nature genetics

Genome-wide association study identifies a susceptibility locus for HCV-induced hepatocellular carcinoma

Vinod Kumar[1,2], Naoya Kato[3], Yuji Urabe[1], Atsushi Takahashi[2], Ryosuke Muroyama[3], Naoya Hosono[2], Motoyuki Otsuka[4], Ryosuke Tateishi[4], Masao Omata[4], Hidewaki Nakagawa[2], Kazuhiko Koike[4], Naoyuki Kamatani[2], Michiaki Kubo[2], Yusuke Nakamura[1,2] & Koichi Matsuda[1]

To identify the genetic susceptibility factor(s) for hepatitis C virus–induced hepatocellular carcinoma (HCV-induced HCC), we conducted a genome-wide association study using 432,703 autosomal SNPs in 721 individuals with HCV-induced HCC (cases) and 2,890 HCV-negative controls of Japanese origin. Eight SNPs that showed possible association ($P < 1 \times 10^{-5}$) in the genome-wide association study were further genotyped in 673 cases and 2,596 controls. We found a previously unidentified locus in the 5′ flanking region of *MICA* on 6p21.33 (rs2596542, $P_{combined} = 4.21 \times 10^{-13}$, odds ratio = 1.39) to be strongly associated with HCV-induced HCC. Subsequent analyses using individuals with chronic hepatitis C (CHC) indicated that this SNP is not associated with CHC susceptibility ($P = 0.61$) but is significantly associated with progression from CHC to HCC ($P = 3.13 \times 10^{-8}$). We also found that the risk allele of rs2596542 was associated with lower soluble MICA protein levels in individuals with HCV-induced HCC ($P = 1.38 \times 10^{-13}$).

It is estimated that more than 170 million people are infected with HCV worldwide[1]. Persistent HCV infection causes CHC and, subsequently, fatal liver diseases such as liver cirrhosis and HCC. Therefore, the treatment of HCV carriers is an issue of global importance. HCC is the third most common cause of cancer-related deaths[2], and HCV infection accounts for 30–70% of the individuals with HCC[3,4]. HCV-induced HCC is a multistep and progressive liver disease in which disease progression may be influenced by both environmental and genetic risk factors. The impact of host genetic variation on progression to CHC after HCV exposure is well documented by recent genome-wide association studies (GWAS)[5–7]. However, no comprehensive analyses have been performed to explore the genetic basis of HCV-induced HCC. Therefore, we conducted a GWAS for HCV-induced HCC.

We genotyped the DNA of 721 individuals with HCV-induced HCC and 2,890 HCV-negative controls (**Supplementary Table 1**) from BioBank Japan[8]. After the initial standard SNP quality filters, we obtained genotyping results for 432,703 SNPs for association analysis. Because progression from CHC to liver cancer is strongly affected by age and gender[3], we performed a logistic regression analysis by including age and gender as covariates at all tested loci in our analyses. The genetic inflation factor (λ) was 1.03, indicating that there is no or little population stratification (**Supplementary Fig. 1**). Although no SNPs cleared the GWAS significance threshold ($P < 5 \times 10^{-8}$) at this stage, we identified eight independent loci showing possible association ($P < 1 \times 10^{-5}$; **Supplementary Fig. 2**).

In the replication stage, 673 cases from an independent HCC cohort from the University of Tokyo and 2,596 HCV-negative controls from BioBank Japan were genotyped at these eight SNPs. We observed a significant replication of association at rs2596542 on chromosome 6p21.33 ($P = 8.62 \times 10^{-9}$, odds ratio (OR) = 1.44, 95% confidence interval (CI) 1.27–1.63; **Table 1**), whereas the remaining seven SNPs failed to replicate the association (**Supplementary Table 2**). Furthermore, the combination analysis of the GWAS and replication study data at rs2596542 revealed a highly significant association in which the frequency of the risk allele A is higher in cases ($P = 4.21 \times 10^{-13}$, OR = 1.39; **Fig. 1** and **Table 1**) after the age and gender adjustment, without any heterogeneity ($P = 0.24$) between the two stages. To further investigate the impact of rs2596542 on the complex nature of the HCV-induced HCC phenotype, we genotyped 1,730 individuals with CHC who had not developed liver cirrhosis or HCC during their recruitment. As a result, rs2596542 was found to have no association with chronic hepatitis C susceptibility ($P = 0.61$) but was significantly associated with progression from CHC to HCC ($P = 3.13 \times 10^{-8}$, OR = 1.36; **Table 2**).

Because heavy alcohol consumption (>50 g per day) as well as poor response to interferon (IFN) treatment were shown to be the major risk factors for HCC among individuals with CHC[9], we evaluated the effect of alcohol consumption as a confounding factor and found that rs2596542 remained highly significant even after adjustment for this factor (non-HCV versus HCC, OR = 1.39, $P = 1.22 \times 10^{-11}$; CHC versus HCC, OR = 1.25, $P = 2.31 \times 10^{-4}$; **Supplementary Table 3**). The major genotypes of HCV can be determined by a serotyping

[1]Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, University of Tokyo, Tokyo, Japan. [2]Center for Genomic Medicine, The Institute of Physical and Chemical Research (RIKEN), Kanagawa, Japan. [3]Unit of Disease Control Genome Medicine, The Institute of Medical Science, University of Tokyo, Tokyo, Japan. [4]Department of Gastroenterology, Graduate School of Medicine, University of Tokyo, Tokyo, Japan. Correspondence should be addressed to K.M. (koichima@ims.u-tokyo.ac.jp).

Received 20 July 2010; accepted 23 March 2011; published online 17 April 2011; doi:10.1038/ng.809

Table 1 Association results of rs2596542 in the GWAS, replication stage and combined analysis

SNP	Chr. (locus)	Stage	Case RAF	Control RAF	*P*	OR (95% CI)
rs2596542 (A/G)	6 (*MICA*)	GWAS[a]	0.388	0.331	4.50×10^{-6}	1.34 (1.16–1.53)
		Replication[a]	0.413	0.331	8.62×10^{-9}	1.44 (1.27–1.63)
		Combined[a]	0.400	0.331	4.21×10^{-13}	1.39 (1.27–1.52)
				MH test	7.76×10^{-12}	1.35 (1.24–1.47)

We analyzed 1,394 cases with HCC (721 in the GWAS and 673 in the replication) and 5,486 controls (2,890 in GWAS and 2,596 in replication). Chr., chromosome; RAF, risk allele frequency (allele A); OR, odds ratio for the minor allele calculated by considering the major allele as a reference; MH, Mantel-Haenszel.
[a]*P* values and ORs are adjusted for age and gender by logistic regression analysis under an additive model.

assay that is based on the type-specific antibodies produced by the infected host[10]. A subgroup analysis for HCV serotypes or history of IFN therapy indicated that this variation is associated with HCC susceptibility independently of HCV genotypes or treatment response (**Supplementary Fig. 3**). Consistent with this result, rs1051796, which had $r^2 = 0.7$ and $D' = 0.95$ with rs2596542, was not associated with IFN response ($P = 0.89$) according to previously published data in the Japanese population[11].

rs2596542 is located within the class I major histocompatibility complex (MHC) region. The human MHC region encompasses the complex and extended linkage disequilibrium (LD) structure[12,13]. Several HLA alleles and genes within MHC region have been implicated in HCV infection or clearance or in response to treatment[14–16]. Therefore, we searched the whole 7.5-Mb extended MHC region using GWAS data to test the possibility of other associated loci. We found a moderate association peak at rs9275572 ($P = 4.99 \times 10^{-5}$), which is located between *HLA-DQA* and *HLA-DQB* loci (**Supplementary Fig. 4**). Subsequent replication and combination analyses at rs9275572 indicated a significant association with HCV-induced HCC ($P = 9.38 \times 10^{-9}$, OR = 1.30; **Supplementary Table 4**). The multiple logistic regression analysis to control for alcohol consumption along with age and gender also indicated a significant association at rs9275572 ($P = 3.21 \times 10^{-8}$, OR = 1.29; **Supplementary Table 5**). However, rs2596542 was not in high LD with rs9275572 ($D' = 0.41$, $r^2 = 0.16$), and both SNPs remained associated with HCC even after conditional analysis on each other and had small reductions in their ORs upon conditioned analysis (OR = 1.23, $P = 4.43 \times 10^{-6}$ and OR = 1.17, $P = 0.00059$, respectively; **Supplementary Table 6**). A haplotype analysis between these two markers showed four possible haplotypes, with haplotype AA showing higher risk (with OR = 1.44) compared to the major haplotype GG (**Supplementary Table 7**). However, the OR for the risk haplotype was 1.32 with $P = 2.31 \times 10^{-10}$ after comparing against all observed haplotypes in the population (**Supplementary Table 7**), which is weaker than that of rs2596542 alone (OR = 1.39, $P = 4.21 \times 10^{-13}$). Hence, the impact of rs2596542 is much stronger than the haplotype of two SNPs, suggesting that rs2596542 is a principal genetic factor in this region. We also found that rs9275572 has a moderate association with CHC susceptibility as well as progression from CHC to HCC ($P = 0.03$ and $P = 2.58 \times 10^{-5}$, OR = 1.09 and OR = 1.29, respectively; **Supplementary Table 8**). Because *HLA-DQ* and *HLA-DR* alleles were shown to be associated with viral persistence and early liver disease among Japanese individuals[16], further study will be needed to confirm whether the association at rs9275572 is because of its LD with *HLA-DQ* or *DR* alleles.

In this regard, it is interesting to note that rs9275572 had a very strong expression quantitative trait locus effect on *HLA-DQB1* ($\log_{10}$ odds (LOD) ≥ 19.48) and *HLA-DRB4* alleles (LOD ≥ 26.88)[17]. Thus, it will be important to test the functional effect of the common haplotype (AA; **Supplementary Table 7**), which tags the risk alleles at these two SNPs.

Two SNPs, rs12979860 and rs8099917, at the *IL28B* locus were reported to be associated with spontaneous clearance of HCV virus[18] and response to pegylated IFN-α and ribavirin therapy[11], respectively. However, we found no association at rs12979860 and rs8099917 in our dataset (**Supplementary Table 9**). Because we used non-HCV control subjects rather than subjects who had cleared HCV infection spontaneously, and because only about 20% of the cases with HCC had been treated with IFN, our study may not be suitable to detect associations at the *IL28B* locus. In addition, the protective C allele at rs12979860 is nearly fixed throughout east Asia, with a frequency of more than 91% in the Japanese population as compared to 67% in European Americans[6], indicating a role for other factors in spontaneous clearance.

The top associated SNP, rs2596542, is located 4.7 kb upstream of *MICA*, the MHC class I polypeptide-related sequence A gene, and 41.7 kb downstream of the *HLA-B* gene (**Supplementary Fig. 5**). The regional association plot at the rs2596542 locus, made using genotype data from the GWAS (**Fig. 1**) and imputation analysis (**Supplementary Fig. 6**), revealed that all of the modestly associated SNPs are tightly

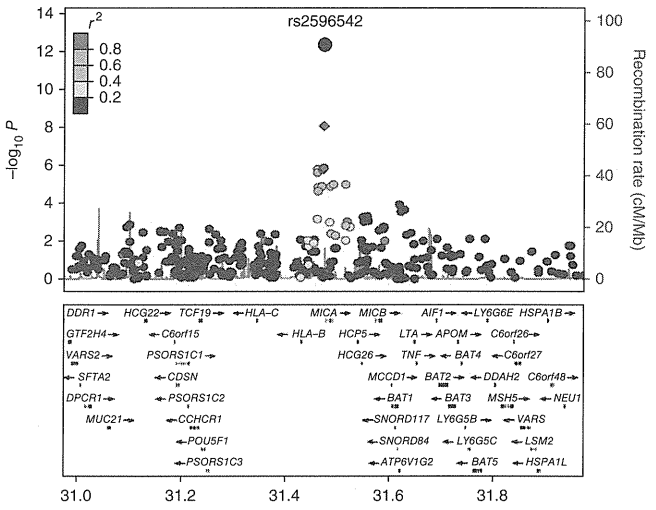


Figure 1 Regional association plot at rs2596542. Above, the *P* values of genotyped SNPs are plotted (as $-\log_{10}$ values) against their physical position on chromosome 6 (NCBI Build 36). The *P* value for rs2596542 at the GWAS stage, replication stage and combination analysis is represented by a purple diamond, circle and diamond, respectively. Estimated recombination rates from the HapMap JPT population show the local LD structure. Inset, the SNP's colors indicate LD with rs2596542 according to a scale from $r^2 = 0$ to $r^2 = 1$ based on pairwise r^2 values from HapMap JPT. Below, gene annotations from the UCSC genome browser.

Table 2 rs2596542 (A/G) is associated with progression from CHC to HCC

Subjects	RAF	(Comparison) *P*[a]	OR[a]	95% CI
Healthy	0.331			
CHC	0.333	(Healthy vs. CHC) 0.61	1.02	0.94–1.10
HCC	0.398	(CHC vs. HCC) 3.13×10^{-8}	1.36	1.22–1.51

We analyzed 5,486 controls, 1,730 CHC cases and 1,394 HCC cases. RAF, risk allele frequency (allele A); OR, odds ratio for the minor allele by considering the major allele as a reference.
[a]Calculated by logistic regression analysis, by PLINK upon age and gender adjustment under additive model.

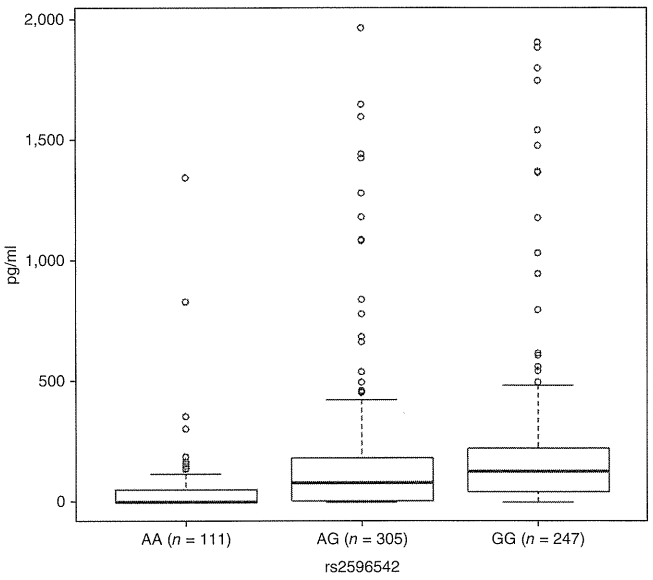


Figure 2 Correlation between soluble MICA levels and rs2596542 genotype. The *x* axis shows the genotypes at rs2596542, and the *y* axis shows the concentration of soluble MICA in pg/ml. The number of independent samples tested in each group is shown in parentheses. Each group is shown as a box plot, and the median values are shown as thick dark horizontal lines (median values of AA = 0, AG = 43.6 and GG = 77.74). The box covers the twenty-fifth to seventy-fifth percentiles, and the whiskers outside the box extend to the highest and lowest value within 1.5 times the interquartile range. Points outside the whiskers are outliers. We tested the difference in the median values among genotypes using the Kruskal-Wallis test ($P = 1.6 \times 10^{-13}$). We plotted the box plots using default settings in R (see URLs).

linked to rs2596542 ($r^2 > 0.4$) and are confined to the *MICA* gene locus. On the other hand, the imputation analysis of *HLA*-tagging SNPs did not show any evidence of linkage with rs2596542 (Online Methods and **Supplementary Table 10**), suggesting that *MICA* is a disease-associated candidate gene at this locus.

MICA is a membrane protein that acts as a ligand for NKG2D to activate anti-tumor effects through natural killer cells and $CD8^+$ T cells[19]. On the other hand, MICA is secreted into the serum by cleavage at the transmembrane domain with matrix metalloproteinases[20,21] and inhibits the anti-tumor effect of natural killer cells and $CD8^+$ T cells by blocking their action[22–24]. Elevated expression of both the membrane-bound and soluble forms of MICA (sMICA) have been reported in several cancers, including HCC[25–27]. Exon 5 of *MICA* encodes the transmembrane domain and contains a variable number of tandem repeats (VNTR) consisting of 4, 5, 6 or 9 repeats of GCT or one additional G nucleotide insertion into the 5-GCT-repeat allele (referred as A4, A5, A6, A9 and A5.1, respectively). The insertion of G (A5.1) causes a premature stop codon and subsequent loss of the transmembrane domain, leading to altered subcellular localization[28]. Therefore, we tested whether rs2596542 is in linkage with functional *MICA* VNTR alleles.

We further genotyped 673 cases with HCV-induced HCC and 890 non-HCV controls for the *MICA* VNTR locus with capillary-based electrophoresis (**Supplementary Fig. 7**). A case-control analysis revealed that the *MICA* VNTR is associated with HCV-induced HCC (global $P = 4.55 \times 10^{-7}$; **Supplementary Table 11**). Particularly, alleles A9 and A6 were associated with conferring a higher risk of HCC (OR = 1.73 and OR = 1.34, respectively), whereas the A5 and A5.1 alleles had a protective effect. Comparison of the genotypes at rs2596542 and the VNTR locus revealed that the A risk allele at rs2596542 is in LD with the A9 and A4 alleles, and the non-risk G allele is in LD with the A5 and A5.1 alleles, whereas we observed no linkage between an A6 allele and rs2596542 (**Supplementary Table 12**). We also genotyped 124 individuals with CHC; however, we observed no significant association between individuals with CHC and controls or individuals with CHC and HCC (**Supplementary Tables 13,14**).

We then tested whether the VNTR alleles, rs2596542 alleles, or VNTR-rs2596542 haplotypes had any association with MICA expression in individuals with HCV-induced HCC. We determined sMICA levels by ELISA using a total of 665 HCC serum samples (**Supplementary Table 15**). Notably, rs2596542 was significantly correlated with sMICA levels, and specifically, the risk genotype AA was associated with low levels of sMICA ($P = 1.38 \times 10^{-13}$; **Fig. 2**), whereas VNTR alleles (**Supplementary Fig. 8**) and VNTR- rs2596542 haplotypes (**Supplementary Table 16**) showed no strong association. The absence of any correlation between *MICA* VNTR alleles and sMICA suggests that sMICA levels are not regulated by post-translational processing or a premature stop codon caused by A5.1 alleles in individuals with HCC. We also examined the sMICA level in different stages of HCV-induced liver disease (in non-HCV subjects and those with CHC and HCV-induced liver cirrhosis) and found that sMICA level was elevated at the early stage of disease and was not correlated with disease progression (**Fig. 3**). Additionally, the risk allele A was also correlated with low sMICA levels in subjects with CHC (**Supplementary Fig. 9**). These findings suggest that MICA expression was induced by factors caused by chronic HCV infection,

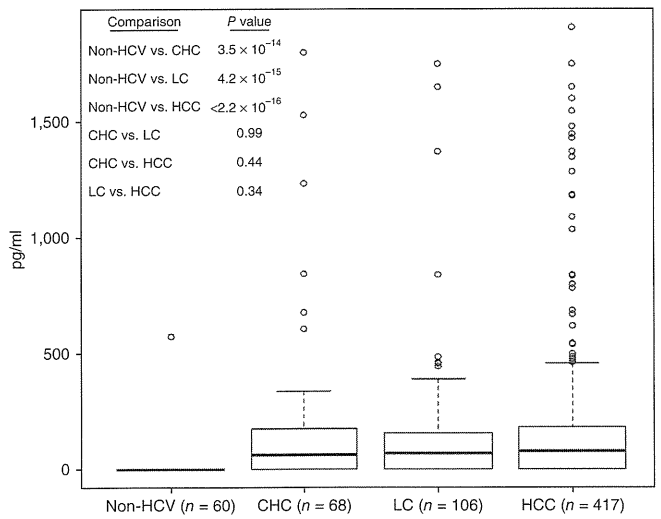


Figure 3 Correlation between soluble MICA and HCV-related diseases. The *x* axis shows the disease stages after HCV infection, and the *y* axis shows the concentration of soluble MICA in pg/ml. The number of independent samples tested in each group is shown in parentheses. Each group is shown as a box plot, and the median values are shown as thick dark horizontal lines (median values of non-HCV = 0, CHC = 64.55, LC = 72.11 and HCC = 77.98). The box covers the twenty-fifth to seventy-fifth percentiles, and the whiskers outside the box extend to the highest and lowest value within 1.5 times the interquartile range. Points outside the whiskers are outliers. We tested the difference in the median values among the disease groups using the Wilcoxon rank test. The box plots were plotted using default settings in R. Non-HCV, individuals not exposed to HCV infection; CHC, individuals with chronic hepatitis C; LC, individuals with liver cirrhosis; HCC, individuals with hepatocellular carcinoma.

similar to various types of stresses such as viral infection, inflammation and heat shock[29,30]. The levels of sMICA were shown to be directly proportional to the level of membrane-bound MICA[25], and membrane bound MICA is essential for activating natural killer cells and $CD8^+$ T cells to eliminate virus-infected cells[19]. Considering the association of the risk allele A with low levels of sMICA, our findings suggest that the individuals who carry the rs2596542 A allele would express low levels of membrane-bound MICA in response to HCV infection, which thus leads to poor or no activation of natural killer cells and $CD8^+$ T cells against virus-infected cells. Eventually, these individuals are likely to progress from CHC to HCC. Notably, several SNPs that are in absolute linkage with rs2596542 are located within the promoter or enhancer region of *MICA* and may alter the binding of stress-inducible transcriptions factors such as heat shock proteins (**Supplementary Table 17**). In this regard, it is important to analyze the factors that regulate *MICA* expression, particularly in the context of CHC. Although, the molecular mechanism whereby *MICA* polymorphisms confer the risk of disease progression should be characterized in the future, our findings reveal a crucial role of genetic variations in the host innate immune system in the development of HCV-induced HCC.

URLs. R, http://cran.r-project.org/; PLINK, http://pngu.mgh.harvard.edu/~purcell/plink/; Primer3 v0.3.0, http://frodo.wi.mit.edu/primer3/; LocusZoom, http://csg.sph.umich.edu/locuszoom/; FastSNP, http://fastsnp.ibms.sinica.edu.tw/pages/input_CandidateGeneSearch.jsp.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturegenetics/.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We would like to thank all the subjects and the members of the Rotary Club of Osaka-Midosuji District 2660 Rotary International in Japan who donated their DNA for this work. We also thank the technical staff of the Laboratory for Genotyping Development, Center for Genomic Medicine, RIKEN, and the Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, the University of Tokyo. This work was conducted as a part of the BioBank Japan Project that was supported by the Ministry of Education, Culture, Sports, Science and Technology of the Japanese government.

AUTHOR CONTRIBUTIONS

K.M. and Y.N. conceived of the study; Y.N., V.K., M.K. and K.M. designed the study; V.K., Y.U., R.M. and N.H. performed genotyping; V.K., Y.N. and K.M. wrote the manuscript; A.T. and N. Kamatani performed quality control at the genome-wide phase; Y.N., K.M., H.N. and M.K. managed DNA and serum samples belonging to BioBank Japan; N. Kato, R.T., M. Otsuka, M. Omata and K.K. managed replication DNA and serum samples; V.K. analyzed the data, performed VNTR genotyping, ELISA and summarized the whole results; Y.N. obtained funding for the study.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/naturegenetics/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. Global Burden of Hepatitis C Working Group. Global burden of disease (GBD) for hepatitis C. *J. Clin. Pharmacol.* **44**, 20–29 (2004).
2. Parkin, D.M., Bray, F., Ferlay, J. & Pisani, P. Global cancer statistics, 2002. *CA Cancer J. Clin.* **55**, 74–108 (2005).
3. Umemura, T., Ichijo, T., Yoshizawa, K., Tanaka, E. & Kiyosawa, K. Epidemiology of hepatocellular carcinoma in Japan. *J. Gastroenterol.* **44** (Suppl 19), 102–107 (2009).
4. Vong, S. & Bell, B.P. Chronic liver disease mortality in the United States, 1990–1998. *Hepatology* **39**, 476–483 (2004).
5. Ge, D. *et al.* Genetic variation in *IL28B* predicts hepatitis C treatment-induced viral clearance. *Nature* **461**, 399–401 (2009).
6. Thomas, D.L. *et al.* Genetic variation in *IL28B* and spontaneous clearance of hepatitis C virus. *Nature* **461**, 798–801 (2009).
7. Rauch, A. *et al.* Genetic variation in IL28B is associated with chronic hepatitis C and treatment failure: a genome-wide association study. *Gastroenterology* **138**, 1338–1345, 1345.e1–7 (2010).
8. Kamatani, Y. *et al.* A genome-wide association study identifies variants in the *HLA-DP* locus associated with chronic hepatitis B in Asians. *Nat. Genet.* **41**, 591–595 (2009).
9. Schütte, K., Bornschein, J. & Malfertheiner, P. Hepatocellular carcinoma–epidemiological trends and risk factors. *Dig. Dis.* **27**, 80–92 (2009).
10. Tanaka, T. *et al.* Significance of specific antibody assay for genotyping of hepatitis C virus. *Hepatology* **19**, 1347–1353 (1994).
11. Tanaka, Y. *et al.* Genome-wide association of *IL28B* with response to pegylated interferon-alpha and ribavirin therapy for chronic hepatitis C. *Nat. Genet.* **41**, 1105–1109 (2009).
12. Anonymous. Complete sequence and gene map of a human major histocompatibility complex. The MHC sequencing consortium. *Nature* **401**, 921–923 (1999).
13. de Bakker, P.I. *et al.* A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. *Nat. Genet.* **38**, 1166–1172 (2006).
14. Kuniholm, M.H. *et al.* Specific human leukocyte antigen class I and II alleles associated with hepatitis C virus viremia. *Hepatology* **51**, 1514–1522 (2010).
15. Wang, J.H. *et al.* Ethnic and geographical differences in HLA associations with the outcome of hepatitis C virus infection. *Virol. J.* **6**, 46 (2009).
16. Singh, R., Kaul, R., Kaul, A. & Khan, K. A comparative review of HLA associations with hepatitis B and C viral infections across global populations. *World J. Gastroenterol.* **13**, 1770–1787 (2007).
17. Dixon, A.L. *et al.* A genome-wide association study of global gene expression. *Nat. Genet.* **39**, 1202–1207 (2007).
18. Ge, D. *et al.* Genetic variation in *IL28B* predicts hepatitis C treatment-induced viral clearance. *Nature* **461**, 399–401 (2009).
19. Bauer, S. *et al.* Activation of NK cells and T cells by NKG2D, a receptor for stress-inducible MICA. *Science* **285**, 727–729 (1999).
20. Salih, H.R., Rammensee, H. & Steinle, A. Cutting edge: down-regulation of MICA on human tumors by proteolytic shedding. *J. Immunol.* **169**, 4098–4102 (2002).
21. Waldhauer, I. *et al.* Tumor-associated MICA is shed by ADAM proteases. *Cancer Res.* **68**, 6368–6376 (2008).
22. Jinushi, M. *et al.* Impairment of natural killer cell and dendritic cell functions by the soluble form of MHC class I-related chain A in advanced human hepatocellular carcinomas. *J. Hepatol.* **43**, 1013–1020 (2005).
23. Groh, V., Wu, J., Yee, C. & Spies, T. Tumour-derived soluble MIC ligands impair expression of NKG2D and T-cell activation. *Nature* **419**, 734–738 (2002).
24. Doubrovina, E.S. *et al.* Evasion from NK cell immunity by MHC class I chain-related molecules expressing colon adenocarcinoma. *J. Immunol.* **171**, 6891–6899 (2003).
25. Kohga, K. *et al.* Serum levels of soluble major histocompatibility complex (MHC) class I-related chain A in patients with chronic liver diseases and changes during transcatheter arterial embolization for hepatocellular carcinoma. *Cancer Sci.* **99**, 1643–1649 (2008).
26. Jinushi, M. *et al.* Expression and role of MICA and MICB in human hepatocellular carcinomas and their regulation by retinoic acid. *Int. J. Cancer* **104**, 354–361 (2003).
27. Groh, V. *et al.* Broad tumor-associated expression and recognition by tumor-derived gamma delta T cells of MICA and MICB. *Proc. Natl. Acad. Sci. USA* **96**, 6879–6884 (1999).
28. Ota, M. *et al.* Trinucleotide repeat polymorphism within exon 5 of the MICA gene (MHC class I chain-related gene A): allele frequency data in the nine population groups Japanese, Northern Han, Hui, Uygur, Kazakhstan, Iranian, Saudi Arabian, Greek and Italian. *Tissue Antigens* **49**, 448–454 (1997).
29. Groh, V. *et al.* Costimulation of CD8αβ T cells by NKG2D via engagement by MIC induced on virus-infected cells. *Nat. Immunol.* **2**, 255–260 (2001).
30. Groh, V. *et al.* Cell stress-regulated human major histocompatibility complex class I gene expressed in gastrointestinal epithelium. *Proc. Natl. Acad. Sci. USA* **93**, 12445–12450 (1996).

ONLINE METHODS

Sample collections. We obtained DNA from 721 HCV-related HCC cases, 1,730 CHC cases and 5,486 HCV-negative controls from the BioBank Japan project[31]. For replication analysis, DNA from 673 HCV-induced HCC cases was obtained from a prospective HCC study cohort of the University of Tokyo. A diagnosis of CHC, liver cirrhosis or HCC were based on histological, clinical and laboratory findings obtained by trained physicians. Case samples with HBV co-infection were excluded from the analysis. Interferon was administrated to 20.4% of HCC cases and 70.1% of cases were not treated. The remaining 9.5% of the cases lacked information about interferon treatment. The non-HCV controls obtained from BioBank Japan contained case-mixed individuals after excluding all individuals with cancer, chronic hepatitis B, diabetes or tuberculosis. All subjects were of Japanese origin and provided written informed consent. The clinical and demographic details of the samples are summarized in **Supplementary Table 1**. We also obtained serum samples from BioBank Japan and the University of Tokyo (**Supplementary Table 12**). This research project was approved by the ethical committees of the University of Tokyo and RIKEN.

SNP genotyping and quality control. In the GWAS, 721 individuals with HCV-related liver cancer and 2,890 controls were genotyped using Illumina HumanHap610-Quad and Illumina HumanHap550v3 Genotyping BeadChip, respectively. In the replication stage, 673 cases with HCV-related disease, 1,730 cases with CHC and 2,596 controls were genotyped by the multiplex PCR-based Invader assay (Third Wave Technologies) and the Illumina HumanHap610-Quad, respectively. The common SNPs between the Illumina HumanHap550v3 and the Illumina HumanHap610-Quad arrays from all autosomal chromosomes were included for the analysis. We applied standard SNP quality control filters to exclude SNPs with low call rate (<99%), a Hardy-Weinberg equilibrium $P < 1.0 \times 10^{-6}$ for controls and minor allele frequency of <0.01. In the end, we obtained 432,703 SNPs for the analysis. In the replication analysis, the allele discrimination plots were validated by two well-trained researchers (the plots are available on request). We excluded samples with low genotyping rate (<99%) and employed principal component analysis to avoid the population stratification issue, in which individuals belonging only to Hondo cluster were included in the analysis (**Supplementary Fig. 10**)[32].

Statistical analysis. The association of SNPs with the disease phenotype in the GWAS, replication stage and combination analyses was tested using multivariate logistic regression analysis after adjusting for age at recruitment (continuous) and gender by assuming an additive model and using PLINK[33]. In the GWAS, the genetic inflation factor (λ) was derived by applying logistic regressed P values for all the tested SNPs. The quantile-quantile plot was drawn using R. The ORs were calculated by considering the major allele as a reference, unless it was stated otherwise elsewhere. The combined analysis of the GWAS and replication stage was verified by conducting the Mantel-Haenszel method. We considered $P < 5 \times 10^{-8}$ as the genome-wide significance threshold, which is the Bonferroni-corrected threshold for the number of independent SNPs genotyped in HapMap Phase 2 (ref. 34). Heterogeneity across the two stages was examined by using the Breslow-Day test[35].

For multiple logistic regression analysis at rs2596542 using the R program, we considered age at recruitment (≤60 or >60 years)[3], gender (male or female) and alcohol consumption (non-drinkers, ≤ 50 g alcohol per day or >50 g alcohol per day) as covariates from both the GWAS and replication stage cases with HCC and non-HCV controls. Association at the *MICA* VNTR locus was analyzed by Fisher's exact test, and the global P value was calculated using a χ^2 test. Statistical comparisons between genotypes and sMICA levels were performed by Kruskal-Wallis test or Wilcoxon rank test using R. We employed the R package haplo.stats to infer haplotypes and to perform haplotype association analysis. P values for association between sMICA levels and haplotype distribution were obtained by score test under an additive model by using the haplo.score function. ORs and 95% confidence intervals were calculated from the coefficients of the GLM model by considering the major haplotype as a reference. We used the haplo.cc function to calculate these statistical values.

HCV serotype. HCV serotype data was available for 531 cases with HCC from the replication stage. HCV serotype was examined by serotyping assay (SRL Laboratory) according to previously reported methods[36]. According to the Simmonds classification[37], serotype 1 corresponded to disease types 1a and 1b, whereas serotype 2 corresponded to disease types 2a and 2b.

***MICA* VNTR locus genotyping.** We followed the method suggested by Applied Biosystems. Briefly, the 5′ end of the forward primer was labeled with 6-FAM, and the 5′ end of reverse primer was labeled with the GTGTCTT non-random sequence to promote addition of As. The primer sequences were previously reported[28]. The PCR products were mixed with Hi-Di Formamide and GeneScan-600 LIZ size standard and separated using a GeneScan system on a 3730xl DNA analyzer (Applied Biosystems). GeneMapper software (Applied Biosystems) was used to assign the repeat fragment size (**Supplementary Fig. 7**).

Quantification of soluble MICA. sMICA levels were measured by sandwich enzyme-linked immunosorbent assay, as described in the manufacturer's instructions (R&D Systems).

Imputation and association analysis at HLA allele tagging SNPs. We obtained a SNP or a combination of SNPs which can tag HLA alleles in the Japanese population from a previous study[13]. The untyped genotypes of these SNPs were imputed in the GWAS samples by using a hidden Markov model programmed in MACH[38] and haplotype information from HapMap JPT samples. We applied the same SNP quality criteria as in the GWAS for selecting SNPs for the analysis. The association was tested on all SNPs that passed the quality control criteria using logistic regression analysis conditioned on age and gender.

Initially, we obtained the pair-wise LD between HLA alleles tagging SNPs and rs2596542. We performed case-control association analysis in our GWAS dataset. As shown in **Supplementary Table 9**, none of the HLA-tagging SNPs showed evidence of linkage or association except rs2844521, and rs2844521 was in absolute linkage with rs2596542 ($r^2 = 1$, $D' = 1$) and thus showed similar association. We obtained actual genotype data at rs2596501, as this SNP is included on the 550K SNP platform, and inferred the haplotype between rs2844521 and rs2596501. However, the haplotype GT (the G allele of rs2844521 and the T allele of rs2596501), which is reported to tag the *HLA-B*3501* allele ($r^2 = 1$, $D' = 1$), was not associated with HCC in our GWAS dataset ($P = 0.39$). We also performed a conditional logistic regression analysis on rs2596501 (data not shown) and found no effect on the association between rs2596542 and HCV-induced HCC. This data suggested that rs2596542 association is independent of *HLA-B*3501*. Although we observed mild association between other *HLA-B* alleles (*HLA-B*5401*, $P = 0.004$; *HLA-B*6701*, $P = 0.012$) and HCV-induced HCC, the association at rs2596542 alone was the most significant. Taken together, we found no strong evidence for linkage of HLA alleles with rs2596542.

Software. For general statistical analysis, we used R statistical environment version 2.6.1 or plink version 1.06. The Haploview software version 4.2 (ref. 39) was used to calculate LD and to draw Manhattan plots. Primer3 v0.3.0 web tool was used to design primers. We used LocusZoom for plotting regional association plots. We used FastSNP[40] web tool for functional annotation of SNPs (see URLs for all software packages).

31. Nakamura, Y. The BioBank Japan Project. *Clin. Adv. Hematol. Oncol.* **5**, 696–697 (2007).
32. Yamaguchi-Kabata, Y. *et al.* Japanese population structure, based on SNP genotypes from 7003 individuals compared to other ethnic groups: effects on population-based association studies. *Am. J. Hum. Genet.* **83**, 445–456 (2008).
33. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
34. Frazer, K.A. *et al.* A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851–861 (2007).
35. Breslow, N. & Day, N. Statistical methods in cancer research. Volume II–The design and analysis of cohort studies. *IARC Sci. Publ.* 1–406 (1987).
36. Tsukiyama-Kohara, K. *et al.* A second group of hepatitis C viruses. *Virus Genes* **5**, 243–254 (1991).
37. Simmonds, P. *et al.* Identification of genotypes of hepatitis C virus by sequence comparisons in the core, E1 and NS-5 regions. *J. Gen. Virol.* **75**, 1053–1061 (1994).
38. Scott, L.J. *et al.* A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* **316**, 1341–1345 (2007).
39. Barrett, J.C., Fry, B., Maller, J. & Daly, M.J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263–265 (2005).
40. Yuan, H.Y. *et al.* FASTSNP: an always up-to-date and extendable service for SNP function analysis and prioritization. *Nucleic Acids Res.* **34**, W635–W641 (2006).

Liver INTERNATIONAL
OFFICIAL JOURNAL OF THE INTERNATIONAL ASSOCIATION FOR THE STUDY OF THE LIVER

Liver International ISSN 1478-3223

CLINICAL STUDIES

Applying a system approach to forecast the total hepatitis C virus-infected population size: model validation using US data

David Kershenobich[1], Homie A. Razavi[2], Curtis L. Cooper[3], Alfredo Alberti[4], Geoffrey M. Dusheiko[5], Stanislas Pol[6], Eli Zuckerman[7], Kazuhiko Koike[8], Kwang-Hyub Han[9], Carolyn M. Wallace[2], Stefan Zeuzem[10] and Francesco Negro[11]

1 Department of Experimental Medicine, School of Medicine, General Hospital of México, Universidad Nacional Autonoma de México, México City, México
2 Center for Disease Analysis, Kromite, Louisville, CO, USA
3 Department of Medicine, Division of Infectious Diseases, University of Ottawa, Ottawa, ON, Canada
4 Department of Histology, Microbiology and Medical Biotechnologies, University of Padova, Padova, Italy
5 Centre for Hepatology, Royal Free Hospital, London, UK
6 Inserm U-1016 et Unité d'Hépatologie, Université Paris Descartes, Assistance Publique Hôpitaux de Paris, Paris, France
7 Liver Unit, Carmel Medical Center, Clalit Healthcare Services, Haifa, Israel
8 Department of Gastroenterology, Graduate School of Medicine, University of Tokyo, Tokyo, Japan
9 Department of Internal Medicine, Yonsei University College of Medicine, Seoul, Korea
10 Medizinische Klinik 1, Klinikum der Johann Wolfgang Goethe-Universität, Frankfurt am Main, Germany
11 Services de Gastroenterologie et d'Hepatologie et de Pathologie Clinique, Hopitaux Universitaires, Genève, Switzerland

Keywords
diagnosis – disease burden – epidemiology – HCV infection – hepatitis C – incidence – mortality – prevalence – sensitivity analysis – system modelling – treatment rate

Abbreviations
CDC, Center for Disease Control; HCC, hepatocellular carcinoma; HCV, hepatitis C virus; I-C3, International Conquer C Coalition; IDU, injection drug use; MODA, multi-objective decision analysis; NHANES, The National Health and Nutrition Examination Survey; SVR, sustained viral response.

Correspondence
Homie A. Razavi, Center for Disease Analysis, Kromite, 901 Front Street, Suite 291, Louisville, CO 80027, USA
Tel: +1 720 890 4848
Fax: +1 303 552 9119
e-mail: hrazavi@kromite.com

Received 2 July 2010
Accepted 22 March 2011

DOI:10.1111/j.1478-3231.2011.02535.x

Abstract
Background: Hepatitis C virus (HCV) infection is associated with chronic progressive liver disease. Its global epidemiology is still not well ascertained and its impact will be confronted with a higher burden in the next decade. *Aim:* The goal of this study was to develop a tool that can be used to predict the future prevalence of the disease in different countries and, more importantly, to understand the cause and effect relationship between the key assumptions and future trends. *Methods:* A system approach was used to build a simulation model where each population was modeled with the appropriate inflows and outflows. Sensitivity analysis was used to identify the key drivers of future prevalence. *Results:* The total HCV-infected population in the US was estimated to decline 24% from 3.15 million in 2005 to 2.47 million in 2021, while disease burden will increase as the remaining infected population ages. During the same period, the mortality rate was forecasted to increase from 2.1 to 3.1%. The diagnosed population was 50% of the total infections, while less than 2% of the total infections were treated. *Conclusion:* We have created a framework to evaluate the HCV-infected populations in countries around the world. This model may help assess the impact of policies to meet the challenges predicted by the evolution of HCV infection and disease. This prediction tool may help to target new public health strategies.

Hepatitis C virus (HCV) infection is associated with chronic, progressive liver disease. Chronic hepatitis C is a leading cause of cirrhosis and hepatocellular carcinoma (HCC) (1, 2), and the latter two are a major indication for liver transplantation (3). A better understanding of HCV infection prevalence and its modelling can help medical communities and government agencies manage the disease burden and develop treatment strategies in light of the emergence of several potent anti-HCV therapies.

Considerable work has been undertaken to forecast the epidemiology of hepatitis C and numerous authors have developed Markov models to estimate the disease burden (4–16). However, these models have limitations resulting from the assumption that the studied cohorts remain homogeneous and traverse through different states at a fixed rate over time when, in reality, populations are very heterogeneous (17). Other investigators have developed multicohort natural history models (17), but all forecasts are complicated by availability and uncertainty of key inputs, especially outside a few well-studied countries.

We addressed this problem by using a system approach to build a simulation model. In this method, each population group

was assessed as a dynamic system with inflows and outflows. When the inflows of population outweighed the outflows, the size of the group increased. Conversely, the size of the population decreased when outflows were higher than inflows. The benefit of this approach was that it did not require accurate information, except for the relative size of populations moving in and out to predict trends. In addition, it enabled us to calculate the size of infected populations when data were missing.

The merits of this approach were verified taking as an example the HCV-infected population in USA, where considerable research has been carried out to understand the historical and future trends. These results allowed us to calibrate our model in order to subsequently apply it to other geographical regions throughout the world.

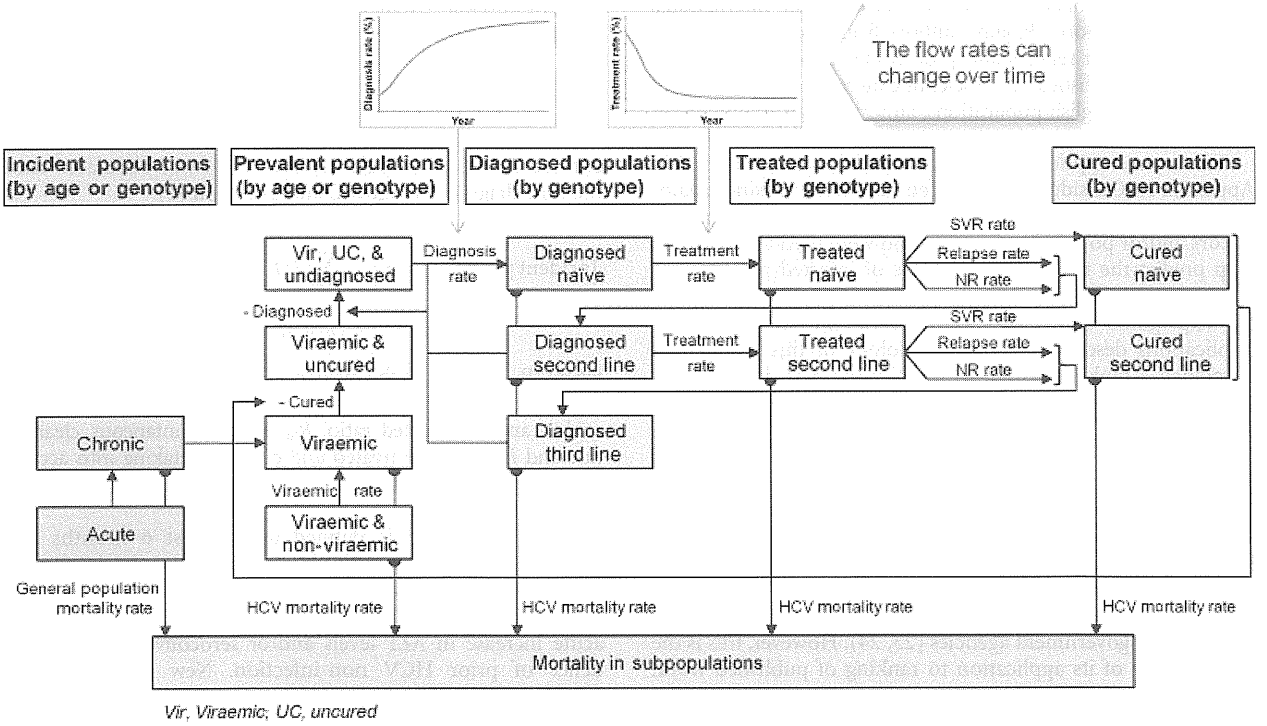
The objective of this paper was to present our methodology and demonstrate the validity of the simulation model using US data. Subsequent publications will describe the application of this methodology and tools to select countries. This work reports the results of one particular scenario: current trends in prevalence, incidence, mortality, diagnosis, treatment and response rates will continue at the same rate. It does not take into account future events like the introduction of new therapies or a significant increase in treatment rates. Our model is designed to analyse numerous scenarios, but this is beyond the scope of this particular article.

Methodology

A number of techniques were used to develop a simulation model and estimate the future size of the HCV-infected population. A system dynamic modelling framework was used to construct the model in Microsoft Excel®. Individual populations (incidence, diagnosed, treated, etc.) were handled as stocks while transitions from one population to another were treated as flows with an associated rate/probability (Fig. 1). Each HCV-infected pool was assessed individually and then linked. To build the model, two tools were used: influence diagrams and flow diagrams (18). Influence diagrams were used to identify all the factors that influenced the HCV-infected pool under consideration and identify the potential key drivers that should be considered in the model. The flow diagram was used to describe the inflow and outflow of infected individuals and to build the final model. Not all factors that were identified in the influence diagram were explicitly incorporated into the model. However, efforts were made to incorporate all factors when gathering data for inputs to the model.

All transition rates/probabilities were modelled as functions to allow change over time. The parameters used to define these functions were start value, end value, start date, years to end value and the curve type, where the curve type value selected from a range of concave to convex sigmoidal curves. Excel® optimization add-in, Solver, was used to calculate the parameters used in the rate functions to fit historical data and predict future trends by minimizing the root mean square value between the actual and the calculated values.

Hepatitis C virus-infected pools were calculated every year given last year's value along with the inflow and outflow rates. The model estimated the infected pools from 2004 to 2021. Reported data from 2004 to 2009 were used to calibrate the rate functions and estimate future trends. Sensitivity analysis was used to identify the transition rates that had the largest impact on future trend as measured by the size of 2021 infected population. The associated infected populations were segmented to enhance



Vir, Viraemic; UC, uncured

Fig. 1. Simplified schematic of the simulation model.

the fit with historical data and the projections. The most significant improvement in the model was observed after the incident, and prevalent populations were segmented by age and gender. Segmenting the populations by genotype also improved fit with historical data, given the response rate and duration of treatment variations in each segment. Further segmentation of the populations by the lines of treatment (naïve – never been treated before; second line – been through one round of treatment; third line – been through two rounds of treatment) had a marginal impact on the fit with historical data. However, it did enable us to forecast the impact of new treatments, as they reach the market, on the size of infected pools.

The model was set up for Bayesian analysis of assumptions with a large uncertainty using Crystal Ball®, an Excel® add-in by Oracle®. Crystal Ball® was also used to run sensitivity analyses. A β-PERT distribution was used for all the rate function parameters using the likeliest value along with 5th and 95th percentile inputs. A β-PERT allowed for non-symmetrical distributions based on the collected data. We assumed that ranges from the literature typically provided a 90% confidence interval. The low end of the range was used as the 5th percentile inputs and the high end of the range was used for the 95th percentile.

A comprehensive global review of the literature was used to gather the size and flow of HCV-infected pools around the world. Data references were identified through two sources: indexed journals and non-indexed sources. Indexed articles were found by searching PubMed and regional databases (e.g. Medigraphic, and Imbiomed in Latin America) using the following terms: 'hepatitis C AND country name AND (incidence OR prevalence OR mortality OR viremia OR genotype OR diagnosis OR treatment OR SVR)'. Furthermore, references cited within the articles were used. More than 27 000 abstracts/full-text articles were reviewed and 2600 references were selected based on relevance. In addition, non-indexed sources were identified through searches of individual countries' ministry of health websites and international health agency reports. Finally, authors from each country provided government reports and proceedings of local conferences that were not published in the scientific literature. The search was not limited to English publications, although they accounted for over 90% of the data sources.

The key assumptions used in our model are summarized in Appendix A. Outside of USA, when data were missing, analogues were used. Data from other countries with similar risk factors and/or population composition were used as a proxy to help predict the future trends in size of infected populations. However, gaps in data in each country were also highlighted to help guide future research. Given the extent of the publications describing the epidemiology of this disease, we developed a ranking system to identify the best data sources. The scoring system did not assess the merits or the quality of the data sources. It simply measured their value for this specific analysis.

A systematic process using multi-objective decision analysis (MODA) (19–22) was used to select the most appropriate data sources, identifying a group of data sources that provided a range for the assumption under consideration rather than relying on a single source. MODA has been used for years to prioritize activities and spending within the pharmaceutical industry and government agencies (23, 24). However, this is the first example of its application to ranking of published work. The key objectives were defined along with a scoring system for each measure. A 0–10 scale along with definitions was developed in which each score indicated the relative contribution to meeting the specific objective. Thus, an article with a score of 10 was twice as valuable as another article with a score of 5 for the same measure, all other measures being equal. The scoring was conducted by the authors, who were familiar with the country-specific disease dynamics and treatment, or by independent epidemiologists.

The weighting for each objective provided the relative importance of that objective as compared with others. A swing weight method was used to estimate the weights (25). Two factors were considered to determine the weight: the importance of the individual measure to the overall objective and the range of scores. Thus, a particular measure that may have been important was weighted less if all data sources examined had very similar scores and the measure could not be used to differentiate one source from another. The overall MODA framework is summarized in Table 1. The data sources with the highest overall score were selected in which the overall score was calculated, as shown in Equation (1):

$$\begin{aligned}\text{Overall score} = (&w_{\text{sample size}} \times score_{\text{sample size}} + w_{\text{extrapolative}} \\ &\times score_{\text{extrapolative}} + w_{\text{analysis type}} \\ &\times score_{\text{analysis type}}) \times score_{\text{relevance}}\end{aligned} \quad (1)$$

The relevant score was used to eliminate articles that did not provide sufficient details for this exercise. Not all articles required scoring. For example, articles describing the risk factors that lead to HCV infection were typically descriptive and did not lend themselves to a scoring system.

The MODA approach was validated by having David Kershenobich score 50 references from Mexico and Stefan Zeuzem score 70 references from Germany. Their ranking was compared with the rating of an independent epidemiologist, Carolyn Wallace. Although there were variations in the absolute scores, the relative rankings had a correlation of 0.98. Further comparisons of ranking of data sources in Canada, Switzerland, Japan, South Korea and Israel provided the same result.

Incident population

We defined incident population as individuals with new chronic and viraemic HCV infections. It was calculated using the flow diagram in Figure 2 and Equation (2):

$$\begin{aligned}\text{Incident population} = &\left[\sum_{\text{Age \& Gender}} ((IR_{\text{Age \& Gender}})(Pop_{\text{Age \& Gender}}))\right] \\ &\times (R_{\text{UU}})(1 - R_{\text{SC}})(1 - R_{T\times C})\end{aligned} \quad (2)$$

where $IR_{\text{Age \& Gender}}$ is the incidence rate by age and gender, $Pop_{\text{Age \& Gender}}$ is the country's population by age and gender, R_{UU} is under- and unreported ratio, R_{SC} is the spontaneous clearance rate and $R_{T\times C}$ is the treated and cured rate taking into account the number of treated patients and the sustained viral response rate.

The acute phase is defined as the first 6 months after infection and is characterized by a sharp increase in the ALT levels and seroconversion (26–28). The US Centers for Disease Control and Prevention (CDC) requires centres to couple an acute increase in ALT levels and/or seroconversion with evidence of prior HCV non-infection. New acute cases are reported to the CDC through sentinel counties and are adjusted for under-reporting. A multiplier is then applied to account for all asymptomatic cases that are not diagnosed or reported. The

Table 1. Multi-objective framework for selecting data sources

Weight (%)	Objective	Rating	Definition
20	Sample size	0–10	Log (sample size) scaled from 0 to 10 maximum at 10k
35	Extrapolative	0	Analysis based on select subpopulations (e.g. IDU)
		4	Analysis based on random but self-selected populations (e.g. blood banks)
		10	Analysis based on a random general population
25	Analysis type	0	Analysis completed before blood screening was implemented
		5	Analysis completed after blood screening was implemented using first- or second-generation screening technology
		10	Analysis completed after blood screening was implemented using the latest screening technology
20	Analysis consistency/quality	0	Analysis completed at a small local research facility
		5	Analysis completed at a hospital lab/research lab
		10	Analysis completed by a national/international reference lab
NA	Relevance	0	Analysis does not provide enough details to be relevant for this exercise (e.g. article breaks out genotypes as G1 and all others)
		1	Analysis provides sufficient details to be relevant to this exercise (e.g. article breaks out specific genotypes: G1, G2, G3, G4, G5, G6 . . .)

IDU, injection drug use.

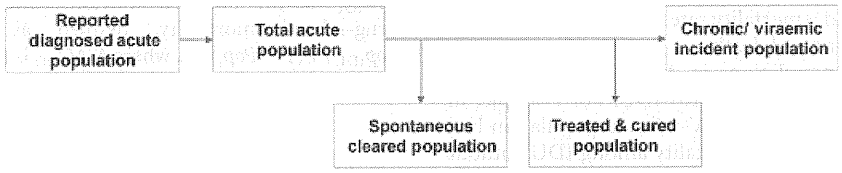


Fig. 2. Hepatitis C virus incident population flow diagram.

Table 2. Center for Disease Control's estimate of the number of new hepatitis C virus infections in USA (37)

	2002	2003	2004	2005	2006	2007
Number of acute clinical cases reported	1223	891	758	694	802	849
Estimated number of acute clinical cases	4800	4500	4200	3400	3200	2800
Estimated number of new infections	29 000	28 000	26 000	21 000	19 000	17 000

unreported multiplier was estimated for USA in a study by Armstrong *et al.* (29).

To estimate true incidence (defined as chronic and viraemic), the number of acute infections spontaneously cleared was extracted. Additionally, those successfully treated during the acute phase of infection had to be accounted. In reality, treatment during the acute phase of infection was estimated to represent a very small proportion of all acute cases, even though there is ample evidence of a high response rate to therapy at this stage of the disease (30–33).

The literature reported that approximately 75–85% of the newly infected persons develop chronic infections (28, 34, 35). The factors that seem to impact the persistence of the infection are an older age at infection, male gender, African-American race, immunosuppressed state, specific HLA subtypes/polymorphisms and blunted innate immune response (28).

For the US model, we used the 2004 CDC reported incidence rates by age and gender and assumed that it remained constant (36). The under- and unreported ratio was calculated from data from the same agency (Table 2) by dividing the annual number of new infections by the number of reported cases (37). This rate was maintained constant throughout the duration of the model. The UN estimates were used for future US population by age and gender (38). A spontaneous clearance rate of 82% was used (28) and it was assumed that the treated and cured rate in the acute phase of infection was zero.

Mortality

We defined mortality as all-cause mortality in order to account for the appropriate portion of all infected individuals who were removed from the infected pool each year. Outside of a handful of countries, little data were available on mortality among HCV-infected individuals. A model was developed to estimate the mortality rate in different countries based on age, liver-related deaths due to HCV infection and per cent of the prevalent population infected by injection drug users (IDU) and transfusion. Although other risk factors also contribute to HCV infections (nosocomial, unsafe injections, etc.), it was assumed that the associated incremental increase in mortality was negligible.

Studies by Duberg *et al.* (39), McDonald *et al.* (40) and Amin *et al.* (41) all show an increase in liver-related deaths in HCV-infected populations. Through personal communications with the authors, the liver-related mortality rates in these studies were compared with the general population mortality in the

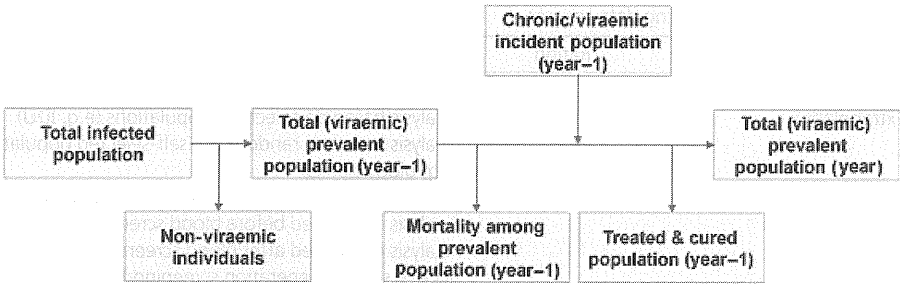


Fig. 3. Prevalent population flow diagram.

same country and age cohorts (42, 43). A marked increase in liver-related deaths was observed starting at ages 40–45. The standard mortality ratio (SMR) was > 1 as compared with the general population until age 70 when mortality in the general population became higher than liver-related deaths. Seeff *et al.* (44) and Guiltinan *et al.* (45) studies among individuals who have been screened for high-risk behaviour (e.g. blood donors and military recruits) showed a SMR of 1.5–2.1. For the purpose of our mortality model, we assumed that all HCV-infected individuals had a mortality rate of 2.0 between the ages 45 and 70 due to liver-related deaths.

The HCV-infected populations in Australia, Scotland and Sweden were heavily weighted towards IDU. For example, in Scotland, 90% of the diagnosed HCV-infected population had injected drugs (14). Analysis of mortality among IDU without HCV infections who received treatment for drug misuse shows an excess mortality ratio of 22.0 between the ages 15 and 44 (46). All three retrospective studies showed a marked increase in drug-related mortality between these same age groups as well. The SMR for drug-related deaths among HCV-infected was 19.3 in Australia, 25.1 in Scotland and 20.7 in Sweden, the latter being specific to deaths due to mental causes due to alcohol and drugs. Even though the HCV infection was not the cause of death in these young age groups from an epidemiological perspective, these individuals were still excluded from the total infected population. In our mortality model, we assumed that the portion of the infected population that was made up of IDU had a mortality rate of 22.0 between the ages 15 and 44.

The final risk group that was identified with a higher mortality rate was blood transfusion recipients. A study from Denmark and Sweden examined 1 118 261 transfusion recipients and found the SMR to be 17.6 in the first 3 months, 2.1 one to four years and 1.3 seventeen years after their first transfusion (47). This study did not separate HCV-infected individuals, but showed a higher mortality rate in blood recipients. In our model, we assumed that the portion of the infected population who contracted HCV from blood transfusion had an SMR of 2.1.

The overall mortality rate among HCV-infected populations was calculated using Equation (3):

$$\text{Mortality rate} = \frac{\sum_{\text{Age \& Gender}}[GPMR_{\text{by Age \& Gender}}(MM_{\text{L}})(AM_{\text{IDU}})(AM_{\text{Trans}})(HCV - Infected\ population_{\text{by Age \& Gender}})]}{\text{Total HCV} - \text{Infected population}} \quad (3)$$

where $GPMR_{\text{by Age \& Gender}}$ is the General Population Mortality Rate by age cohort for all causes provided by United Nations' Demographic Yearbook's Table 19 (42) and MM_{L} is the adjusted mortality multiplier for liver-related morbidity and is defined as

$$MM_{\text{L}} \begin{cases} 1 & \text{Age cohort} \leq 44 \\ 2.0 & 45 \leq \text{Age cohort} \leq 70 \\ 1 & 71 \leq \text{Age cohort} \end{cases}$$

AM_{IDU} is the adjusted mortality multiplier for injection drug-related morbidity, defined as $AM_{\text{IDU}} = (MM_{\text{IDU}} \times Pop_{\text{IDU}}) + (1 - Pop_{\text{IDU}})$, where MM_{IDU} is the injection drug user mortality multiplier defined as

$$MM_{\text{IDU}} \begin{cases} 1 & \text{Age cohort} \leq 14 \\ 22 & 15 \leq \text{Age cohort} \leq 44 \\ 1 & 45 \leq \text{Age cohort} \end{cases}$$

and Pop_{IDU} is the proportion of the population who were injection drug users, AM_{Trans} is the adjusted mortality multiplier for transfusion-related morbidity, defined as $AM_{\text{Trans}} = (MM_{\text{Trans}} \times Pop_{\text{Trans}}) + (1 - Pop_{\text{Trans}})$, where MM_{Trans} is the transfusion mortality multiplier (2.1) and Pop_{Trans} is the proportion of the population who contracted the disease from transfusions.

In USA, we assumed that 47% of the infected population had experimented with IDU and 7% of the cases were due to transfusion (27, 48).

Prevalent population

In our analysis, we defined the prevalent population as the total viraemic uncured HCV-infected population. It was calculated according to the flow diagram shown in Figure 3, Equations (4) and (5):

$$2004\ \text{Prevalent population} = \sum_{\text{Age \& Gender}} ((PR_{\text{Age \& Gender}})(2004\ POP_{\text{Age \& Gender}}))(R_{\text{Viraemic}}) \quad (4)$$

where $PR_{\text{Age \& Gender}}$ is the prevalence rate by age and gender, 2004 $Pop_{\text{Age \& Gender}}$ is the US population by age and gender in 2004, and R_{Viraemic} is the viraemic rate.

$$\text{Prevalent population} = \sum_{\text{Age \& Gender}}^{(yr)} (PP_{\text{Age \& Gender}(yr-1)} + IP_{\text{Age \& Gender}(yr-1)} - MP_{\text{Age \& Gender}(yr-1)}) TxC_{(yr-1)} \quad (5)$$

where $PP_{\text{Age \& Gender}(yr-1)}$ is the previous year's viraemic prevalence population by age and gender, $IP_{\text{Age \& Gender}(yr-1)}$ is the previous year's viraemic incident population by age and gender and $MP_{\text{Age \& Gender}(yr-1)}$ is the size of the mortality population among the previous year's prevalent population (*PP*) defined as $MP_{\text{Age \& Gender}(yr-1)} = PP_{\text{Age \& Gender}(yr-1)} \times Mortality\ Rate_{\text{Age \& Gender}(yr-1)}$, and $TxC_{(yr-1)}$ is the number of individuals who were treated and cured in the previous year as defined in the treated population section.

Our model started in 2004. As shown in Equation (4), the 2004 prevalent population was estimated by applying the 1999–2002 National Health and Nutrition Examination Survey (NHANES) prevalence rates by age and gender (49) to the 2004 US population (38) although a recent study suggested that the total number of infections could be 27% higher if subgroups not sampled by NHANES (e.g. homeless, incarcerated, veterans, active military personnel, etc.) were also considered (50). A viraemic rate of 79.7% was applied to estimate the starting population size with HCV RNA (49). After 2004, the prevalent population was calculated each year using the formula shown in Equation (5). The previous year's new cases were added to the last year's total infected pool and the mortality among the same pool was subtracted. Finally, the number of individuals treated and cured was also excluded.

Outside of USA, HCV prevalence data were available for most countries. However, not all data were representative of the country's population, and the years for which estimates were provided differed. Furthermore, HCV prevalence was determined by local practices, risk factors and access to care in particular communities. Thus, extreme care was required in extrapolating the disease burden from a few well-studied data points.

Studies based on first- and second-generation immuno-assay tests provided the upper bound of the potential prevalence. Because of the number of false-positive outcomes in early tests, older HCV prevalence studies overestimated the total population infected with the disease (51).

In many countries, blood bank data were available and used as a basis for the country's HCV prevalence. However, due to selection bias, sero-epidemiological studies using blood donors may not reflect the epidemiological reality (52). In addition, studies in the USA (53) have suggested that direct questioning related to behavioural habits before blood donation results in decreased blood donor HCV prevalence (0.63% → 0.4% between 1991 and 1996). Blood bank data were useful, but likely provided an estimate that was rather low compared with the HCV prevalence in the general population (54).

Diagnosed population

We defined the diagnosed population as the total viraemic individuals who have been diagnosed with HCV infection. Our definition excludes all who were diagnosed and cured or those where were removed due to mortality. It was calculated, as shown in Figure 4 and Equation (6), where this year's diagnosed population started with last year's pool, added newly viraemic

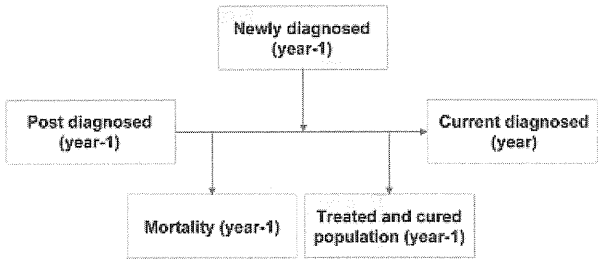


Fig. 4. Diagnosed population flow diagram.

diagnosed patients, and subtracted diagnosed patients who died that year and any who were treated and cured:

$$Dx_{yr} = Dx_{yr-1} + [UnDx_{yr-1} \times NDxR_{yr-1} \times R_{\text{Viraemic}}] - TxC_{(yr-1)} - MP_{yr-1} \quad (6)$$

where Dx_{yr} is the size of the current year's diagnosed population, Dx_{yr-1} is the size of the previous year's diagnosed population, $UnDx_{yr-1}$ is the size of the previous year's undiagnosed population, $NDxR_{yr-1}$ is the previous year's newly diagnosed rate, R_{Viraemic} is the viraemic rate, $TxC_{(yr-1)}$ is the number of previous year's individuals who were treated and cured as described in the treated population section and MP_{yr-1} is the size of last year's mortality population in the diagnosed pool defined as $MP_{yr-1} = Dx_{yr-1} \times Mortality\ Ratio$ where mortality ratio was defined in the mortality section.

In USA, a NHANES survey found that 50% of randomly selected individuals who tested positive for HCV were already aware of their infection (5). Similar studies in subpopulations by NYC HANES, Veterans Affairs' Mental Illness Treatment and Research Center, and Philadelphia Department of Public Health found that between 50 and 72% of those who tested positive to HCV already knew of their infection (55–57), suggesting that a high percentage of the total infected population is already diagnosed in USA.

The newly diagnosed rate was estimated by a recent CDC study, which reported that 69.2 per 100 000 (or 211 600 in 2007) of new diagnosed HCV infections were reported between 2006 and 2007 (58). In our model, we assumed that before 2004, 795 000 viraemic individuals were diagnosed in USA. A viraemic rate of 79.7% was applied to estimate the starting population size with HCV RNA (49).

Treated population

The numbers reported by Volk *et al.* (5) were used for the number of treated individuals in USA. This comprehensive analysis looked at longitudinal data from 2002 to 2007 to estimate the number of treated HCV patients in USA.

Disease burden

We used the number of infected individuals over 60 years old as a proxy for disease burden. A number of studies have evaluated the HCV disease burden as a function of time (7, 14, 17, 59–61), age, gender and risk factors. They forecasted the progression of HCV infection to cirrhosis, and the specific values for complications such as encephalopathy, ascites or variceal haemorrhage (8, 15, 17). Previously unpublished data by Poynard (Fig. 5) showed an increase in cirrhosis and fibrosis progression with age.

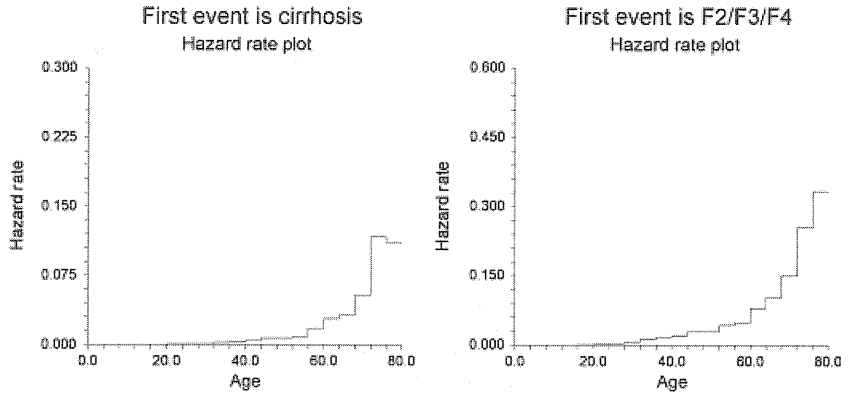


Fig. 5. Progression of fibrosis assessed according to age in 2313 hepatitis C virus patients with a biopsy who were not treated (Thierry Poynard 2010, Université Pierre et Marie Curie Liver Center, Paris Frano pers. comm.).

The methodology described here was designed to be applied to numerous countries around the world. Outside of USA and Western Europe, data were insufficient to estimate the disease burden with any accuracy using more sophisticated modelling. Thus, the age of the infected population was used as a proxy for disease burden. Previous work (62, 63) already demonstrated a marked increase in the probability of disease progression around age 60. For our purposes, we assessed the percentage of infected population over the age of 60 as a surrogate for the disease burden.

Results

A key advantage of the system approach was that it allowed us to quickly assess the impact on key populations without significant programming. To illustrate, consider the change in size of the viraemic infected population between 2007 and 2008. In 2007, we estimated that there were 3 million individuals infected with HCV. In that same year, the CDC estimated that 17 000 new US-based HCV infections occurred (Table 2) (37). With a spontaneous clearance rate of 18% (28, 64), 13 940 went on to have a chronic infection. Volk *et al.* (5) estimated that 83 000 patients were treated in that same year. With a 50% sustained viral response (SVR) (all genotypes combined), this would translate to 41 500 cured persons. For the same year, our mortality model estimated that 61 500 people infected with HCV died in 2007.

The system approach predicted that viraemic prevalence will decline in USA between 2007 and 2008 as shown in the balance beam in Figure 6. The true incidence, mortality and number of cured individuals were all uncertain. However, the conclusion held even if mortality, spontaneous clearance or cured rates were off by more than a factor of two. The likelihood that the true incidence was higher than the sum of mortality and cured individuals was negligible. Thus, it was safe to conclude that the overall viraemic prevalence will decline in USA. The trend will only reverse if there is a significant increase in incidence through a very large outbreak. This analysis showed that in order to estimate the future total infected population size, it was more important to develop a robust mortality model rather than fine-tune incidence estimates. Outside of USA, the reported incidence was most often newly diagnosed, which included both new and existing infections. However, using the above approach, we were able to estimate the number of new cases when the trends in prevalence, mortality and the number of treated and cured populations were available.

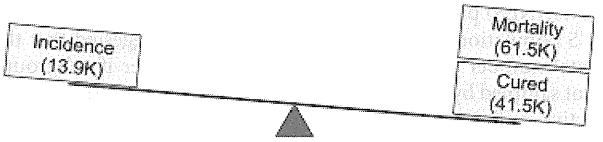


Fig. 6. The 2007 total infected population balance beam for hepatitis C virus in USA.

Prevalent population

The simulation model predicted a decline in the total number of infected individuals in USA in the absence of new events including the utilization of more effective HCV antiviral medications in the future (Fig. 7). The viraemic prevalent population was projected to decline by 24% from 3.15 million in 2005 to 2.47 million in 2021. This trend was consistent with forecasts by Davis *et al.* (17), who modelled the total number of individuals ever infected using a multicohort natural history model. In their model, the total chronic HCV population declined 22% from 3.53 million in 2005 to 2.75 million in 2021. Overall, the true infected population is likely to be higher than the results shown here, because published reports exclude data on incarcerated and homeless persons or active IDU.

The expected decline in prevalence does not mean that the burden of the disease will decline as well. The dichotomy with HCV infection is that as prevalence declines in many countries around the world, the burden of the disease is expected to increase due to an ageing HCV population.

Diagnosed population

We estimated that 50% of the HCV-infected population was diagnosed by 2009 and close to 80% will know of their infection by 2021 (Fig. 7). The literature suggested that approximately half of the infected population already knew of their diagnoses before 2009. However, we started with a more conservative diagnosed population to account for sampling in the studies. Relaxing this assumption would simply accelerate the increase in the size of the diagnosed population in USA.

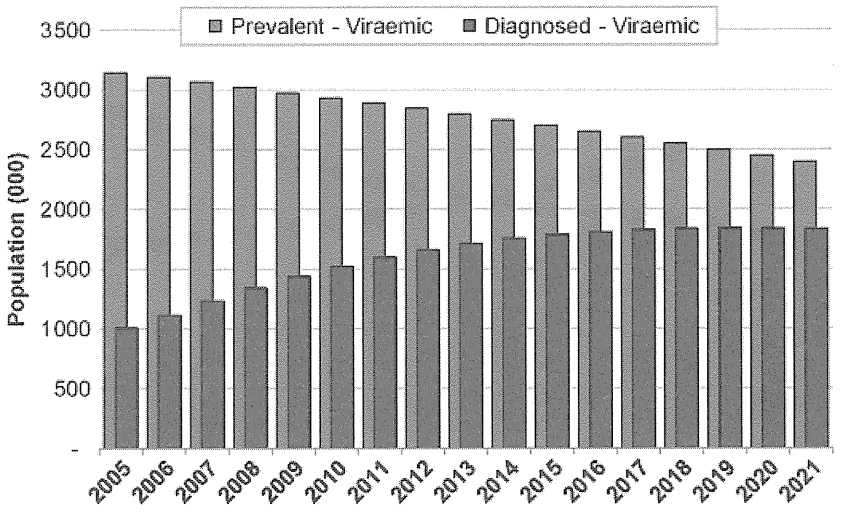


Fig. 7. Projected number of hepatitis C virus-infected and diagnosed individuals in USA.

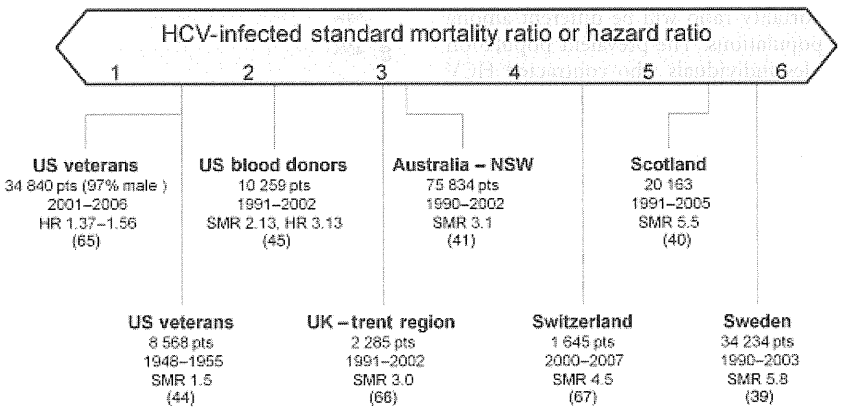


Fig. 8. Select hepatitis C virus-infected mortality studies.

The model predicted a diminishing growth in the size of this population due to fewer undiagnosed individuals available as more are diagnosed and an increase in mortality as the infected population ages.

Mortality

A comparison of select mortality studies is shown in Figure 8 (39–41, 44, 45, 65–67). We used our mortality model to explain the differences in the studies by the risk profile of the infected population in the studies. The output of our model for the US-infected populations is shown in Figure 9. We forecasted that the SMR will decline from 2.5 in 2005 to 1.6 in 2021. SMR was calculated by dividing the forecasted mortality in the HCV-infected population by age and gender by the expected deaths in the general population with the same age and gender profile. During the same period, the mortality rate (total deaths in HCV-infected population/total infected population) was forecasted to increase from 2.1 to 3.1%. Both trends were consistent with an ageing infected population that would cause the mortality rate to increase. At the same time, the SMR declined as the mortality rate in the general population (the denominator in SMR) increased as a result of an ageing population. Our SMRs were similar to previous US studies (range of 1.5–2.1), although the two sets cannot be compared directly because our estimates were for the total infected population (44, 45, 65). In a further validation of the mortality model, when the risk factors and general population mortality rates in Sweden were used, the forecasted SMR was within 12% of the actual numbers reported by Duberg *et al.* (39). The model underestimated the number of deaths among the HCV-infected population between the ages of 45 and 75. This could be due to the fact that we did not consider the impact of alcohol or body mass. However, for most countries without a robust mortality study, this would provide a reasonable approximation.

Our analysis was not consistent with studies by Harris *et al.* (68, 69), where it was shown that the mortality rate among HCV-infected transfusion recipients was the same as the control group after 16 years of follow-up. However, a more recent study from Denmark showed a clear increase in mortality in the HCV-infected population as compared with the cured population in line with our model (70).

It is worth noting that our analysis predicts that the mortality rate will change over time as the infected population

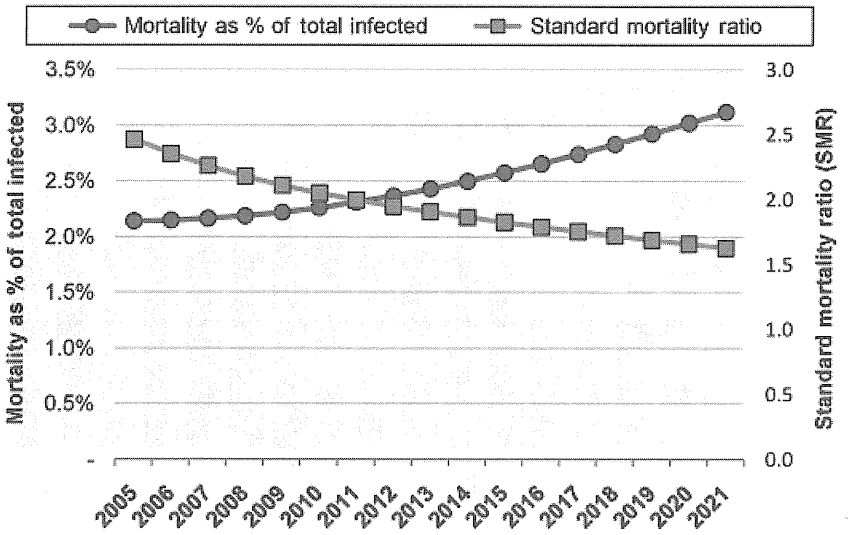


Fig. 9. Projected US standard mortality ratio and mortality rate.

ages. In addition, the mortality ratio will be different among incident and prevalent populations. The prevalent population in most countries includes individuals who contracted HCV from contaminated blood supplies or unintentional contaminated injections. Today, in most western countries, the new cases (incidence) are due to IDU, which means that mortality among newly infected individuals will be significantly higher.

Our analysis showed that it is possible to estimate the mortality rate in a country using the risk factors present. In addition, it predicted a lower SMR in USA as compared with the European studies, consistent with the literature. Finally, it demonstrated that rather diverse reports in the literature are actually consistent when the risk factors are considered.

Disease burden

Consistent with other studies (17, 59), we predict an increase in disease burden over the next decade despite a decreasing prevalence.

A recent study of veterans showed that ageing of the HCV-infected patients accounted for a significant proportion of the observed increase in the prevalence of cirrhosis and HCC, although not all (71). By 2011, 20% of the infected viraemic population will be over 60 (Fig. 10). This population was forecasted to more than double within 10 years. The objective of our analysis was not to calculate the estimated number of individual morbidities associated with hepatitis C, but rather to provide directional insight. According to Davis *et al.* (17), the number of HCV-related cirrhosis cases is expected to increase by 24% in the same period; decompensated cirrhosis cases and HCC will increase by 50%. A retrospective study among veterans reported a significantly higher increase in the burden of HCC and a lower increase in the prevalence of cirrhosis than predicted by mathematical models (71). Directionally, our model was consistent with previously reported estimates.

Treated population

The Volk analysis reported that the number of treated patients had been declining in USA. This trend was observed at a time when the number of diagnosed individuals was increasing. The factors that

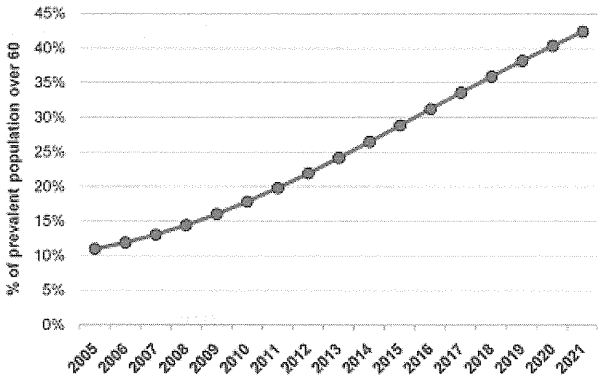


Fig. 10. Per cent of US viraemic hepatitis C virus-infected population over 60 years old.

could explain this observed historical decrease in USA were clinical trials, eligibility, warehousing (physicians holding back patients until new therapies become available), physician perception and the treatment capacity, which may vary for each country.

There were more patients enrolled and treated in clinical trials as new drugs were developed to treat this disease. According to National Institute of Health (http://www.clinicaltrials.gov), in 2007, approximately 10 000 patients were enrolled in the US hepatitis C clinical trials. The Volk study did not include treated patients in trials. Alternatively, as more patients were treated, the remaining pool had a higher percentage of ineligible patients. However, this could not fully explain the downward trend, as the number of newly diagnosed individuals far outweighed the treated patients. The impact of warehousing could explain the downward trend after 2007 as more clinical trial data on new treatment became available. Another likely explanation for the decrease in the number of treated patients is treating physicians' perception of the need to treat given the slow progression of the disease, corroborated by our improved knowledge of the natural history of chronic hepatitis C and other cofactors of fibrogenesis.

Finally, there is a treatment capacity in some countries in which there are not enough physicians and other health care

providers available to treat the infected individuals. This has been observed in parts of Brazil and Eastern Europe. In USA, on the other hand, the medical system can handle at least 144 000 treated patients, the number treated in 2004 (5). Historically, treatment capacity is an unlikely cause of decline in treated patients in USA.

The treatment rate of HCV remains very low. According to our analysis, $< 5\%$ of the diagnosed population and close to 2% of the viraemic prevalent population are treated in USA today. Without an increase in the treatment (and cured) rate, the disease burden is expected to increase.

Sensitivity analysis

As evident from this analysis, forecasting the epidemiology of HCV at a regional or a national level requires numerous assumptions. This is further complicated by the uncertainty in individual inputs. To determine which assumptions have the largest impact on the future prevalence of the disease, sensitivity analysis was used. As described in Figure 11, there are typically two potential explanations for an assumption to be highlighted as important in a sensitivity analysis: (i) high level of uncertainty related to the assumption/input and (ii) output is highly sensitive to small changes in the assumption. In a typical sensitivity analysis, the two are confounded.

Our analysis focused on identifying assumptions that have the largest impact on future prevalence by selecting a fixed range $\pm 10\%$ above and below the base assumption. The assumptions that were varied were all the transition rate/probabilities and starting populations as shown in Figure 12.

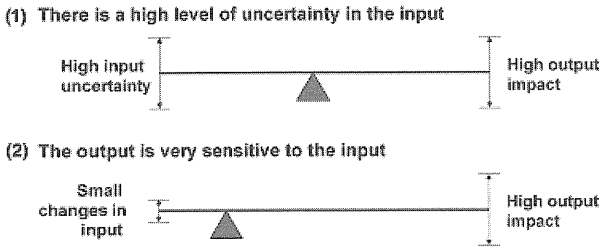


Fig. 11. Reasons for assumptions showing up as important in a sensitivity analysis.

In the base scenario, the 2021 prevalence was calculated (2.47 million) while keeping all assumptions at their base value. One assumption was increased by 10% while keeping all others at their base value, and the new 2021 prevalence was recorded. The same assumption was reduced by 10% and the new 2021 prevalence was recorded again. This was repeated with all assumptions under consideration. In case of mortality, the mortality rate in the general population was changed by $\pm 10\%$, which had the same effect as changing the mortality rate in the HCV-infected pool by the same percentage. The resulting prevalence numbers were divided by the original value to determine per cent change in the 2021 prevalence as a result of $\pm 10\%$ change in each assumption. The assumptions were ranked from the highest to the lowest impact and the results are plotted in Figure 12.

The tornado diagram shown below ranks the assumptions by the magnitude of their impact on 2021 prevalence. When all assumptions were set to their base value (centre line), the forecasted 2021 prevalence was the same as the original value (100%). When the starting prevalence rate of 1.37% was increased by 10% (1.51%) or decreased by 10% (1.23%), the future prevalence increased or decreased by 10% as well. The same behaviour was observed when per cent viraemic (base value of 79.70%) was changed by 10%, implying that future prevalence is very sensitive to both of these assumptions and any change in these assumptions will have an equal impact on the forecasted prevalence.

An interesting observation was the importance of mortality rate on the future prevalence of the disease. When the mortality rate was increased by 10%, the 2021 prevalence decreased by 5%. When it was decreased by 10%, the 2021 prevalence increased by 5%. This implies that in our model, mortality is the third key driver of future prevalence after prevalence rate and per cent viraemic. After mortality, incidence rate and per cent viraemic among the incident population had the largest impact on the 2021 prevalence. Varying all other assumptions (SVR, % genotype 1, diagnosis rate, treatment rate and relapse rate) by $\pm 10\%$ had a $< 2\%$ impact on future prevalence.

Discussion

Although hepatitis C is considered an important cause of chronic liver disease, accurate and representative

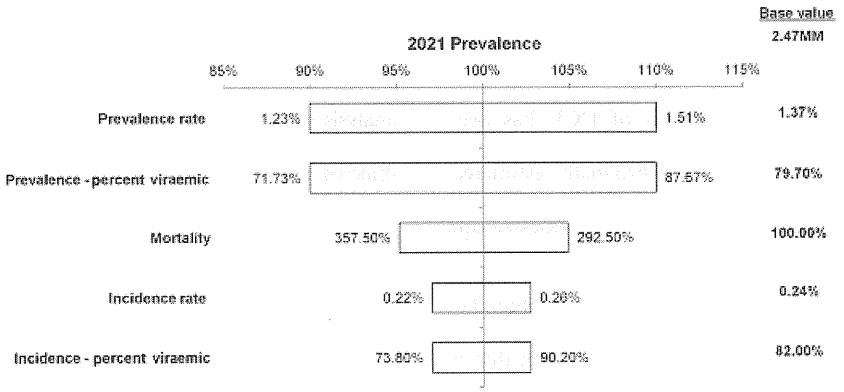


Fig. 12. Key drivers of uncertainty for 2021 prevalence.

epidemiological data are difficult to obtain. Several models have been developed, but all predictive models are limited by the uncertainty of key assumptions. In this paper, we utilized a system-based approach for compiling epidemiological data and validating the model. These data are the result of the first phase of an ongoing disease-specific project. In this manuscript, we have described a system dynamic modelling framework that can be used to predict future prevalence of the disease in different regions; our methodology provides a unique means to understand the cause and effect relationships between the key assumptions that need to be made and future trends in the prevalence and incidence of the disease. Using a system approach, it is possible to estimate the future trend in the absence of perfect information.

A comprehensive global review of the literature was used to gather size and flow data of infected pools in 37 countries. A weighting was considered to rank the importance of the reported data. Sensitivity analyses were used to adjust for key drivers of the future prevalence. Where data were not available, a model was used to estimate mortality rates. This method provides an approach to modelling the epidemiology of HCV infection that can be used in countries where detailed data are not available, where each population group is first assessed as a dynamic system with inflows and outflows to and from the pool. The unique benefit of such an approach is that it does not require accurate information, except for the relative size of populations moving in and out.

We take into account trends showing that, for example, the prevalence is expected to decline in several regions including USA, as shown by Davis *et al.* (17). The benefits of this approach and the simulation model were validated using data from USA, where considerable research has been carried out to understand the historical and future trends of HCV infection.

There are many difficulties in determining the prevalence of HCV. Many countries have reported cross-sectional data in unrepresentative populations – for example, blood donors. Others have relied on expert opinion. Similarly, incidence is one of the most problematic factors in estimating the future epidemiology of HCV infection for any given country, and reported prevalence rates may not be representative of a country's population. Even when robust studies are available, it is possible that the overall prevalence is underestimated. Studies in USA have shown that the true prevalence is higher than NHANES' estimates due to undersampling in high prevalence subgroups (50, 72). When underrepresented subgroups were taken into account (e.g. homeless, incarcerated, veterans, active military personnel, etc.), HCV prevalence was shown to be significantly higher. For the purpose of this work, NHANES data were used, given that the trends and conclusions will remain the same.

Our results reveal that the diagnosis of HCV has been increasing while treatment rates have been declining. The diagnosis rate in USA is higher than originally estimated. Although the exact reasons for this are not clear, it is possible that an increasing number of patients are being assessed and treated within clinical trials or that treatment is being temporarily withheld until the advent of protease inhibitors in combination with pegylated interferon and ribavirin. Another key finding of this study is that the mortality rate among HCV-infected individuals was determined to be higher than that of the general population and is a key driver of future prevalence; also, the disease burden resulting from HCV infection will increase even though the total size of the infected population will decrease.

Our model shows that mortality rates among HCV-infected individuals are higher than those of the general population, and predicts a noticeable increase in liver-related mortality rates after the age of 45 as the disease progresses, a finding that is in keeping with previous studies. For example, recently, McDonald and colleagues observed a significant excess for any diagnosis as well as liver-related diagnosis in a large cohort of HCV-infected individuals (73). However, it is important to note that extraneous mortality due to IDUs overshadows all other causes of mortality in individuals between ages 15 and 44.

It should be pointed out that the pool of existing cases of liver disease is due to past infection. HCV was not identified until 1989, but by then had infected tens of millions of people worldwide. Thus, the prevalence of HCV infection may decrease despite the prevalence of liver disease increasing because of a significant lag between the onset of infection and advanced disease. The impact of the disease is predicted to be substantial, however. A many-fold increase in health care utilization is likely as a result of the current and impending disease burden. Hepatitis C is now a common reason for a hepatology consultation and is the single leading indication for liver transplantation in USA and several other countries.

The launch of new drugs, higher treatment levels and increased efficacy rates will accelerate the reduction in prevalence with time. Sensitivity analysis was used to further understand the cause and effect relationship between the different assumptions. Most factors that were identified as part of this analysis – prevalence, viraemic and incidence rates – were predictable. However, a critical factor highlighted as a result of this work, which has not been identified in other studies, was the large impact of mortality rate among infected individuals. Small changes in the mortality rate seem to have a substantial impact on the future prevalence. A review of the prevalence flow diagram highlighted the importance of mortality where it outweighs both new cases (incidence) and the number of cured patients. There are two logical reasons why mortality is now showing up as a key driver of future prevalence. First, the overall prevalent population is ageing and their overall mortality rate is increasing even in the absence of liver-related deaths. Second, most new cases in USA (and many other countries) are among IDUs, who, although younger, have a much higher mortality rate between the ages 15 and 44 due to their high-risk life style. These data are supported by recent Australian, Scottish and Swedish publications. The addition of the liver-related deaths at older ages (45+) in both populations only exacerbates the overall impact.

It is possible to suggest that our analysis predicts HCV infections will simply disappear with time as the result of low incidence and high mortality rates. This is not the case. Our analysis suggests that approximately 2 million infected individuals will remain in USA by 2021. In this same timeframe, the infected population over the age of 60 will more than double, resulting in a proportional increase in liver-related diseases. Treatment rates remain extremely low for this disease and they have been declining in USA since 2004. We predict an increase in the disease burden unless there is a substantial increase in treatment and cured rates.

From a system approach, the ageing of the infected population could result in some unexpected trends. An increase in nosocomial infections could occur among elderly populations as the prevalence in this population increases and a higher

percentage of the older population requires medical treatment. In fact, a review of data among US blood donors (74) already shows an increase in incidence among 50+-year-old first-time blood donors. However, this assumption is based on a continued risk of transmission in hospital settings at a time when transmission by blood products is no longer occurring in regions where sensitive testing is utilized.

The model suggests that it is feasible to eradicate HCV infections. Accelerating the decline in the prevalent population will ultimately lead to a reduction in incidence, which will further reduce prevalence. This can be accomplished through segmentation of the high-risk and infected populations and developing unified strategies for each segment. While strategies to eradicate HCV are beyond the scope of this work, there is ample evidence that such strategies could work. In France, a multidisciplinary approach to treating IDUs has significantly reduced the number of new cases in this population. However, a close collaboration and communication between the medical and public health communities responsible for diagnosis, tracking and treatment (and the development of more effective and palatable treatments by the pharmaceutical industry) will be required. Our model suggests a means for control of the infection, in the absence of a vaccine.

Acknowledgements

This study was completed through the International Conquer C Coalition (I-C3) organization. Funding for this programme was provided through an educational grant provided by Merck and support from the Center for Disease Analysis. We are indebted to all I-C3 and Regional Conquer C Coalitions members for their contributions and comments. We would also like to thank David Goldberg and Scott McDonald of Health Protection Scotland, Ann-Sofi Duberg of Department of Infectious Diseases, Örebro University Hospital, Örebro, Sweden, and Janaki Amin, of Epidemiology and Clinical Research, University of New South Wales, Australia, for their contributions and insights to the mortality section of this article. We are also grateful to Dr. Thierry Poynard of Groupe Hospitalier Pitié-Salpêtrière in Paris, France, for providing us with previously unpublished hazard ratios for untreated HCV-infected patients. Finally, we would like to acknowledge Regina Klein and Angie Largen of Center for Disease Analysis for their assistance with data gathering and analysis in the preparation of this document.

Disclosures: **DK** Advisory Board: MSD. **HAR** Grant: Merck. **CLC** Research: Merck, Roche, Boehringer Ingelheim, Vertex, Tibotec and Gilead. Speaker: Merck, Roche. Advisor: Merck, Roche and Vertex. **AA** Consultant: Roche, Gilead, Novartis, BMS, J&J, Merck, Schering Plough, Research Grants: Gilead, Merck, BMS. **GMD** Financial relationship with a commercial interest: GlaxoSmithKline. Grant/Research Support: Vertex. **SP** Board Member: BMS, Boehringer Ingelheim, Tibotec/Janssen-Cilag, Gilead, Roche, Merck/Schering-Plough and Abbott. Speaker: GSK, BMS, Boehringer Ingelheim, Tibotec/Janssen Cilag, Gilead, Roche and Schering-Plough. Grants: BMS, Gilead, Roche and Merck/Schering-Plough. **EZ** Nothing to disclose. **KK** Nothing to disclose. **KHH** Nothing to disclose. **SZ** Consultant: Abbott, Achillion, Anadys, BMS, Gilead, Itherix, Merck, Novartis, Pfizer, Roche, Santaris, Tibotec and Vertex. **FN** Advisor: Schering-Plough, Roche, Novartis, Abbott and Gilead.

Appendix A

Table A1. Summary of key assumptions

Model input	Assumption	Data source
Incidence rate	2004 incidence rate by age and gender	Center for Disease Control, Hepatitis Surveillance Report (36)
Acute to chronic rate	82%	Thomas and Seeff (28)
Prevalence rate	2002 prevalence by age and gender applied to 2004 population	Armstrong *et al.* (49)
Viraemic rate	79.7%	Armstrong *et al.* (49)
Mortality model input – % of infected population with risk factors	IDU (Pop_{IDU}) – 47% Transfusion (Pop_{Trans}) – 7%	Daniels *et al.* (27) Alter *et al.* (48)
Genotype distribution	G1 – 73.7% G2 – 14.9% G3 – 7.4% Other – 4%	Alter *et al.* (48)
Annual newly diagnosed rate	69.2 per 100 000	Klevens *et al.* (58)
Treated population	144k in 2004 to 83k in 2007	Volk *et al.* (5)
Relapse rate	G1 – 23% G2 – 4.2% G3 – 10.6%	WINR Study, Jacobson *et al.* (75)
Sustained viral response rate	G1 – 40% G2 & 3 – 72%	IDEAL Trials, McHutchison *et al.* (76) WINR Study, Jacobson *et al.* (75)

References

1. El-Serag HB, Mason AC. Rising incidence of hepatocellular carcinoma in the United States. *N Engl J Med* 1999; **340**: 745–50.
2. Poynard T, Yuen MF, Ratziu V, Lai CL. Viral hepatitis C. *Lancet* 2003; **362**: 2095–100.
3. Seeff LB, Hoofnagle JH. National Institutes of Health Consensus Development Conference: management of hepatitis C: 2002. *Hepatology* 2002; **36**: S1–20.
4. Pereira A. Health and economic consequences of HCV lookback. *Transfusion* 2001; **41**: 832–9.
5. Volk ML, Tocco R, Saini S, Lok AS. Public health impact of antiviral therapy for hepatitis C in the United States. *Hepatology* 2009; **50**: 1750–5.
6. Deuffic-Burban S, Mathurin P, Valleron AJ. Modelling the past, current and future HCV burden in France: detailed analysis and perspectives. *Stat Methods Med Res* 2009; **18**: 233–52.
7. Deuffic-Burban S, Poynard T, Sulkowski MS, Wong JB. Estimating the future health burden of chronic hepatitis C and human immunodeficiency virus infections in the United States. *J Viral Hepat* 2007; **14**: 107–15.
8. Salomon JA, Weinstein MC, Hammitt JK, Goldie SJ. Cost-effectiveness of treatment for chronic hepatitis C infection in an evolving patient population. *JAMA* 2003; **290**: 228–37.
9. Sweeting MJ, De Angelis D, Hickman M, Ades AE. Estimating hepatitis C prevalence in England and Wales by synthesizing evidence from multiple data sources. Assessing data conflict and model fit. *Biostatistics* 2008; **9**: 715–34.
10. Sweeting MJ, De Angelis D, Neal KR, *et al.* Estimated progression rates in three United Kingdom hepatitis C cohorts differed according to method of recruitment. *J Clin Epidemiol* 2006; **59**: 144–52.
11. Terrault NA, Im K, Boylan R, *et al.* Fibrosis progression in African Americans and Caucasian Americans with chronic hepatitis C. *Clin Gastroenterol Hepatol* 2008; **6**: 1403–11.
12. Tramarin A, Gennaro N, Compostella FA, *et al.* HCV screening to enable early treatment of hepatitis C: a mathematical model to analyse costs and outcomes in two populations. *Curr Pharm Des* 2008; **14**: 1655–60.
13. Wong JB, Bennett WG, Koff RS, Pauker SG. Pretreatment evaluation of chronic hepatitis C: risks, benefits, and costs. *JAMA* 1998; **280**: 2088–93.
14. Hutchinson SJ, Bird SM, Goldberg DJ. Modeling the current and future disease burden of hepatitis C among injection drug users in Scotland. *Hepatology* 2005; **42**: 711–23.
15. Wong JB, McQuillan GM, McHutchison JG, Poynard T. Estimating future hepatitis C morbidity, mortality, and costs in the United States. *Am J Public Health* 2000; **90**: 1562–9.
16. Buti M, San Miguel R, Brosa M, *et al.* Estimating the impact of hepatitis C virus therapy on future liver-related morbidity, mortality and costs related to chronic hepatitis C. *J Hepatol* 2005; **42**: 639–45.
17. Davis GL, Alter MJ, El-Serag H, Poynard T, Jennings LW. Aging of hepatitis C virus (HCV)-infected persons in the United States: a multiple cohort model of HCV prevalence and disease progression. *Gastroenterology* 2010; **138**: 513–21.
18. Howard R, Matheson JE. Influence diagram retrospective. *Decis Anal* 2005; **2**: 144–7.
19. Keeney RL, Raiffa H. *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*. Cambridge: Cambridge University Press, 1993.
20. Keeney RL. *Value Focused Thinking: A Path to Creative Decisionmaking*. President and Fellow of Harvard College, 1992. Harvard University Press, Cambridge, MA, USA.
21. Hammond JS, Keeney RL, Raiffa H. *Smart Choices: A Practical Guide to Making Better Life Decisions*. Harvard Business School Press, Boston, MA, USA 1999.
22. Kirkwood GW. *Strategic Decision Making: Multiobjective Decision Analysis with Spreadsheets*. Belmont, CA: Duxbury Press, 1997.
23. Grelk BJ, Kloeber JM, Jackson JA, Parnell GS, Deckro RF. Making the CERCLA criteria analysis of remedial alternatives more objective. *Remediation* 1998; **8**: 87–105.
24. Hamill JT, Deckro RF, Kloeber JM. Evaluating information assurance strategies. *Decis Support Sys* 2005; **39**: 463–84.
25. Edwards W, Barron FH. SMARTS and SMARTER: improved simple methods for multiattribute utility measurements. *Organ Behav Hum Decis Process* 1994; **60**: 306–25.
26. Centers for Disease Control and Prevention. *Guidelines for Viral Hepatitis Surveillance and Case Management*. Atlanta, GA: US Department of Health, Education, and Welfare, Public Health Service, Atlanta, GA, USA 2005.
27. Daniels D, Grytdal S, Wasley A. Surveillance for acute viral hepatitis – United States, 2007. *MMWR Surveill Summ* 2009; **58**: 1–27.
28. Thomas DL, Seeff LB. Natural history of hepatitis C. *Clin Liver Dis* 2005; **9**: 383–98.
29. Armstrong GL, Alter MJ, McQuillan GM, Margolis HS. The past incidence of hepatitis C virus infection: implications for the future burden of chronic liver disease in the United States. *Hepatology* 2000; **31**: 777–82.
30. Wiegand J, Deterding K, Cornberg M, Wedemeyer H. Treatment of acute hepatitis C: the success of monotherapy with (pegylated) interferon alpha. *J Antimicrob Chemother* 2008; **62**: 860–5.
31. Corey KE, Ross AS, Wurcel A, *et al.* Outcomes and treatment of acute hepatitis C virus infection in a United States population. *Clin Gastroenterol Hepatol* 2006; **4**: 1278–82.
32. Kamal SM, Moustafa KN, Chen J, *et al.* Duration of peginterferon therapy in acute hepatitis C: a randomized trial. *Hepatology* 2006; **43**: 923–31.
33. Morin T, Pariente A, Lahmek P, *et al.* Acute hepatitis C: analysis of a 126-case prospective, multicenter cohort. *Eur J Gastroenterol Hepatol* 2010; **22**: 157–66.
34. Villano SA, Vlahov D, Nelson KE, Cohn S, Thomas DL. Persistence of viremia and the importance of long-term follow-up after acute hepatitis C infection. *Hepatology* 1999; **29**: 908–14.
35. Alter MJ, Margolis HS, Krawczynski K, *et al.* The natural history of community-acquired hepatitis C in the United States. The sentinel counties chronic non-A, non-B hepatitis study team. *N Engl J Med* 1992; **327**: 1899–905.
36. Centers for Disease Control and Prevention. *Hepatitis Surveillance*. Hepatitis Surveillance/Center for Disease Control, Atlanta, GA, USA 2006.
37. Centers for Disease Control. *Disease Burden from Viral Hepatitis A, B, and C in the United States. CDC Division of Viral Hepatitis – Statistics and Surveillance*. Centers for Disease Control, 2007. Available at http://www.cdc.gov/hepatitis/statistics/index.htm.
38. United Nations, Department of Economic and Social Affairs, Population Division. *World Population Prospects. 2002 Population Database*. United Nations, Department of Economic and Social Affairs, Population Division, 2005. Available at http://esa.un.org/unpp/index.asp?panel=2.
39. Duberg AS, Torner A, Davidsdottir L, *et al.* Cause of death in individuals with chronic HBV and/or HCV infection, a nationwide community-based register study. *J Viral Hepat* 2008; **15**: 538–50.
40. McDonald SA, Hutchinson SJ, Bird SM, *et al.* A population-based record linkage study of mortality in hepatitis C-diagnosed persons with or without HIV coinfection in Scotland. *Stat Methods Med Res* 2009; **18**: 271–83.
41. Amin J, Law MG, Bartlett M, Kaldor JM, Dore GJ. Causes of death after diagnosis of hepatitis B or hepatitis C infection: a large community-based linkage study. *Lancet* 2006; **368**: 938–45.
42. United Nations, Statistical Office. *Demographic yearbook/Annuaire dâemographique. Demographic Yearbook*, 59th edn. United Nations, Statistical Office, New York, NY, USA 2007. Available at http://unstats.un.org/unsd/demographic/products/dyb/dyb2007.htm (accessed 9 October 2009).
43. United Nations and Statistical Office. *Demographic Yearbook/Annuaire dâemographique 2006*, 58th edn. Department of Economic and Social Affairs, Statistical Office, United Nations, 2008.
44. Seeff LB, Miller RN, Rabkin CS, *et al.* 45-year follow-up of hepatitis C virus infection in healthy young adults. *Ann Intern Med* 2000; **132**: 105–11.
45. Guiltinan AM, Kaidarova Z, Custer B, *et al.* Increased all-cause, liver, and cardiac mortality among hepatitis C virus-seropositive blood donors. *Am J Epidemiol* 2008; **167**: 743–50.
46. Frischer M, Goldberg D, Rahman M, Berney L. Mortality and survival among a cohort of drug injectors in Glasgow, 1982–1994. *Addiction* 1997; **92**: 419–27.
47. Kamper-Jorgensen M, Ahlgren M, Rostgaard K, *et al.* Survival after blood transfusion. *Transfusion* 2008; **48**: 2577–84.
48. Alter MJ, Kruszon-Moran D, Nainan OV, *et al.* The prevalence of hepatitis C virus infection in the United States, 1988 through 1994. *N Engl J Med* 1999; **341**: 556–62.
49. Armstrong GL, Wasley A, Simard EP, *et al.* The prevalence of hepatitis C virus infection in the United States, 1999 through 2002. *Ann Intern Med* 2006; **144**: 705–14.
50. Chak E, Talal AH, Sherman KE, Schiff ER, Saab S. Hepatitis C virus infection in the United States: an estimate of true prevalence. *Liver Int* 2011. DOI: 10.1111/j.1478-3231.2011.02494.x. (in press).
51. Abdel-Hamid M, El-Daly M, El-Kafrawy S, *et al.* Comparison of second- and third-generation enzyme immunoassays for detecting antibodies to hepatitis C virus. *J Clin Microbiol* 2002; **40**: 1656–9.

52. Silva L, Parana R, Mota E, *et al.* Prevalence of hepatitis C virus in urban and rural populations of northeast Brazil-pilot study. *Arq Gastroenterol* 1995; **32**: 168–71.
53. Glynn SA, Kleinman SH, Schreiber GB, *et al.* Trends in incidence and prevalence of major transfusion-transmissible viral infections in US blood donors, 1991 to 1996. Retrovirus Epidemiology Donor Study (REDS). *JAMA* 2000; **284**: 229–35.
54. Santos-Lopez G, Sosa-Jurado F, Vallejo-Ruiz V, Melendez-Mena D, Reyes-Leyva J. Prevalence of hepatitis C virus in the Mexican population: a systematic review. *J Infect* 2008; **56**: 281–90.
55. Bornschlegel K, Berger M, Garg RK, *et al.* Prevalence of hepatitis C infection in New York City, 2004. *J Urban Health* 2009; **86**: 909–17.
56. Brady KA, Weiner M, Turner BJ. Undiagnosed hepatitis C on the general medicine and trauma services of two urban hospitals. *J Infect* 2009; **59**: 62–9.
57. Kilbourne AM, McCarthy JF, Himelhoch S, *et al.* Guideline-concordant hepatitis C virus testing and notification among patients with and without mental disorders. *Gen Hosp Psychiatry* 2008; **30**: 495–500.
58. Klevens RM, Miller J, Vonderwahl C, *et al.* Population-based surveillance for hepatitis C virus, United States, 2006–2007. *Emerg Infect Dis* 2009; **15**: 1499–502.
59. Davis GL, Albright JE, Cook SF, Rosenberg DM. Projecting future complications of chronic hepatitis C in the United States. *Liver Transpl* 2003; **9**: 331–8.
60. Deuffic-Burban S, Wong JB, Valleron AJ, *et al.* Comparing the public health burden of chronic hepatitis C and HIV infection in France. *J Hepatol* 2004; **40**: 319–26.
61. Grant WC, Jhaveri RR, McHutchison JG, Schulman KA, Kauf TL. Trends in health care resource use for hepatitis C virus infection in the United States. *Hepatology* 2005; **42**: 1406–13.
62. Poynard T, Ratziu V, Charlotte F, *et al.* Rates and risk factors of liver fibrosis progression in patients with chronic hepatitis c. *J Hepatol* 2001; **34**: 730–9.
63. Massard J, Ratziu V, Thabut D, *et al.* Natural history and predictors of disease severity in chronic hepatitis C. *J Hepatol* 2006; **44**: S19–24.
64. Thomas DL, Astemborski J, Rai RM, *et al.* The natural history of hepatitis C virus infection: host, viral, and environmental factors. *JAMA* 2000; **284**: 450–6.
65. Butt AA, Wang X, Moore CG. Effect of hepatitis C virus and its treatment on survival. *Hepatology* 2009; **50**: 387–92.
66. Neal KR, Ramsay S, Thomson BJ, Irving WL. Excess mortality rates in a cohort of patients infected with the hepatitis C virus: a prospective study. *Gut* 2007; **56**: 1098–104.
67. Prasad L, Spicher VM, Negro F, Rickenbach M, Zwahlen M. Little evidence that hepatitis C virus leads to a higher risk of mortality in the absence of cirrhosis and excess alcohol intake: the Swiss hepatitis C cohort study. *J Viral Hepat* 2009; **16**: 644–9.
68. Harris HE, Ramsay ME, Andrews NJ. Survival of a national cohort of hepatitis C virus infected patients, 16 years after exposure. *Epidemiol Infect* 2006; **134**: 472–7.
69. Harris HE, Ramsay ME, Andrews N, Eldridge KP. Clinical course of hepatitis C virus during the first decade of infection: cohort study. *BMJ* 2002; **324**: 450–3.
70. Omland LH, Krarup H, Jepsen P, *et al.* Mortality in patients with chronic and cleared hepatitis C viral infection: a nationwide cohort study. *J Hepatol* 2010; **53**: 36–42.
71. Kanwal F, Hoang T, Kramer JR, *et al.* Increasing prevalence of HCC and cirrhosis in patients with chronic hepatitis C virus infection. *Gastroenterology* 2011; **140**: 1182–8.e1.
72. Gish RG, Afdhal NH, Dieterich DT, Reddy KR. Management of hepatitis C virus in special populations: patient and treatment considerations. *Clin Gastroenterol Hepatol* 2005; **3**: 311–8.
73. McDonald SA, Hutchinson SJ, Bird SM, *et al.* Excess morbidity in the hepatitis C-diagnosed population in Scotland, 1991–2006. *Epidemiol Infect* 2011; **139**: 344–53.
74. Zou S, Dorsey KA, Notari EP, *et al.* Prevalence, incidence, and residual risk of human immunodeficiency virus and hepatitis C virus infections among United States blood donors since the introduction of nucleic acid testing. *Transfusion* 2010; **50**: 1495–504.
75. Jacobson IM, Brown RS Jr, Freilich B, *et al.* Peginterferon alfa-2b and weight-based or flat-dose ribavirin in chronic hepatitis C patients: a randomized trial. *Hepatology* 2007; **46**: 971–81.
76. McHutchison JG, Lawitz EJ, Shiffman ML, *et al.* Peginterferon alfa-2b or alfa-2a with ribavirin for treatment of hepatitis C infection. *N Engl J Med* 2009; **361**: 580–93.

Liver INTERNATIONAL
OFFICIAL JOURNAL OF THE INTERNATIONAL ASSOCIATION FOR THE STUDY OF THE LIVER

Liver International ISSN 1478-3223

A systematic review of hepatitis C virus epidemiology in Asia, Australia and Egypt

William Sievert[1], Ibrahim Altraif[2], Homie A. Razavi[3], Ayman Abdo[4], Ezzat Ali Ahmed[5], Ahmed AlOmair[6], Deepak Amarapurkar[7], Chien-Hung Chen[8], Xiaoguang Dou[9], Hisham El Khayat[10], Mohamed elShazly[11], Gamal Esmat[12], Richard Guan[13], Kwang-Hyub Han[14], Kazuhiko Koike[15], Angela Largen[3], Geoff McCaughan[16], Sherif Mogawer[17], Ali Monis[18], Arif Nawaz[19], Teerha Piratvisuth[20], Faisal M. Sanai[21], Ala I. Sharara[22], Scott Sibbel[3], Ajit Sood[23], Dong Jin Suh[24], Carolyn Wallace[3], Kendra Young[3] and Francesco Negro[25]

1 Monash Medical Centre and Monash University, Melbourne, Vic., Australia
2 Department of Hepatobiliary Science and Liver Transplantation, King Abdulaziz Medical City, Riyadh, Saudi Arabia
3 Center for Disease Analysis, Kromite, Louisville, CO, USA
4 King Khaled University Hospital, King Saud University, Riyadh, Saudi Arabia
5 Gastroenterology Unit, Department of Internal Medicine, Alexandria University, Alexandria, Egypt
6 GI/Hepatology Division, Main Hospital, King Fahad Medical City, Riyadh, Saudi Arabia
7 Department of Gastroenterology, Bombay Hospital and Medical Research Centre, Mumbai, India
8 Department of Internal Medicine, National Taiwan University Hospital and National Taiwan University College of Medicine, Taipei, Taiwan
9 Department of Infectious Disease, Sheng Jing Hospital, China Medical University, Shenyang, China
10 Department of Gastroenterology, Theodore Bilharz Research Institute, Giza, Egypt
11 Department of Tropical Medicine, Alexandria University, Alexandria, Egypt
12 Department of Tropical Medicine and Hepatology, Cairo University, Cairo, Egypt
13 Medical Clinic One, Mount Elizabeth Medical Centre, Singapore, Singapore
14 Department of Internal Medicine, Yonsei Liver Cancer Special Clinic, Yonsei Institute of Gastroenterology, Yonsei University College of Medicine, Seoul, Korea
15 Department of Gastroenterology, Graduate School of Medicine, University of Tokyo, Tokyo, Japan
16 A.W. Marrow Gastroenterology and Liver Center, Royal Prince Alfred Hospital, Centenary Institute, University of Sydney, Sydney, Australia
17 Department of Internal Medicine, Cairo University, Cairo, Egypt
18 Department of Internal Medicine, Ain Sahms University, Cairo, Egypt
19 Department of Gastroenterology, Fatima Memorial Hospital College of Medicine and Dentistry, Shadman, Lahore, Pakistan
20 Department of Internal Medicine, NKC Institute of Gastroenterology and Hepatology, Songklanagarind Hospital, Prince of Songkla University, Hat Yai, Thailand
21 Department of Hepatobiliary Science and Liver Transplantation, King Abdulaziz Medical City, Riyadh, Saudi Arabia
22 Division of Gastroenterology, American University of Beirut, Beirut, Lebanon
23 Department of Gastroenterology, Dayanand Medical College and HospitalLudhiana, India
24 Asan Medical Center, University of Ulsan, Seoul, South Korea
25 Services de Gastroenterologie et d'Hepatologie et de Pathologie Clinique, Hopitaux Universitaires, Genève, Switzerland

Keywords
diagnosis – disease burden – epidemiology – HCV – hepatitis C – incidence – mortality – prevalence – systems modeling – treatment rate

Abbreviations
EDHS, egypt demographic and health survey; HCV, hepatitis C virus; HCVPWG, hepatitis C virus projections working group; I-C3, international conquer C coalition; IDU, injection drug use; IV, intravenous; MOH, ministry of health; NGHA, national guard health affairs in Saudi Arabia; NNDSS, national notifiable diseases surveillance system; RNA, ribonucleic acid.

Correspondence
Homie A. Razavi, Center for Disease Analysis, Kromite, 901 Front Street, Suite 291, Louisville, CO 80027, USA.
Tel: +1 720 890 4848
Fax: +1 303 552 9119
e-mail: hrazavi@kromite.com

Received 27 March 2011
Accepted 1 April 2011

DOI:10.1111/j.1478-3231.2011.02540.x

Abstract

Background: The hepatitis C pandemic has been systematically studied and characterized in North America and Europe, but this important public health problem has not received equivalent attention in other regions. *Aim:* The objective of this systematic review was to characterize hepatitis C virus (HCV) epidemiology in selected countries of Asia, Australia and Egypt, i.e. in a geographical area inhabited by over 40% of the global population. *Methodology:* Data references were identified through indexed journals and non-indexed sources. In this work, 7770 articles were reviewed and 690 were selected based on their relevance. *Results:* We estimated that 49.3–64.0 million adults in Asia, Australia and Egypt are anti-HCV positive. China alone has more HCV infections than all of Europe or the Americas. While most countries had prevalence rates from 1 to 2% we documented several with relatively high prevalence rates, including Egypt (15%), Pakistan (4.7%) and Taiwan (4.4%). Nosocomial infection, blood transfusion (before screening) and injection drug use were identified as common risk factors in the region. Genotype 1 was common in Australia, China, Taiwan and other countries in North Asia, while genotype 6 was found in Vietnam and other Southeast Asian countries. In India and Pakistan genotype 3 was predominant, while genotype 4 was found in Middle Eastern countries such as Egypt, Saudi Arabia and Syria. *Conclusion:* We recommend implementation of surveillance systems to guide effective public health policy that may lead to the eventual curtailment of the spread of this pandemic infection.

The hepatitis C pandemic has been systematically studied and characterized in North America and Europe, but in other areas of the world this important public health problem has not received equivalent attention. The objective of this systematic review is to characterize hepatitis C virus (HCV) epidemiology in Egypt, Australia and selected countries in Asia, i.e. in a geographical area inhabited by over 40% of the global