m ² – 29.9 kg/m ²)	169 (17%)	239 (24%)
Obese (30.0 kg/m ² – 39.9 kg/m ²)	83 (9%)	121 (12%)
Very Obese (>40.0 kg/m ²)	6 (0.6%)	16 (2%)
Missing	284 (29.1%)	226 (23%)
Socioeconomic deprivation, in quintiles (no., %)*		
Quintile 1	69 (7%)	188 (19%)
Quintile 2	104 (11%)	158 (16%)
Quintile 3	152 (16%)	169 (17%)
Quintile 4	224 (23%)	187 (19%)
Quintile 5	328 (34%)	195 (20%)
Missing	99 (10%)	81 (8%)
Alcohol consumption (no., %)		
Lifetime non-drinker	145 (15%)	193 (20%)
Past drinker without a diagnosis of abuse or dependence	58 (6%)	19 (2%)
Past drinker with history of abuse or dependence	13 (1%)	7 (1%)
Current drinker without a diagnosis of abuse or dependence	383 (39%)	474 (48%)
Current abuse and dependence	178 (18%)	97 (10%)
Missing	199 (20%)	188 (19%)

* Based on Townsend index score (in quintiles, with higher quintiles representing worsening socioeconomic status).

Table 4

Availability of data on hepatitis C-related laboratory tests, antiviral medications, and liver-related complications throughout patients' follow-up between 1 January 2000 and 31 December 2007.

Finding	Hepatitis C (n=976)	Viral Hepatitis NOS (n=978)
Laboratory test data available in THIN during follow-up (no., %)		
Hepatitis C test performed [*]	259 (34%)	158 (19%)
Hepatitis C test result available [*]	69 (7%)	23 (2%)
Hepatic aminotransferase test performed [‡]	659 (68%)	792 (81%)
Hepatic aminotransferase test result available [‡]	659 (68%)	791 (81%)
Antiviral medication usage (no., %)		
Standard interferon alfa	3 (0.3%)	2 (0.2%)
Pegylated interferon alfa	4 (0.4%)	0 (0%)
Ribavirin	5 (0.5%)	1 (0.1%)
Liver-related complications (no., %)		
Ascites	20 (2.0%)	13 (1.3%)
Spontaneous bacterial peritonitis	6 (0.6%)	4 (0.4%)
Esophageal/gastric variceal hemorrhage	21 (2.2%)	24 (2.5%)
Hepatic encephalopathy	7 (0.7%)	2 (0.2%)
Hepatocellular carcinoma	9 (0.9%)	4 (0.4%)

^{*} Hepatitis C tests include HCV antibody test, HCV recombinant immunoblot assay, HCV RNA, and/or HCV genotype

[‡] Hepatic aminotransferase tests include alanine aminotransferase and aspartate aminotransferase