Empirically Calibrated Model of Hepatitis C Virus Infection in the United States

Joshua A. Salomon¹,², Milton C. Weinstein², James K. Hammitt², and Sue J. Goldie²

¹ Global Programme on Evidence for Health Policy, World Health Organization, Geneva, Switzerland.
² Harvard Center for Risk Analysis, Department of Health Policy and Management, Harvard School of Public Health, Boston, MA.

Received for publication December 12, 2001; accepted for publication March 27, 2002.

This study presents a comprehensive epidemiologic model of hepatitis C in the United States. Through empirical calibration of model parameter values, the objectives were to gain insights into uncertain aspects of the natural history of hepatitis C and to improve the basis for projecting the future course of the epidemic. A systematic review of the published literature was conducted to define plausible ranges around model parameters, and multiple simulations of the model were undertaken using sampled values from these ranges. Model predictions produced by each set of sampled values were compared with available epidemiologic data on infection prevalence and mortality from liver cancer, and various goodness-of-fit criteria were used to identify the range of parameter values that were consistent with these data. The results of the study indicate that rates of progression to advanced liver disease may be lower than previously assumed. The authors also found that a wide range of plausible assumptions about heterogeneity in these rates, beyond that explained by age and sex, is consistent with observed epidemiologic trends. These findings have important implications both for individual clinical decisions and for broader public health policy.

carcinoma, hepatocellular; hepatitis B; hepatitis C; natural history

Abbreviations: HBsAg, hepatitis B surface antigen; HBV, hepatitis B virus; HCC, hepatocellular carcinoma; HCV, hepatitis C virus; NHANES III, Third National Health and Nutrition Examination Survey; SEER, Surveillance, Epidemiology, and End Results.

Hepatitis C virus (HCV) is the most common blood-borne infection in the United States and is a leading cause of chronic hepatitis, cirrhosis, and hepatocellular carcinoma (HCC) (1). Approximately four million Americans have been exposed to HCV, of whom an estimated 2.7 million remain chronically infected (2). Since the identification of the virus in the late 1980s (3), significant progress has been made in understanding the epidemiology of HCV (4, 5), but substantial uncertainty persists regarding risks of progression to advanced liver disease (6, 7). In addition, available treatment regimens are effective in only approximately 30–60 percent of patients (8–14), and there is limited evidence on valid prognostic factors (15–17). Treatment decisions for individual patients thus involve difficult trade-offs between the costs and potential side effects of current regimens and the uncertain magnitude of the benefits.

From the population perspective, several public health interventions have received considerable attention, including a government “look back” campaign launched in 1998 to notify people who received blood from potentially infected donors (18) and the Surgeon General’s July 2000 letter warning the public about the “silent epidemic” and encouraging at-risk persons to get tested (19). These broader public policy decisions, much like individual clinical decisions, will inevitably proceed before all uncertainties are resolved. Decisions at both levels must consider the future course of the disease in infected persons and the collective impact of HCV on the population. Several mathematical models have been developed to forecast the trajectory of the HCV epidemic by using various data sources and different assumptions about disease progression (20–23). Interpretation of these projections and assessments of their validity must be guided by two fundamental criteria: 1) that a model should accurately reflect both the knowledge and uncertainty about major parameters relating to the natural history of the disease on the basis of the best available data, and 2) that
modeled outcomes should be consistent with epidemiologic data describing the population impact of the epidemic. One limitation of previous models has been the failure to impose the latter constraint in choosing the values of the model parameters. Previous studies have calibrated past trends in incidence of HCV infection using information on seroprevalence (20) or projected mortality using current patterns as a starting point (23), but to our knowledge, none have taken into account data on prevalence and mortality simultaneously. Given the important uncertainties around the natural history of HCV, the confrontation of multiple data sources may serve to provoke a critical examination of key assumptions in a model.

The objectives of this study were to develop a comprehensive model of HCV infection in the United States and to fit this model to observed epidemiologic trends relating to HCV and its consequences. Through a model-fitting strategy that emphasized empirical calibration of parameter values, our goals were to gain insights into key uncertainties around the natural history of hepatitis C and to improve the basis for projections of the future course of the epidemic.

MATERIALS AND METHODS

Analytic overview

We developed a mathematical model of HCV infection, including acquisition of infection, probability of persistence, and risk of progression to end-stage liver disease. The model includes the natural history of hepatitis B virus (HBV) infection because of key interactions, shared risk factors, and related etiologic roles in the development of cirrhosis and HCC. A comprehensive review of the published literature was conducted to define plausible ranges for the model parameter values. Numerical simulations were undertaken based on sampling jointly from these ranges to examine the different outcomes implied by various sets of parameter values. For each set of sampled parameter values, modeled outcomes were compared with available epidemiologic data on HCV and HBV seroprevalence and HCC mortality, and goodness-of-fit was evaluated. Guided by various fit criteria, we identified sets of parameter values that were consistent with the observed trends in the population.

Model structure

The model consists of a set of mutually exclusive compartments stratified by age and sex and defined in terms of serologic infection status (HCV or HBV) and clinical liver disease status (relating to HCV, HBV, or other causes such as alcohol consumption).

For both HCV and HBV, the model distinguishes between uninfected, chronically infected, and previously, but no longer, infected persons. Early stages of clinical disease are defined using the METAVIR scoring system (24), which characterizes the extent of fibrosis (scarring) that results as damaged liver cells are repaired (e.g., “no fibrosis,” “portal fibrosis without septa,” “portal fibrosis with few septa,” and “numerous septa without cirrhosis”). Advanced stages of disease are defined clinically as compensated cirrhosis, decompensated cirrhosis (portal hypertension leading to ascites, variceal bleeding, or hepatic encephalopathy), and HCC.

Model transitions

Infection. Acquisition of HCV or HBV infection occurs through three major routes: mother-to-child transmission, transfusion-associated infection, and community infection, including sexual transmission and exposure to infected blood through routes other than transfusion, such as injection drug use (see Appendix). Acute infection may become chronic or may resolve (figure 1). The model allows for delayed resolution of chronic HCV infections, marked by the loss of HCV RNA, and subsequent waning of antibody to HCV (the primary marker of past exposure) over time. Prior infection with HCV does not confer immunity against new infection (25). Similar transitions are modeled for HBV, although, unlike HCV, acute HBV may result in fatal fulminant hepatitis, and a prior HBV infection confers immunity from reinfection (26).

Clinical progression. Progression of liver disease is represented as a series of stepwise movements (figure 2), allowing the model to capture the lag between the time of infection and development of cirrhosis (27, 28). Rates of progression through stages of liver fibrosis vary by age, sex, and infection status and possibly by additional, unidentified factors (7, 28). To represent this potential heterogeneity, we delineate slow- and fast-progressing substates in the model.

HBV infection may progress directly to HCC in the absence of cirrhosis (29), but more commonly, cirrhosis precedes HCC. In the setting of HCV-induced liver disease, HCC appears to occur almost exclusively after cirrhosis (30). The model also allows for progression to cirrhosis and HCC in persons not infected with either HCV or HBV.
Birth and death. Persons enter the model at birth without fibrosis. The total number of births is determined by population age structure and age-specific fertility rates. The distribution of newborns across different infection categories depends on the infection category of the mother, the probabilities of vertical transmission of HCV and HBV, and the probabilities that acute infections lead to chronic infection (see Appendix). Persons exit the model upon death, which may occur from causes unrelated to hepatitis (i.e., via background mortality rates that depend on age, sex, and calendar year), from fulminant hepatic failure after acute HBV infection, or from decompensated cirrhosis or liver cancer.

Model simulation

A set of differential equations describes changes in the population of each model compartment over time, starting from an initial distribution. Because of the complexity of the model, analytic solutions are not possible, so numerical simulations trace movements between compartments at discrete 5-day time steps according to the rates and proportions defined by the parameter values. The model is written in the C programming language. Model simulations were undertaken for the period 1900–2000. The age and sex structure of the initial population in year 1900 was based on historical census data. Since a very small proportion of the population alive in 1900 would still be alive in 1970 (the first year of data to which the model was fit), we made a simplifying assumption that this initial population had no hepatitis or fibrosis.

Definition of parameter ranges

Our approach to model fitting involved multiple simulations using alternative sets of parameter values sampled from specified ranges. The first step was to define a range for each parameter, selected to be as inclusive as possible, based on available clinical and epidemiologic studies. We conducted a review of English-language papers on the natural history of HCV, HBV, and cirrhosis by using the MEDLINE database and the search terms “hepatitis C,” “hepatitis B,” “cirrhosis,” “natural history,” and “progression,” augmented by additional citations from these papers (2, 6, 7, 25–80) (table 1). The Appendix provides details on parameter values relating to infection risks.

Fibrosis progression. The key model parameters governing progression are the age- and sex-specific transition rates between consecutive fibrosis stages. Rates of fibrosis progression after HCV are difficult to define precisely because acute infection often remains undetected, chronic infection may last for several decades before major sequelae develop, and most longitudinal studies have had relatively short durations of follow-up. Most natural history studies have not included liver biopsies and therefore report on the incidence of clinically diagnosed cirrhosis rather than on histologic assessment of fibrosis stages; results have varied widely—from rapid progression in studies of patients with established liver disease to much more favorable outcomes in several long-term prospective studies (6, 27, 41, 64–76) (table 2). A smaller number of studies have estimated progression rates through intermediate histologic stages, including cross-sectional studies in which fibrosis stages were determined through biopsy for patients with known infection durations (27, 28, 77, 78) and prospective studies of untreated controls in clinical trials (79, 80).

Given the variability in findings across different natural history studies, we selected wide ranges around rates of fibrosis progression to allow the fitting procedure to identify the rates that are most consistent with observed data on infection prevalence and liver cancer mortality (table 3). In more intuitive terms, the ranges in table 3 allow for the 30-year cumulative probability of cirrhosis to range from 6 to 71 percent in males infected at age 30 years and from less than 1 to 54 percent for females, which accommodates the wide range of results reported in table 2.
To capture the heterogeneity in rates of fibrosis progression, in addition to that explained by age and sex (27, 28, 78, 79), and to allow for potential referral biases in some natural history studies, we assume that only a proportion of a newly infected cohort will be eligible to progress at the rates described in table 3, while the rest will instead progress at the same (slower) rates as uninfected persons.

### TABLE 1. Parameter ranges

<table>
<thead>
<tr>
<th>Infection parameters</th>
<th>Range</th>
<th>Sources (reference no.)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Infection parameters</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Risks of transfusion-associated or community infection</td>
<td>†</td>
<td></td>
</tr>
<tr>
<td>Relative risk of HBV‡ infection for persons with past HCV infection</td>
<td>3.0–8.0</td>
<td>2</td>
</tr>
<tr>
<td>Relative risk of HCV‡ infection for persons with past HBV infection</td>
<td>1.5–5.0</td>
<td>2</td>
</tr>
<tr>
<td>Probability of HCV vertical transmission</td>
<td>0.01–0.06</td>
<td>31–33</td>
</tr>
<tr>
<td>Probability of HBV vertical transmission</td>
<td>0.2–0.4</td>
<td>34–37</td>
</tr>
<tr>
<td>Probability that HCV infection will become chronic</td>
<td>0.6–0.9</td>
<td>38–41</td>
</tr>
<tr>
<td>Probability that HBV infection will become chronic by age (years)</td>
<td>26, 42, 43</td>
<td></td>
</tr>
<tr>
<td>&lt;1</td>
<td>0.80–0.95</td>
<td></td>
</tr>
<tr>
<td>1–2</td>
<td>0.40–0.70</td>
<td></td>
</tr>
<tr>
<td>3–4</td>
<td>0.20–0.30</td>
<td></td>
</tr>
<tr>
<td>5–7</td>
<td>0.10–0.20</td>
<td></td>
</tr>
<tr>
<td>&gt;7</td>
<td>0.05–0.10</td>
<td></td>
</tr>
<tr>
<td>Probability that acute HBV infection will lead to fatal fulminant hepatitis</td>
<td>0.001–0.003</td>
<td>44, 45</td>
</tr>
<tr>
<td>Rate of loss of HCV RNA</td>
<td>0–0.01</td>
<td>40, 46</td>
</tr>
<tr>
<td>Rate of loss of anti-HCV‡</td>
<td>0.006–0.018</td>
<td>39, 40</td>
</tr>
<tr>
<td>Rate of loss of HBsAg‡</td>
<td>0.01–0.03</td>
<td>42, 47–50</td>
</tr>
<tr>
<td><strong>Progression parameters</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proportion of new chronic HCV infections entering the fast-progression category</td>
<td>0.2–1.0</td>
<td>7</td>
</tr>
<tr>
<td>Rate of fibrosis progression</td>
<td>§</td>
<td></td>
</tr>
<tr>
<td>Rate of progression from compensated to decompensated cirrhosis</td>
<td>0.03–0.06</td>
<td>51–55</td>
</tr>
<tr>
<td>Rate of progression from cirrhosis to HCC‡</td>
<td>0.015–0.03</td>
<td>51–53, 55, 56</td>
</tr>
<tr>
<td>Rate of progression to HCC for chronic HBV without cirrhosis by age (years)¶</td>
<td>49, 57</td>
<td></td>
</tr>
<tr>
<td>&lt;20</td>
<td>0.0001–0.001</td>
<td></td>
</tr>
<tr>
<td>20–49</td>
<td>0.0005–0.005</td>
<td></td>
</tr>
<tr>
<td>≥50</td>
<td>0.003–0.015</td>
<td></td>
</tr>
<tr>
<td>Death rate related to decompensated cirrhosis</td>
<td>0.14–0.4</td>
<td>51, 58</td>
</tr>
<tr>
<td>Death rate related to HCC</td>
<td>0.3–0.7</td>
<td>59, 60</td>
</tr>
<tr>
<td>Relative risk of dying from other causes for HCV-infected persons</td>
<td>2.0–5.0</td>
<td>61–63</td>
</tr>
</tbody>
</table>

* Some parameters may depend on age or infection categories. Probabilities and proportions are expressed per person, while rates are expressed per person per year.
† See Appendix.
‡ HBV, hepatitis B virus; HCV, hepatitis C virus; anti-HCV, antibody to hepatitis C virus; HBsAg, hepatitis B surface antigen; HCC, hepatocellular carcinoma.
§ See table 3 and text for description of fibrosis progression rates in persons with chronic HCV. Ranges in table 3 were also used for chronic HBV. For coinfected persons, it was assumed that progression rates would range from the higher of the HBV and HCV rates in each sampled set of parameter values to the sum of the two rates. For uninfected persons, rates of cirrhosis incidence were modeled directly, using ranges that increase with age from <0.001% per year in those less than age 20 years to 0.01–0.06% per year in those aged 60 years and older.
¶ Ranges reported for males. Model also includes sex-specific adjustment that allows rates in females to be lower than those for males, with the proportional reductions ranging from 0 to 50%.
Model fitting

The goal of the model-fitting exercise was to identify, within the ranges defined around each model parameter, the subset of values that produced results consistent with epidemiologic data on HCV and HBV infection and mortality from primary HCC. Data from the Third National Health and Nutrition Examination Survey (NHANES III), conducted from 1988 to 1994, included serologic examinations for markers of HCV and HBV infection in approximately 20,000 persons (81). Age- and sex-specific prevalence of infection was estimated by using sampling weights to account for the complex multistage sampling plan of the survey (2). Age- and sex-specific mortality rates from primary HCC (International Classification of Diseases, Ninth Revision, code 155.0) were obtained from vital registration data for the years 1979–1996 (82).

Assessing goodness-of-fit. By sampling from the defined parameter ranges in multiple rounds, we generated thousands of different sets of candidate values. For a given set of parameter values, numerical simulation of the model was undertaken, and the HCV and HBV seroprevalence and HCC mortality rates produced by those particular parameter values were calculated. These predicted values were compared with the observed values, and goodness-of-fit was assessed by using separate likelihood statistics for HCV infection prevalence, HBV infection prevalence, and HCC mortality, as well as a composite score based on all of these likelihood statistics. Each likelihood statistic was computed by treating the observed data (counts of infection numbers or deaths) as realizations of binomial processes, and the composite score was calculated as the sum of the individual likelihood measures after rescaling each measure in reference to its maximum possible value.

Searching across parameter ranges. The search across the multidimensional parameter space defined by the specified ranges was implemented in three stages. In the first stage, intended as a broad search to identify particular regions for more precise focus, 50,000 parameter combinations were generated by sampling from independent uniform distributions spanning the specified ranges. For the second round of simulations, these 90 sets were used to generate 27,000 new sets of parameter values by first categorizing all model parameters as relating primarily to HCV infection, HBV infection, or progression and then generating all possible combinations of the HCV infection

<table>
<thead>
<tr>
<th>Author (reference no.)</th>
<th>No. of subjects</th>
<th>Mean follow-up (years)</th>
<th>% with cirrhosis</th>
<th>Study population</th>
</tr>
</thead>
<tbody>
<tr>
<td>Koretz et al. (65)</td>
<td>80</td>
<td>16.0</td>
<td>18.0</td>
<td>Non-A, non-B</td>
</tr>
<tr>
<td>Hopf et al. (67)</td>
<td>86</td>
<td>8.0</td>
<td>24.3</td>
<td>Non-A, non-B with presumed chronic liver disease</td>
</tr>
<tr>
<td>Tremolada et al. (68)</td>
<td>135</td>
<td>7.5</td>
<td>15.6</td>
<td>Posttransfusion non-A, non-B</td>
</tr>
<tr>
<td>Di Bisceglie et al. (69)</td>
<td>39</td>
<td>9.7</td>
<td>20.5</td>
<td>Chronic posttransfusion non-A, non-B</td>
</tr>
<tr>
<td>Seeff et al. (70)</td>
<td>568</td>
<td>18.0</td>
<td>21.0</td>
<td>Posttransfusion non-A, non-B</td>
</tr>
<tr>
<td>Tong et al. (27)</td>
<td>131</td>
<td>3.9</td>
<td>51.1</td>
<td>Chronic posttransfusion HCV* at referral center</td>
</tr>
<tr>
<td>Gronbaek et al. (71)</td>
<td>178</td>
<td>23.0</td>
<td>9.0</td>
<td>Community-acquired non-A, non-B</td>
</tr>
<tr>
<td>Takahashi et al. (72)</td>
<td>57</td>
<td>9.0</td>
<td>35.1</td>
<td>Chronic HCV</td>
</tr>
<tr>
<td>Kobayashi et al. (73)</td>
<td>136</td>
<td>9.6</td>
<td>27.9</td>
<td>Chronic HCV</td>
</tr>
<tr>
<td>Muller (64)</td>
<td>152</td>
<td>15.0</td>
<td>0</td>
<td>Women infected with HCV-contaminated immunoglobulin</td>
</tr>
<tr>
<td>Kenny-Walsh (74)</td>
<td>376</td>
<td>17.0</td>
<td>1.9</td>
<td>Women infected with HCV-contaminated immunoglobulin</td>
</tr>
<tr>
<td>Datz et al. (75)</td>
<td>20</td>
<td>18.0</td>
<td>20.0</td>
<td>Non-A, non-B outbreak in plasmapheresis center</td>
</tr>
<tr>
<td>Seeff et al. (76)</td>
<td>17</td>
<td>45.0</td>
<td>11.8</td>
<td>HCV-positive military recruits</td>
</tr>
</tbody>
</table>

* HCV, hepatitis C virus.

<table>
<thead>
<tr>
<th>Age (years)</th>
<th>Males</th>
<th>Females</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;20</td>
<td>0.00–0.05</td>
<td>0.00–0.04</td>
</tr>
<tr>
<td>20–29</td>
<td>0.00–0.10</td>
<td>0.00–0.08</td>
</tr>
<tr>
<td>30–49</td>
<td>0.03–0.15</td>
<td>0.01–0.12</td>
</tr>
<tr>
<td>50–59</td>
<td>0.05–0.20</td>
<td>0.01–0.16</td>
</tr>
<tr>
<td>60–69</td>
<td>0.10–0.40</td>
<td>0.02–0.32</td>
</tr>
<tr>
<td>≥70</td>
<td>0.20–0.50</td>
<td>0.04–0.40</td>
</tr>
</tbody>
</table>

* Rates are presented per person per year. Rates are fixed across fibrosis classes; that is, transitions from no fibrosis to portal fibrosis without septa occur at the same rate as transitions from portal fibrosis without septa to few septa, etc.
parameters from the 30 best fits to HCV prevalence with the HBV infection parameters from the 30 best fits to HBV prevalence and the progression parameters from the 30 best fits to HCC mortality \((30 \times 30 \times 30 = 27,000)\). From these second-round simulations, we selected the 50 combinations with the best composite fit scores to all of the data and generated 1,000 new parameter combinations from each by sampling from independent, uniform distributions spanning 90–110 percent of the starting values (within the constraints of the original ranges). From these 50,000 simulations, we identified a final set of 50 good-fit parameter combinations with the highest composite fit scores.

RESULTS

Model fit

The model produces a close match to NHANES III data on HCV prevalence (figure 3). Prevalence is significantly higher in males than in females but peaks in the age group 30–39 years in both sexes.

The model fit to vital registration data on HCC mortality is summarized in figure 4. The model fits best to the data in adults between ages 30 and 59 years, while it fails to reproduce the steepest rise in the data at the oldest age groups. The sharp increase in modeled mortality in younger adults results from aging and disease progression in persons infected between ages 20 and 30 years during the late 1960s and 1970s. The rise in mortality at these ages is consistent with the trends in infection inferred from fitting to the seroprevalence data. The lack of fit at the oldest ages, on the other hand, suggests that it may not be possible to reconcile reported mortality trends at these ages with the patterns of HCV and HBV infection inferred from NHANES III.

Implications of good-fit parameter sets

Past HCV infection trends. Across the range of good-fit models, the incidence of new infections rises sharply starting around 1960, then declines gradually after peaking in the 1980s (figure 5). Infection risks are substantially higher among males than among females, with a slightly older age pattern in females. The advent of screening for HBV markers in blood donors produces a sharp reduction in transfusion-related HCV infection after 1972 because of the relatively high occurrence of coinfection. The models depict dramatic declines in both community infection and transfusion-related infection since 1990, although recent estimates are more uncertain.

Fibrosis progression rates. Like others (27, 28, 79), we find that the rate of progression from HCV infection to cirrhosis depends strongly on age and sex. In the model with the highest composite fit score, the median time from infection to cirrhosis is 46 years for males infected at age 25, while in a cohort of females infected at age 25, fewer than 30 percent will progress to cirrhosis even after 50 years of infection.

The set of good-fit models indicates that a wide variety of plausible assumptions regarding progression rates is consistent with observed epidemiologic data. As an example, figure 6 illustrates two progression assumptions from among the 50 models with good fit, summarized in terms of the cumulative probability of progressing to various different endpoints in a cohort of males infected at age 25 years. In the first example, fewer than one fourth of the infected cohort will progress to cirrhosis before dying from other causes, while in the second example more than three fourths of the cohort will eventually progress to cirrhosis. This wide variation in the range of parameter values that provide a good fit to the data is consistent with the empirical findings in table 2. For the parameter governing the proportion of chronic HCV infections eligible for faster progression, the good-fit
values range from 0.26 to 0.90, nearly the full range across which this parameter was allowed to vary.

**HCV and HCC mortality.** Previous projections of a two- to fourfold increase in the number of deaths attributable to hepatitis C over the next 2–3 decades (21, 23) depend on progression assumptions drawn from the high end of the plausible range, but our model shows that more optimistic estimates may be equally plausible. Across the span of

---

**FIGURE 4.** Model fit to US vital registration data on death rates from primary hepatocellular carcinoma (HCC), 1970–1996. Each graph presents a comparison of the observed data (diamonds) with the range of model predictions for one age-sex group (upper and lower curves). Note that each graph is presented on a different scale.

---

**FIGURE 5.** Trends in the estimated incidence of acute hepatitis C virus infection in the United States, 1950–1997. The two lines show the range in the number of new infections across the 50 good-fit models.
models with good fit, the projected increase in liver cancer deaths between the years 2000 and 2020 ranges from approximately 60 percent to approximately 120 percent of the number of deaths in the year 2000 (figure 7). These projections may represent the upper bound of future mortality, since the model excludes the effects of treatment that might attenuate the predicted mortality trends.

The model results also indicate a shift over time in the relative contributions of HCV, HBV, and other causes to HCC mortality. In 1970, nearly 50 percent of HCC deaths occurred in persons with chronic HBV infection compared with approximately 20 percent in persons with chronic HCV infection and 30 percent in uninfected persons. In 2000, these proportions were 36, 39, and 24 percent, respectively.

DISCUSSION

We developed an empirically calibrated model of the HCV epidemic with the aim of taking full advantage of available clinical and epidemiologic information. The model was fit to data on prevalence of infection with HCV and HBV and mortality from primary HCC to investigate critical uncertainties surrounding the natural history and epidemiology of HCV infection. In developing future projections of the epidemic and examining the potential impact of different policy interventions, a minimal requirement might be that models use parameter values that are consistent with past epidemiologic trends.

Our results indicate that rates of progression to advanced liver disease may be lower than previously assumed. For example, retrospective studies of transfusion-associated infection have estimated the median duration between infection and cirrhosis to be approximately 21–22 years (27, 83), while a large cross-sectional study imputed the expected duration between infection and cirrhosis to be 38 years for persons infected between ages 21 and 30 years (28). In contrast, we found a median duration of 46 years for men infected at age 25 years and estimated that fewer than 30 percent of women infected at this age would ever develop cirrhosis. Higher progression rates in the previous studies may relate to older ages at infection in the case of the post-transfusion studies or to referral biases from the focus on patients with chronic hepatitis who had undergone biopsy in the cross-sectional study. Our finding of lower progression rates is consistent with the results of several recent cohort studies (66, 74, 76). As public health campaigns encourage persons with potential risk factors to be tested for HCV infection (19), the possibility that progression rates are lower than previously estimated will be an important consideration when making decisions about treatment recommendations and health policy.

Rates of fibrosis progression in the setting of chronic hepatitis C evidently depend on age and sex, but there appear to be other important, as yet unidentified, sources of heterogeneity in progression. We found that it was possible to fit observed epidemiologic trends under a wide range of different assumptions about heterogeneity of rates, which implies that conventional sensitivity analyses undertaken in

---

**FIGURE 6.** Cumulative probabilities of progressing to various endpoints after hepatitis C virus infection at age 25 years for males. The two graphs give two different examples drawn from the set of 50 parameter combinations with good fit. HCC, hepatocellular carcinoma.

**FIGURE 7.** Projected increase in deaths from primary hepatocellular carcinoma, United States, 2000–2020. Increases are expressed as a percentage of the number of deaths in the year 2000. The two lines in the graph represent the upper and lower bounds in each year across the 50 parameter sets with good fit.
various decision analytic models may inadequately reflect the true uncertainty about progression to advanced sequelae of HCV infection.

In this study, we also found that it may be impossible to reconcile reported patterns of liver cancer mortality at the oldest ages with observed data on the age pattern of HCV infection in the population. Vital registration data indicate a steep rise in death rates from primary HCC, even in persons aged 85 years and older, yet the past trends in infection that are consistent with infection seroprevalence data would be unlikely to produce concurrent increases at all ages. One possibility is that deaths attributable to other causes, such as colorectal cancer, have been misclassified as HCC. A comparison between the mortality rates from vital registration data and the expected mortality rates computed from the Surveillance, Epidemiology, and End Results (SEER) Program national cancer registry indicates higher reported death rates from liver cancer at the oldest ages in vital registration data compared with those of the SEER Program but the opposite result for colorectal cancer (84). Given the high rates of colorectal cancer mortality in the United States, the misclassification of a relatively small proportion of these deaths would produce a relatively large number of miscoded liver cancer deaths.

Another possibility is that changes in the use of diagnostic imaging techniques since the 1980s have led to an apparent increase in the diagnosis of HCC. This hypothesis is supported by a study of liver transplant recipients that found that unsuspected HCC was discovered by computed tomography in up to 10 percent of cirrhotic patients undergoing transplantation (85). Comparison with Canadian vital registration data on HCC mortality rates offers further support for this hypothesis, since death rates in Canada are similar in magnitude to those in the United States, but there has been little increase in Canadian rates in the 1990s, perhaps due to more limited use of imaging technologies.

This study adds to a growing body of research on developing analytic methods to improve model calibration (86–88), but it has several limitations. The specification of ranges around the model parameters, as a first step in the calibration of parameter values, required the synthesis of numerous disparate sources of data from various types of studies. We sought to define wide ranges around each parameter to accommodate the full span of plausible values, but more formal methods for developing prior distributions for each parameter might allow us to adopt a full Bayesian approach to model fitting. As additional information becomes available, some of the uncertainties identified in this paper may be clarified. For example, results from the ongoing fourth round of the National Health and Nutrition Examination Survey will provide critical information on the changing patterns of infection over the last decade, which remain poorly understood at present. Our multistage search procedure was one of many possible choices, and more systematic search algorithms might produce greater efficiency in identifying good-fit regions in parameter space.

The clinical management of patients with chronic HCV infection and the public health policy response to the epidemic both depend on having a clear understanding of the natural history and epidemiology of HCV. A range of mathematical models has provided a set of valuable tools to aid in the planning of a prudent response to the HCV epidemic (20–23, 89, 90). These models, however, have typically assumed inevitable and rapid progression after chronic HCV infection without validating this assumption in reference to observed disease patterns in the population. For example, the prediction that liver cancer deaths will increase almost fourfold between 1998 and 2018 (21) is based on the assumption that nearly everybody with chronic HCV infection will eventually progress to cirrhosis. Elsewhere, Alter and Seeff (7) have proposed that only a minority of infections lead to severe, progressive liver disease, and the empirical calibration of our model progression parameters using reported trends in primary liver cancer mortality lends support to this hypothesis. This study has allowed us to explore some of the key uncertainties that remain in our understanding of the natural history of HCV and its impact in the US population. The next steps will be to use the insights from the model to develop forecasts of the future burden of HCV and to examine the likely costs and benefits of alternative interventions.

ACKNOWLEDGMENTS

Supported in part by a training grant from the former Agency for Health Care Policy and Research.

The authors gratefully acknowledge helpful comments and advice from Dr. Grace Lee, Dr. Christopher Murray, Dr. Craig Earle, Dr. Cynthia Boschi-Pinto, Dr. Catherine Michaud, Maria Bulzacchelli, Dr. Karen Kuntz, and Dr. Jean Marie Arduino.

REFERENCES


APPENDIX

There are three categories of infection risks in the model: vertical (mother-to-child) transmission, transfusion-associated infection, and community infection.

Vertical transmission

The distribution of newborn babies across different infection categories depends on the infection category of the mother, the probabilities of vertical transmission of HCV and HBV infection, and the probabilities that acute infection will lead to chronic infection. For example, all babies born to mothers who do not have an HCV or HBV infection will enter the model without infection, while babies born to mothers with chronic HCV infection but no HBV infection may enter the model with no infection, chronic HCV infection, or resolved HCV infection. In the case of coinfected mothers, we assume independent probabilities for HCV and HBV transmission.

Transfusion-associated infection

Risks of transfusion-associated infection with HCV or HBV depend on the incidence of blood use and the probability that a unit of blood is infected. In principle, the probability that a blood unit is infected is a function of the prevalence of infection in the population; however, the per-unit prevalence is unlikely to equal the population prevalence, given blood screening protocols and nonrandom donor selection. We have therefore treated transfusion risks as exogenous parameters based on empirical trends in total utilization of blood and blood components in the United States (91), the age pattern of incidence of blood use (92–94), results from serologic screening of blood donors (95–97), and a multicenter prospective study of cardiac surgery patients (98). Ranges for parameter values governing trends in risk levels were specified as follows: 1) prior to 1970, HCV and HBV infection risks were approximately 0.8–1.2 percent per unit of blood transfused (99); 2) with the advent of an assay for the hepatitis B surface antigen (HBsAg) in 1970, a Food and Drug Administration mandate to screen all donations for HBsAg in 1972, and the shift from paid to volunteer blood donation in the early 1970s, there was a total decline between 1970 and 1972 of 30–75 percent in both HCV and HBV transfusion-related infection risks (99); 3) improved assays for HBsAg and the exclusion of high-risk donors (those with acquired immunodeficiency syndrome-related symptoms or risk behaviors, such as injection drug use) produced average declines in risk through the late 1970s and early 1980s of 25–29 percent per year for HBV and 0.5–2 percent per year for HCV (98, 99); 4) testing for surrogate HCV markers beginning in 1986 led to accelerated declines in HCV infection risks of 30–40 percent per year and similar rates of continued decline in HBV (98); 5) licensing and implementation of HCV screening starting in 1990, with continued improvements in HCV assays since that time, have led to dramatic reductions in the risk of HCV infection, at rates of approximately 70–85 percent per year (96, 97) starting in 1990.

Community infection

As with transfusion risks, levels of community transmission risks should, in theory, relate endogenously to infection prevalence in the model; however, modeling transmission endogenously would add considerable complexity and require a number of additional parameters and assumptions for which there is scant empirical basis, including average contact rates and mixing patterns for uninfected and infected persons in different subgroups and per-contact infection probabilities. We have therefore modeled community infection risks without specifying an explicit transmission model.

A flexible parametric curve was used to define distinct age patterns of incidence of HCV and HBV infection for males and females, with the peak age and shape of the curve varying according to four parameters that determine the location and width of a plateau in risk levels and the rate of (exponential) decline moving in either direction away from this plateau. Although risk factor studies for HCV and HBV infection offer solid evidence that community risks for both...
are concentrated predominantly in young adults, the exact age pattern remains uncertain. In the model-fitting procedure, we have allowed the peak age group for risk to begin between ages 20 and 30 years for both HCV and HBV and the width of the plateau to range from 5 to 10 years or 7 to 20 years for HCV and HBV, respectively, which accommodates slightly older age patterns of HBV infection due to differences in risk factors. Rates of decline from the peak levels ranged from 80 to 95 percent per year of age below the peak and from 70 to 90 percent per year of age above the peak for both HCV and HBV.

There is considerable uncertainty about time trends in community risk levels. For HCV infection, the dominant hypothesis postulates a steep increase in the 1960s and 1970s due to high rates of injection drug use, but the exact timing and levels are unknown. For HBV, population seroprevalence data from 1976 to 1994 suggest less dramatic changes (100). We explored two different formulations for community infection trends: 1) constant risk up to 1960, followed by linear change over each subsequent decade; and 2) a smooth infection curve, adapted from the study by Armstrong et al. (20), which may change direction twice based on six parameters (the years in which risk begins to rise, peaks, and attains a new postpeak equilibrium level; and the levels of risk before the rise, at the peak, and in the new equilibrium). The smoothed curve provided better fits to observed prevalence data for HCV, while the linear model fit better for HBV. Results presented in this paper reflect the selection of the most appropriate model in each case.

Ranges around the parameters determining community infection trends were specified as follows, with values referring to risk levels in the peak age group, and risks in other age groups were computed based on the relative age patterns described above: 1) for HCV, we assumed a prerise level of between 0 and 0.2 percent for males and between 0 and 0.1 percent for females, a peak level of between 0.2 and 1 percent for males and between 0.1 and 0.5 percent for females, and a postpeak equilibrium level of 0–1 percent for males and 0–0.5 percent for females. We assumed that the rise in risk began between 1955 and 1965, the peak occurred between 1975 and 1985, and the new equilibrium would occur between 2000 and 2010; 2) for HBV, we assumed that the risk level before 1960 was 0.3–0.7 percent in males and 0.15–0.5 percent in females and that the slope of change during each subsequent 10-year period has ranged from –5 to 3 percent of the starting level for that decade.