Journal of Korean Medical Science
Vol 6, No. 4, 319-324 December 1991

A Reconstructed Cohort Study on the Hepatitis B Virus Infection as a Risk Factor of Liver Cancer in Korea[1]

Keun-Young Yoo, M.D., Heon-Kim, M.D., Moo-Song Lee, M.D., Byung-Joo Park, M.D., Yoon-Ok Ahn, M.D., Hyo-Suk Lee, M.D.[1)], Chung-Yong Kim, M.D.[1)] and Tae-Soo Park, M.P.H.[2)]

Department of Preventive Medicine and Department of Internal Medicine[1)], Seoul National University College of Medicine, Seoul, Korea, and Korea Medical Insurance Corporation[2)]

A stable cohort (n = 369,725) was reconstructed to test the association of hepatitis B virus (HBV) infection with subsequent occurrence of liver cancer in a healthy Korean population. The cohort consisted of male beneficiaries of the Korea Medical Insurance Corporation over 30 years of age, living nationwide. The HBV infection was tested by the reversed passive hemagglutination method for the HBsAg and by the passive hemagglutination method for the anti-HBs at the time of recruitment in 1984. Admissions due to liver diseases were identified through a computerized system for reimbursements on medical insurance claims from January 1, 1985 to December 31, 1987. For a more valid estimate of incidence, a sample survey was carried out, obtaining correction coefficients against misclassification of the diagnosis, as well as those for duplicate claims in a year. The incidence rate of liver cancer steadily increased with age ($\chi^2_{TREND} = 51.1$, df = 1, p = 0.00). The highest rate was estimated to be in the age group of 55-59 (71.8 per 100,000 person-year), which was 2.46 (95% CI = 1.79-3.89) times higher than that of the age group under 34. The rate showed a decreasing trend when the income level increased, which was quite linear ($\chi^2_{TREND} = 221$, df = 1, p = 0.00). The incidence rate of liver cancer was strongly related with the infection status by the HBV at the time of recruitment ($\chi^2_{LR} = 200$, df = 3, p = 0.00), when the potential confounding effects of both age and income level were simultaneously controlled with a log-linear model for a constant hazard. Compared with the susceptibles to the HBV, the risk of liver cancer was significantly reduced among the anti-HBs positives (adjusted RR = 0.71, 95% CI = 0.54-0.93), while the relative risk of liver cancer among the HBsAg positives was the greatest (adjusted RR = 5.71, 95% CI = 4.59-7.10). This study confirmed the relationship between HBV infection and liver cancer in Korea.

Key Words: *Cohort, Hepatitis B virus, Liver cancer, Incidence rate, Record linkage study*

INTRODUCTION

Liver cancer might be causally related to the hepatitis B virus (HBV) infection in man. There is much evidence that HBV is the most important factor worldwide in the occurrence of liver cancer. However, such a hypothesis has not yet been proven with definite evidence due to the lack of a proper animal model for the demonstration of the causality. Most of the efforts, therefore, are based on observations of the human population in epidemiologic studies.

A prospective cohort study conducted with 22,707 male governmental servants in Taiwan supported a

Address for correspondence: *Keun-Young Yoo; Department of Preventive Medicine, Seoul National University College of Medicine, 28 Yongon-dong, Chongno-gu, Seoul 110-460, Korea. Phone: (02) 760-3303.*

[1]This work was supported in part by a research grant from Seoul National University Hospital (1988) and in part a grant from Jinro cultural Foundation (1989).

positive relationship between the HBV infection and liver cancer (Beasley et al., 1981). Similar studies on the same topic have been conducted in various countries to observe the relationship among different populations (Obata et al., 1980; Sakuma et al., 1982; Heyward, et al., 1982; Feinman et al., 1982; Musca et al., 1983). Although a vaccine against HBV had already been developed in Korea (Kim, 1979), incidence and prevalence of liver cancer among Koreans are very high (IARC, 1987; Ahn et al., 1990), and the HBV infection rate happens to be also very high in Korea (Choi, 1986; Park et al., 1987; Yoo et al., 1988; Yoo et al., 1990). Based on the fact of such correlation, it may be speculated that these 2 entities seem to be associated with each other in this population. The objective of this study was to test the association of HBV infection with the subsequent occurrence of liver cancer through a reconstructed cohort study of a male adult population in Korea.

MATERIALS AND METHODS

The study population was selected among the beneficiaries of the Korea Medical Insurance Corporation (KMIC), consisting of governmental employees, private schoolteachers and staffs, and their pensioners, living nationwide. The eligible population for the male cohort over 30 years of age was 516,668 persons in 1984. Physically unhealthy persons who needed further medical attention by physician and those with abnormal liver function, those who had vaccination against HBV prior to the recruitment, those without records on HBV markers, and those who migrated in or out of the population during the observation period of 1984-1986 were excluded. A stable and liver disease-free study population (n=369,742) was finally reconstructed as of 1984.

The data on the occurrence of liver cancer were collected from a data file of the KMIC. When an insurance claim for medical service is reimbursed, a computerized record is entered into an automated "Treatment File," which contains information such as ID number, name of the patient, a 3-digit code of ICD of the 2 presumptive diagnoses, date of admission, names of the hospitals and clinics, etc. Records on any medical problem were collected from January 1, 1985 to December 31, 1987.

Linking was done by an ID number between the "Treatment File of 1985-1987" and the "Physical Examination File of 1984." The latter contained records of each individual's HBV seromarkers, tested by the reversed passive hemagglutination method (RPHA) for the HBsAg, and by the passive hemagglutination method (PHA) for the anti-HBs. Admitted cases with a diagnosis of ICD-070 (acute viral hepatitis) or with ICD-571 (chronic active hepatitis, chronic persistent hepatitis and liver cirrhosis) or with ICD-155 (liver cancer) were regarded as presumptive cases of liver cancer.

In order to select the true cases among the presumptives and to unify the diagnostic criteria of the disease, a sample survey was carried out in 1988. Eight hospitals located in the Seoul area were chosen for the survey. Seventeen senior students of the Seoul National University College of Medicine visited each hospital, in which a presumptive case had once been admitted. They abstracted 572 medical records on a standard dictation sheet. All medical abstracts were reviewed by a hepatologist.

The incidence rate was measured by the estimated number of liver cancer cases divided by the 100,000 person-year observation. The actual number reported to the KMIC was corrected by the coefficient against the misclassification error between the true cases and cases with ICD-070 (α), ICD-571 (β), and ICD-155 (γ). Correction against duplicate claims (ϵ_1, ϵ_2, ϵ_3) was also done for the valid estimation of annual incidences. Correction coefficients were obtained from the sample survey as described above. Values and calculation procedure for the correction were as follows:

Incidence rate of liver cancer:

$$\frac{\{C_{070}\times\alpha\times\epsilon_1\}+\{C_{571}\times\beta\times\epsilon_2\}+\{C_{155}\times\gamma\times\epsilon_3\}}{N_1\times k}\times 100{,}000$$

N_1 : number of cohort observed
k : years of observation (=3)
C_{070} : number of claims with the presumptive diagnosis of ICD 070
C_{571} : number of claims with the presumptive diagnosis of ICD 571
C_{155} : number of claims with the presumptive diagnosis of ICD 155
α : proportion of the most probable liver cancer cases among C_{070} (=0.000)
β : proportion of the most probable liver cancer cases among C_{571} (=0.037)
γ : proportion of the most probable liver cancer cases among C_{155} (=0.944)
ϵ_1 : ratio of the number of admissions to the number of claims with C_{070} (=1.000)
ϵ_2 : ratio of the number of admissions to the number of claims with C_{571} (=0.813)
ϵ_3 : ratio of the number of admissions to the number of claims with C_{155} (=0.659)

Given the incidence rate of the lowest stratum of each variable as a reference, the relative risk in each stratum and its 95% confidence limits were calculated from the regression coefficient, β, and its standard error in the log-linear model.

The model had been developed for the analysis of rates and constant hazards in a follow-up study (Holford, 1980). The relative risk of HBV infection was measured as an adjusted risk for both age and income level. The likelihood ratio test was applied to test the global null hypothesis of $H_0: \beta_1 = \beta_2 = \ldots = \beta_i = 0$ by the log-linear model. A likelihood ratio test for linear trend was done to assess a dose-dependent relationship between the risk factor and the logit risk (Breslow and Day, 1980; Holford, 1984). The SAS and the GLIM system were statistical systems used for the analysis (SAS Institute Inc., 1988; The GLIM Working Party, 1987).

RESULTS

Incidence rate of liver cancer by age group

Incidence rates of liver cancer by age group are summarized in Table 1. The rate under the age of 34 was 27.2 per 100,000 person-years. The age group of 35-39 showed 30.5 per 100,000 person-years, which was 1.12 times higher than the reference level of the age group under 34. However, it was statistically nonsignificant (95% CI = 0.80-1.57). The relative risk steadily increased with age groups afterwards; the risk was 1.76 (95% CI = 1.29-2.38) for the age group of 40-44, 2.23 (95% CI = 1.66-3.00) for the age group of 45-49, 2.50 (95% CI = 1.82-3.42) for the age group of 50-54, 2.64 (95% CI = 1.79-3.89) for the age group of 55-59, and 1.85 (95% CI = 0.88-3.89) for the age group of 60-64, respectively. As a whole, the incidence rate of liver cancer was strongly related to age ($\chi^2_{LR} = 65.6$, df = 7, p = 0.00) and showed a linearly increasing pattern with age ($\chi^2_{TREND} = 51.1$, df = 1, p = 0.00).

Incidence rate of liver cancer by income level

Using the lowest income group as a reference, all the relative risks of the remaining groups were significantly reduced to around 0.2 (Table 2). It is likely that income level was strongly associated with the risk of liver cancer ($\chi^2_{LR} = 345$, df = 4, p = 0.00).

It apparently decreased with the increase of income level, which was quite linear ($\chi^2_{TREND} = 221$, df = 1, p = 0.00). However, such significances fell into nonsignificant level when adjustment for HBV infection was done. It implies that HBV infection status was definitely a confounder in the association between income level and liver cancer.

Table 1. Relative Risk with 95% Confidence Intervals for Liver Cancer by Age Group from a Stable Cohort Consisting of Korean Males over 30 Years of Age Reconstructed as of 1984 in Korea, 1985-1987

Age Groups	No. of Cohort[1]	No. of Cases[2]	I.R.[3]	R.R. (95% CI)[4]
-34	83,258	68	27.2	1.00
35-39	76,441	70	30.5	1.12 (0.80-1.57)
40-44	73,250	105	48.0	1.76 (1.29-2.38)
45-49	68,647	125	60.7	2.23 (1.66-3.00)
50-54	43,174	88	67.9	2.50 (1.82-3.42)
55-59	19,028	41	71.8	2.64 (1.79-3.89)
60-64	5,293	8	50.4	1.85 (0.88-3.89)
65-	651	0	0.0	0.00 (-)
			χ^2_{LR} [5] = 65.6, df = 7, p = 0.00	
			χ^2_{TREND}[6] = 51.1, df = 1, p = 0.0	

1) Number of cohorts at the time of recruitment, 1984
2) Number of clinical liver cancer cases during 1985-1987 estimated by correction coefficients for both misclassification of disease and duplicate claims
3) Incidence rate per 100,000 person-year
4) Unadjusted relative risk and its 95% confidence intervals estimated from regression coefficient and its standard error
5) Likelihood ratio statistic by log-linear model for constant hazard
6) Chi-square for linear trend calculated by log-linear model for constant hazard

Relative risk of liver cancer by the HBV infection status

The incidence of liver cancer was strongly related with the infection status by the HBV at the time of recruitment ($\chi^2_{LR}=200$, df=3, p=0.00), when the potential confounding effects of age and income level were simultaneously controlled by a log-linear model (Table 3). Compared to the susceptibles to HBV, the relative risk of the anti-HBs positives was significantly reduced (adjusted RR=0.71, 95% CI=0.54-0.93), while the risk among the HBsAg positives was the greatest (adjusted RR=5.71, 95% CI=4.59-7.10).

Table 2. Relative Risk with 95% Confidence Intervals for Liver Cancer by Income Levels[1] from a Stable Cohort Consisting of Korean Males over 30 Years of Age Reconstructed as of 1984 in Korea, 1985-1987

Income Levels	No. of Cohort[2]	No. of Cases[3]	I. R.[4]	R.R. (95% CI)[5]
Lowermost	41,969	269	213.6	1.00
Lower	51,240	69	44.9	0.21 (0.16-0.27)
Middle	197,842	239	40.3	0.19 (0.16-0.22)
Upper	80,254	133	55.2	0.26 (0.21-0.32)
Uppermost	291	0	0.0	0.03 (0.00-96.1)
			χ^2_{LR} [6] = 345, df = 4, p = 0.00	
			χ^2_{TREND} [7] = 221, df = 1, p = 0.00	

1) Income levels were classified by standard monthly wages
2) Number of cohorts at the time of recruitment, 1984
3) Number of clinical liver cancer cases during 1985-1987 estimated by correction coefficients for both misclassification of disease and duplicate claims
4) Incidence rate per 100,000 person-year
5) Unadjusted relative risk and its 95% confidence intervals estimated from regression coefficient and its standard error
6) Likelihood ratio statistic by log-linear model for constant hazard
7) Chi-square for linear trend calculated by log-linear model for constant hazard

Table 3. Relative Risk with 95% Confidence Intervals for Liver Cancer by Serologic Profiles[1] of Hepatitis B Virus Infection Tested at the Time of Recruitment from a Stable Cohort Consisting of Korean Males over 30 years of Age Reconstructed as of 1984 in Korea, 1985-1987

Profiles		No. of Cohort[2]	No. of Cases[3]	I.R.[4]	R.R. (95% CI)[5]
HBsAg	antiHBs				
(−)	(−)	274,029	332	40.4	1.00
(−)	(+)	78,523	65	27.6	0.71 (0.54-0.93)
(+)	(−)	16,382	107	217.7	5.71 (4.59-7.10)
(+)	(+)	544	3	183.8	5.11 (1.70-15.3)
				χ^2_{LR} [6] = 200, df = 3, p = 0.00	
				χ^2_{TREND} [7] = 100, df = 1, p = 0.00	

1) Tested by RPHA for HBsAg and by PHA for anti-HBs
2) Number of cohorts at the time of recruitment, 1984
3) Number of clinical liver cancer cases during 1985-1987 estimated by correction coefficients for both misclassification of disease and duplicate claims
4) Incidence rate per 100,000 person-year
5) Adjusted relative risk for age and income levels and its 95% confidence intervals estimated from regression coefficient and its standard error
6) Likelihood ratio statistic by log-linear model for constant hazard
7) Chi-square for linear trend calculated by log-linear model for constant hazard

The risk of positives for both HBsAg and antiHBs was 5.11 times greater than the reference level (95% CI = 1.70-15.3). It was also evident that the risk of subsequent occurrences of liver cancer steadily increased with the HBV infection status ($\chi^2_{TREND} = 100$, df = 1, p = 0.00).

DISCUSSION

This study confirmed the relationship between the HBV infection and liver cancer in Korea. The results can be validated based on the fact that those findings were drawn from a cohort, which was one of the largest ever conducted in the world. It was possible to make a reconstructed-type cohort for such a short period of time, since every record on HBV infection status at the time of recruitment was well-preserved in a computer file.

The transmissibility of malignant sarcoma by cell-free tumor filtrates was demonstrated in 1911 in the chicken embryo by Rous, who was awarded the Nobel Prize in 1972. Since that notable report, Ebstein-Barr virus (EBV) in the cultured cells of an African tumor of children, EBV in infectious mononucleosis and in nasopharyngeal cancer, herpes simplex type 2 (HSV-2) in cervical neoplasia, papillomavirus in invasive cervical carcinoma, HBV and hepatocellular carcinoma, and, recently, human retrovirus (HTLV-I) in adult T-cell leukemia have been recognized or suspected as etiologic agents of various cancers in humans (Evans, 1982; Blattner, 1990).

The HBV infection as a risk factor of liver cancer had long been suspected due to earlier observations on an international correlation between hepatocellular carcinoma incidences and prevalences of HBV infection, particularly in the HBsAg carrier state; the increased frequency of chronic HBV infection in liver cancer patients; and the association of chronic HBV infection with cirrhosis and chronic liver disease, often in conjunction with liver cancer (Szmuness, 1978). Several case-control studies on the relationship revealed that the chronicity of the HBV infection, rather than the acute status of the infection, can be more attributable to the disease (Szmuness, 1978).

Methodologically, temporal ambiguity between the exposure and the occurrence of the disease must be the most serious disadvantage of case-control study (Schlesselman,1982). Moreover, many epidemiologists believe that serious biases that may arise from the case-control setting, for example, the selection bias or the recall bias, can be avoided in the cohort study.

In order to obtain a valid estimate of the health index using the insurance data, a verification procedure should be applied, both to uncover inaccuracies in the diagnostic report and any double counting (Yoo et al.,1990). Correction coefficients were obtained from a sample survey in Seoul, which may result in an biased estimates of incidence. If misclassifications could occur similarly between Seoul and the other area, there might be no bias;otherwise, underestimation or overestimation might come out. Another source of invalidity in estimating true incidences in this population is the problem of predictability of RPHA-PHA tests for HBV infection. The RPHA-PHA method is known to be less specific than RIA method (Kim et al., 1984; Park, 1987; Yoo,1988). In addition, there was no record on the anti-HBc serologic marker in the study population. If they were false-negative, and if our hypothesis was true, the results of this study then would be falsely biased towards the null value, an underestimation of the genuine relative risk. Since those findings were from the physically-healthy population at the time of recruitment, the incidence rates in the results could underestimate the true incidences of the whole population.

Results from a cohort study in Taiwan reported a surprisingly high incidence of hepatocellular carcinoma among the HBsAg positives (relative risk = 223) compared to the negatives (Beasley et al., 1981). They followed 22,707 persons among a general population composed of male governmental servants for an average of 3.3 years. Differences in the cohort selection, the case ascertainment procedure, and the number of person-year observations may be attributed to the big differences in relative risks. Otherwise, those are compatible with our study results of positive association with liver cancer in Orientals. It should be pointed out that the subsequent occurrence of liver cancer was detected only after 2 to 4 years, so that a part of the incident cases, if ever, might be undetected at the end of the study. Further observation for longer period of time should be undertaken to establish a more comprehensive result on the epidemiologic relationship between HBV and liver cancer.

The current concept of the pathogenesis of liver cancer is that HBV infection in early life, possibly perinatally transmitted from the mother, leads to chronic antigenemia, to chronic hepatitis, to cirrhosis, and finally to hepatocellular carcinoma. Host and environmental factors may play a role in the HBV-liver cancer relationship. The age of infection and genetic susceptibility have been mentioned. Aflatoxin may act synergistically with the HBV. Malnutrition, malaria, and intestinal parasites may be cofactors. Whether HBV is directly oncogenic, with or without cofactors, or whether

a tumor can arise as an effect of a prolonged low-grade infection is not known (Evans, 1982).

Acknowledgements

The authors would like to thank the Korea Medical Insurance Corporation for its cooperation, as well as Mr. Dong-In Park and Mr. Poong-Min Kim in the Systems Engineering Research Institute, the Korea Advanced Institute of Science and Technology (KAIST) for their technical assistance in the data preparation, Mrs. Mi-Hyung Lee for her sincere effort in data processing, and, lastly, the 17 senior medical students of the Seoul National University College of Medicine for their data collection in verification procedure of diagnosis.

REFERENCES

Ahn YO, Park BJ, Yoo KY, et al.: *Incidence estimation of primary liver cancer among Koreans. J Korean Cancer Asso 21:241-248, 1989.*

Blattner WA: *Human retrovirology: HTLV. Raven Press, New York, pp1-14, 1990.*

Beasley RP, Hwang LY, Lin CC, et al.: *Hepatocellular carcinoma and hepatitis B virus; a prospective study of 22,707 men in Taiwan. Lancet 2:1129-1132, 1981.*

Blumberg BS, Larouze B, London WT, et al.: *The relation of infection with the hepatitis B agent to primary hepatic carcinoma. Am J Pathol 81:669-682, 1975.*

Breslow NE, Day NE: *Statistical methods in cancer research. Vol 1. The analysis of case-control studies. (IARC scientific publication number 32). Lyon: International Agency for Research on Cancer, 1980.*

Choi BY: *Incidence of hepatitis B virus infection of the school children in a rural area of Korea. Korean J Prev Med 19:281-292, 1986.*

Evans AF: *Viruses. In: Schottenfeld D, Graumeni JF, eds. Cancer epidemiology and prevention. W.B. Saunders Company, Philadelphia, London, Toronto, Mexico City, Rio de Janeiro, Sydney, Tokyo. 364-390, 1982.*

Feinman SV, Berris B, Cooter N, et al.: *Results of a long term prospective study of the hepatitis B surface antigen carrier state. Hepatogastroenterology 29:58-61, 1982.*

Holford TR: *The analysis of rates and of survivorship using log-linear models. Biometrics 36:299-305, 1980.*

Holford TR: *Strategies for the analysis of case-referent and cohort studies. In Perinatal epidemiology edited by Bracken MB. New York, Oxford, Oxford University Press, 1984.*

Heyward WL, Bender TR, Lanier AP, et al.: *Serologic markers of hepatitis B virus and alpha-fetoprotein levels preceding primary hepato-cellular carcinoma in Alaskan Eskimos. Lancet 2:889-891, 1982.*

International Agency for Research on Cancer: *Cancer incidence in five continents, Vol. V. Lyon: IARC Scientific Publications No. 88, 1987.*

Kim CY: *Prevention of viral hepatitis, type B, by vaccination with purified hepatitis B surface antigen. J Korean Med Assoc 22:1013-1025, 1979.*

Kim CY, Lee HS, Yu PC, et al.: *Comparison of radioimmunoassay and reversed passive hemagglutination assay in the detection of hepatitis B surface antigen. J Korean Med Assoc 27:49-53, 1984.*

Musca A, Cordova C, Barnaba V, et al.: *HBeAg/antiHBe system and development of primary hepatocellular carcinoma in patients with HBsAg-positive liver cirrhosis. Hepato-gastroenterol 30:3-5, 1983.*

Obata H, Hayashi N, Motoike Y, et al.: *A prospective study on the development of hepato-cellular carcinoma from liver cirrhosis with persistent hepatitis B virus infection. Int J Cancer 25:741-747, 1980.*

Park BJ: *A study on the validity of passive hemagglutination test for HBsAb. Korean J Prev Med 20:114-119, 1987.*

Park JH, Youn TH, Chun BY, et al.: *Hepatitis B virus infection rate of medical school students in Taegu. Korean J Prev Med 10:129-136, 1987.*

Sakuma K, Takahashi T, Okuda K, et al.: *Prognosis of hepatitis B virus surface antigen carriers in relation to routine liver function tests. A prospective study. Gastroenterol 83:114-117, 1982.*

SAS Institute Inc.: *SAS/STAT user's guide, release 6.03 edition. Carry, NC: SAS Institute Inc., 1988.*

Schlesselman JJ: *Case-control studies. Design, conduct, analysis. New York, Oxford: Oxford University Press, 1982.*

Szmuness W: *Hepatocellular carcinoma and hepatitis B virus; evidence for causal association. Prog Med Virol 24:40-69, 1978.*

The GLIM Working Party: *The generalised linear interactive modelling system, release 3.77. Oxford: Numerical Algorithms Group Ltd., 1987.*

Yoo KY: *Predictability of the test methods used in the detection for hepatitis B virus infection in a defined population. Seoul J Med 29:143-148, 1988.*

Yoo KY, Park BJ, Ahn YO: *Seroepidemiology of hepatitis B virus infection in healthy Korean adults in Seoul. Korean J Prev Med 21:89-98, 1988.*

Yoo KY, Park BJ, Ahn HS, et al.: *Incidence rate of hepatitis B: results of a record linkage study among healthy Korean males. J Korean Med Sci 5:179-183, 1990.*