



OPEN ACCESS

EDITED BY

Olumide Babatope Longe,
Academic City University
College, Ghana

REVIEWED BY

Kayode Oshinubi,
Bielefeld University, Germany
Yuzhu Dai,
The 903rd Hospital of the PLA, China

*CORRESPONDENCE

Mijeong Kim
✉ m.kim@ewha.ac.kr
Jungsoon Choi
✉ jungsoonchoi@hanyang.ac.kr

SPECIALTY SECTION

This article was submitted to
Infectious Diseases: Epidemiology and
Prevention,
a section of the journal
Frontiers in Public Health

RECEIVED 13 October 2022

ACCEPTED 06 December 2022

PUBLISHED 20 January 2023

CITATION

Jeong J, Kim M and Choi J (2023)
Investigating the spatio-temporal
variation of hepatitis A in Korea using a
Bayesian model.
Front. Public Health 10:1085077.
doi: 10.3389/fpubh.2022.1085077

COPYRIGHT

© 2023 Jeong, Kim and Choi. This is
an open-access article distributed
under the terms of the [Creative
Commons Attribution License \(CC BY\)](#).
The use, distribution or reproduction
in other forums is permitted, provided
the original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution
or reproduction is permitted which
does not comply with these terms.

Investigating the spatio-temporal variation of hepatitis A in Korea using a Bayesian model

Jaehong Jeong^{1,2}, Mijeong Kim^{3*} and Jungsoon Choi^{1,2*}

¹Department of Mathematics, Hanyang University, Seoul, Republic of Korea, ²Research Institute for Natural Sciences, Hanyang University, Seoul, Republic of Korea, ³Department of Statistics, Ewha Womans University, Seoul, Republic of Korea

Hepatitis A is a water-borne infectious disease that frequently occurs in unsanitary environments. However, paradoxically, those who have spent their infancy in a sanitary environment are more susceptible to hepatitis A because they do not have the opportunity to acquire natural immunity. In Korea, hepatitis A is prevalent because of the distribution of uncooked seafood, especially during hot and humid summers. In general, the transmission of hepatitis A is known to be dynamically affected by socioeconomic, environmental, and weather-related factors and is heterogeneous in time and space. In this study, we aimed to investigate the spatio-temporal variation of hepatitis A and the effects of socioeconomic and weather-related factors in Korea using a flexible spatio-temporal model. We propose a Bayesian Poisson regression model coupled with spatio-temporal variability to estimate the effects of risk factors. We used weekly hepatitis A incidence data across 250 districts in Korea from 2016 to 2019. We found spatial and temporal autocorrelations of hepatitis A indicating that the spatial distribution of hepatitis A varied dynamically over time. From the estimation results, we noticed that the districts with large proportions of males and foreigners correspond to higher incidences. The average temperature was positively correlated with the incidence, which is in agreement with other studies showing that the incidences in Korea are noticeable in spring and summer due to the increased outdoor activity and intake of stale seafood. To the best of our knowledge, this study is the first to suggest a spatio-temporal model for hepatitis A across the entirety of Korean. The proposed model could be useful for predicting, preventing, and controlling the spread of hepatitis A.

KEYWORDS

hepatitis A virus, spatio-temporal analysis, spatio-temporal models, zero-inflated Poisson, Bayesian hierarchical modeling, Korea

1. Introduction

Unlike Hepatitis B or C, Hepatitis A virus (HAV) is not transmitted through blood, but by consuming food or water contaminated with HAV or by contact with an infected person (1). World Health Organization (1) reported that the number of deaths from HAV in 2015 was estimated to be about 11,000 worldwide, accounting for 0.8% of deaths from viral infections. A case-control study of the HAV outbreak in Shanghai in 1988 supported that clams were a carrier of the virus (2). In Korea, HAV is often reported to spread through shellfish consumption (3). In India, a case-control study showed a high association between pipe water contamination and HAV infection (4). It particularly occurs in underdeveloped areas where personal hygiene management is poor, and HAV infection cases are decreasing in countries with an improvement in socioeconomic level, clean water management system, and HAV vaccination (5). However, the incidence rate has recently increased rapidly, in young adults who grew up in a hygienic environment in Korea (6).

Several studies have investigated the effects of socioeconomic and epidemiological factors such as age, medical level, and hygiene level on HAV in various countries. For example, it was noted that a significantly lower rate of HAV infection in people coupled with moderate to high socioeconomic conditions in Brazil, Argentina, and Mexico in Tapia-Conyer et al. (7). Mantovani et al. (8) discovered that a region with a high incidence of HAV had a weak socioeconomic condition in Brazil, thus emphasizing the need for hygiene improvement and better water treatment in the western Brazilian Amazon to reduce infectious disease outbreaks. Dogru et al. (9) classified children under the age of 15 in Turkey into three categories and analyzed the spatial patterns of HAV occurrence. In Turkey, the incidence of HAV was reported to be high in areas where water and sewage facilities are not well equipped. Copado-Villagrana et al. (10) pointed out that HAV infections were mainly found in the metropolitan areas of southern and western Mexico, noting that it may be associated with poor medical services in the most marginalized areas. Zheng et al. (11) characterized changes in the incidence and mortality of HAV in various age groups and regions in China from 1990 to 2018 and evaluated the effectiveness of the nationwide expanded program on immunization. The spread of the disease was decreased by expanding vaccinations and improving hygiene facilities. Shanmugam et al. (12) showed that primary infection of HAV among the older age group in India has recently decreased with improved living conditions.

Weather-related variables are associated with the incidence of HAV. According to Cann et al. (13), in extreme weather conditions such as hurricanes, cross-contamination of water supply and sewage may affect the transmission of waterborne diseases. In Brazil, cases of HAV infection increased during the

rainy season (14). In state of Pará, Brazil, monthly accumulated precipitation was positively correlated with the incidence of HAV (15). Tosepu (16) found a strong relationship between HAV and weather change, particularly rainfall and floods, in several areas, such as Spain, India, China, and Egypt. Baek et al. (17) showed that weekly precipitation and maximum temperature tended to decrease the incidence rate ratio of HAV in Seoul, Korea from the analysis of time series with past 1–6 week lags. In Seoul, the capital of South Korea, nearly 100% of households receive sterilized tap water and most citizens drink purified or clean spring water. In addition, since Seoul is geographically less affected by typhoons and floods, HAV is not likely to be transmitted due to cross-contamination of water supply and sewerage due to heavy seasonal rains as in other regions (17). In this respect, HAV incidence seems to be related to weather conditions, but the pattern may differ across regions. Fares (18) pointed out that some specific months are associated with a higher incidence of HAV in most countries around the world, but the exact reason for the seasonality of HAV has yet to be known. Several researchers have suggested climatic and behavioral factors such as swimming and traveling may play an important role in seasonal disease incidence (18). Moon et al. (19), based on the dataset in Korea from 2011 to 2013, reported that most HAV cases occurred from March to June. People have more outdoor activities as the weather becomes warmer during this period; thus, they are at risk of being exposed to tainted drinking water and uncooked seafood. Both are well-known risk factors for HAV infections (20). Thus, weather-related variables such as temperature and precipitation should be considered when studying HAV occurrences because weather conditions can affect people's behavior.

Various spatio-temporal analyses have been conducted to explore and understand the risk of HAV in terms of spatial and temporal structures. Gomez-Barroso et al. (21) analyzed the space-time risk of HAV using standardized incidence ratios (SIR; the ratio between actual and expected cases) and the posterior probability of the smoothed relative risk (RR; the ratio of the outcome probability for the exposed group to the probability for the unexposed group) in Spain at the municipal level from 1997 to 2007. Stoitsova et al. (22) applied the Global Moran's I index for spatial autocorrelation across Bulgaria concerning the risk of HAV infection and SIR across the nation for the whole period from 2003 to 2013 and two divided periods (2003–2008 and 2008–2013). Leal et al. (23) explored the spatio-temporal patterns of HAV outbreaks before (2008–2013) and after (2014–2017) the implementation of the national public immunization program in Pará State, a region of Brazil with severe endemic disease. Space-time scan statistics were applied to detect spatio-temporal clusters. Moreover, Leal et al. (15) investigated the association between environmental and socio-demographic data in HAV transmission in Pará State, Brazil, using various models, including generalized linear

models, multilayer perceptron (MPL) deep-learning algorithm, gradient boost, decision tree, and histogram gradient boost (HGB). To reflect the spatial variation, the longitude and latitude of each municipality were used as covariates in the model.

As discussed above, HAV is related to many factors, such as socioeconomic and epidemiological factors, weather-related factors, and spatio-temporal variations; therefore, referring to the status of diseases in other countries is not enough. For a better understanding of the HAV of the country, we should consider not only its local and national characteristics but also its social and hygienic situation.

With rapid urbanization, Korea has become cleaner. Since the first sewage treatment plant was established in Korea in 1976 (24), the number of sewage treatment plants has gradually increased, and accordingly, the number of HAV infections caused by contaminated water has decreased rapidly. As of 2020, the water supply rate is 97% nationwide, 100% in Seoul and 45.7% in Cheongyang-gun, Chungcheongnam-do. In the same year, the sewage supply rate was 94.5% nationwide, 100% in Seoul, and 5.4% in Ulleung-gun, Gyeongsangbuk-do. Although Korea's large cities have well-equipped water and sewage facilities, some areas of rural and fishing villages are in poor condition. Children with acute HAV are asymptomatic or mildly symptomatic, and antibodies (IgG anti-HAV) develop, resulting in lifelong immunity (6). According to Yoon et al. (6), IgG anti-HAV seropositivity in the Korean young adult population was low and a clean environment may lead to a decrease in the natural immune system population. Since the HAV vaccination began in 1997, the vaccination rate for 3-year-old infants in Korea exceeded 95% in 2019, according to the Korea Disease Control and Prevention Agency. People without antibodies are at an increased risk of exposure to HAV during active adolescence because they do not contract HAV as a child. When infected as an adult, the immune response is severe and symptoms such as jaundice appear; in severe cases, acute liver failure can lead to death (6). People born after 1980, when the environment was cleaner and the HAV vaccine was yet developed, have been reported to be more susceptible to HAV (25, 26). Other similar studies on the seroprevalence of HAV antibodies have been reported steadily over time (27–30).

Some efforts have been made to investigate and understand the status of HAV in Korea using statistical approaches. Research on the frequency analysis of the number of HAV infections by year, region, and age have been steadily reported (19, 31–33). Moon et al. (19) studied the epidemiological status of HAV cases in Korea between 2011 and 2013. They described significant differences in the incidence of HAV between months, regions, sexes, and age groups. In particular, they classified regions into five clusters according to the RR. Using Moran's I and scan statistics, they found clear and existing regional differences in the incidence of HAV; however, their approach

was limited to exploratory data analysis. In addition, RR does not always show correct risks (34). Seo et al. (25) studied the effect of socioeconomic status and environmental hygiene by region on the incidence of HAV based on the registered national population of Korea and national health insurance data from 2004 to 2008 using a Poisson regression model. Choi (35) conducted spatial hotspot detection of monthly incidences of HAV using spatial scan statistics and investigated the effects of socioeconomic factors using the Bayesian spatial Poisson regression model. Even though the HAV data were monthly, the proposed spatial model considered yearly incidence data and was independently applied. Thus, this study did not consider temporal and spatial variations for consecutive periods. Choi (36) analyzed HAV incidence data from 2007 to 2012 in Korea using a Bayesian spatio-temporal Poisson regression model. In Korea, HAV occurs more in summer and less in winter; therefore, seasonal factors are reflected in the model using a sine/cosine function. However, this study did not consider socioeconomic or weather factors. Baek et al. (17) conducted a time series analysis to explain the influence of factors, such as temperature and precipitation, on the incidence of HAV in Seoul, Korea. By minimizing the influence of other factors and limiting the study region to a place with a similar lifestyle, they could explain the association between weather-related factors and HAV incidence. However, it is questionable whether the same result can hold for other regions rather than Seoul in Korea. Also, further investigation of the association between HAV cases and variables other than weather is required.

In this study, we aimed to investigate the spatio-temporal variation of HAV in Korea and effects of socioeconomic and weather-related factors using a flexible spatio-temporal model. We proposed a Bayesian spatio-temporal zero-inflated Poisson regression model of weekly HAV incidence in Korea to estimate the effects of risk factors. This study is the first to develop a spatio-temporal model of HAV incidence across the entirety of Korean, with various socioeconomic factors. The advantage of this study is that the proposed model could be useful in predicting, preventing, and controlling the spread of hepatitis A.

2. Materials and methods

2.1. Description of data

We considered HAV cases as a response variable and socioeconomic, environmental, and weather-related factors as explanatory variables. The dataset for 250 nationwide districts from 2016 to 2019 was obtained from the Korea Disease Control and Prevention Agency. The weekly HAV cases at the district level (called si/gun/gu) had many zero counts (73.5%), suggesting zero-inflated statistical modeling.

We considered income, education level, and fertility rate as socioeconomic factors, because they affect the quality of life. The income variable is defined as the average monthly income per person at the district level. We calculated the high education rate as the proportion of educational attainment of a university degree or higher among the population in their 20s or an older age. The fertility rate was obtained from the actual fertility rate of women 15–49 years in Statistics Korea. We also considered the male proportion because people with active social activities are more likely to be exposed to HAV, and previous studies, including Moon et al. (19), found that the infection occurred more often among men than that among women. Since people born around the 1980s in Korea have a weak tendency toward HAV immunity (26), we considered the age group of 30–49 years. Jacobsen (20) mentioned that diverse epidemiological profiles should be treated as risk factors for HAV, and the number of registered foreigners was used for analyzes. Based on the result of Choi (35), we included the number of medical doctors per 1,000 people as a factor.

For environmental factors, we considered the water supply and sewage treatment facility rates obtained annually from Statistics Korea. Water supply plays an important role not only in terms of health and sanitation but also in industry and firefighting. Waterworks are essential for daily life, but the water system capacity varies depending on several conditions, such as the residential environment (including the housing structure) and the financial status of the local government. Therefore, the water supply rate can be used to evaluate the quality of the local living environment. According to statistics published by the Ministry of Environment of Korea, the water supply rate in Korea increased from 80.1% in 1991 to 97.3% in 2019 due to the government's continuous infrastructure expansion. However, the water supply rate varied between regions. For example, Seoul and Daegu reached 100%, but Jeju Island did not even reach 90%. The gap between the city and rural (and small town) areas is sufficiently large to be ignored. The public sewage treatment facility rate for the population is the ratio of the population beneficiaries of public sewerage services, and the closer it is to 100%, the higher the ratio of the population beneficiaries of public sewage services.

The three weather variables of interest, average temperature, total precipitation, and average humidity, were obtained from the Korea Meteorological Administration. The weather datasets were measured using two systems, an automatic weather system (AWS) and automated synoptic observing system (ASOS), at distinct weather stations (up to 510 and 103 stations, respectively). These measured values should be located in the same spatial domain as other factors; thus, we need to fit a surface to irregularly spaced weather values. Here, we combined the datasets from the two systems and subsequently predicted the weather values at each time and location of interest. To achieve this goal, we used the Kriging model based on a Gaussian

TABLE 1 Description of data set (outcome, socioeconomic factors, environmental factors, and weather-related factors).

Variable type	Variable	Time unit
Outcome	HAV cases	Weekly
Socioeconomic factor	Total income per person (1 million won)	Yearly
	High education rate (%)	2015
	Total fertility rate per woman aged 15–49 years	Yearly
	Proportion of males (%)	Yearly
	Proportion of people aged 30–49 years (%)	Yearly
	Log(number of foreigners)	Yearly
Environmental factor	Number of doctors per thousand people	Yearly
	Water supply rate (%)	Yearly
Weather-related factor	Sewage treatment facility rate (%)	Yearly
	Average temperature (°C)	Weekly
	Total precipitation (mm)	Weekly
	Average humidity (%)	Weekly

spatial random process with a Matérn covariance function at a given time (37).

We report a list of factors for this study in Table 1 and their data sources in Supplementary Table S1. Here, HAV cases and weather-related factors were measured weekly, but socioeconomic and environmental factors were collected yearly. For the socioeconomic and environmental factors, we obtained datasets from different statistics from various institutions. However, we can conveniently access all statistics through the Korean Statistical Information Service (KOSIS).

2.2. Spatial and temporal association measures

First, weekly HAV cases in a given district are now becoming time series data. Thus, we considered the autocorrelation function (ACF) and partial autocorrelation function (PACF) to examine the temporal association. The ACF corresponds to the correlation between the time series with a lagged version of itself, whereas the PACF measures the additional correlation explained by each successive lagged term. Although these two functions are slightly different, they are both measures of the association between current and past series values. For more details on the ACF and PACF, we refer to Brockwell and Davis (38).

In contrast, HAV cases in districts given at a time are areal data. In this case, we used Moran's I index measuring the strength of spatial associations among districts (39) to examine the spatial association. It is a spatial analog of the measure of association in a time series. In addition, Moran's I index explores a specific type of spatial clustering (40). The proximity matrix consists of weights that spatially connect two districts in a certain manner. Here, two districts closer to one another have more weight than those farther away. Moran's I index coupled with the proximity matrix can be interpreted as follows: a negative value corresponds to some clustering of dissimilar values, a zero value corresponds to perfect randomness, and a positive value indicates some clustering of similar values.

All the analyzes were performed using R software (version 4.1.0; <https://www.r-project.org>). We used "ape" (41) and "fields" (42) packages to compute the distances among districts and the global Moran's I index.

2.3. Statistical model

A Bayesian space-time regression model was developed to investigate the association between socioeconomic, environmental, and weather-related factors and HAV cases and to account for the space-time-dependent structures in the data. As there were many zero values in the weekly district-level HAV cases data, we used a zero-inflated Poisson (ZIP) distribution. Moreover, we considered the two-stage framework proposed by Lawson et al. (43) to overcome the spatial confounding bias problem.

In the first stage, the number of cases for district area s ($= 1, 2, \dots, S$) and weekly time index t ($= 1, 2, \dots, T$), $y_{s,t}$, follows a zero-inflated Poisson distribution as follows:

$$y_{s,t} \sim ZIP(p_{s,t}, \lambda_{s,t}),$$

where $p_{s,t}$ is the probability of structural zeros and $\lambda_{s,t}$ is the mean term of the Poisson distribution without structural zeros. The hierarchical structure of the ZIP model can be expressed as

$$\begin{aligned} y_{s,t} | z_{s,t} &\sim \text{Poisson}(\mu_{s,t} = \lambda_{s,t}(1 - z_{s,t})), \\ z_{s,t} &\sim \text{Bernoulli}(p_{s,t}), \\ \text{logit}(p_{s,t}) &= \gamma_0 + \mathbf{X}_{s,j(t)}^T \boldsymbol{\gamma} + \mathbf{W}_{s,t}^T \boldsymbol{\delta}, \end{aligned}$$

where $j(t)$ is the yearly time index for socioeconomic and environmental factors and $z_{s,t}$ is a binary variable with probability $p_{s,t}$, representing whether it is a structural zero or not. The logit($p_{s,t}$) is the linear combination of the intercept γ_0 , and the fixed socioeconomic and environmental factors $\mathbf{X}_{s,j(t)}$ and weather-related factors $\mathbf{W}_{s,t}$ with the corresponding coefficient vectors $\boldsymbol{\gamma}$ and $\boldsymbol{\delta}$, respectively. The log RR, $\log(\lambda_{s,t})$, is

modeled using fixed factors with the corresponding coefficient vectors $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$.

$$\log(\lambda_{s,t}) = \beta_0 + \mathbf{X}_{s,j(t)}^T \boldsymbol{\beta} + \mathbf{W}_{s,t}^T \boldsymbol{\alpha} + \log(N_{s,j(t)}), \quad (1)$$

where β_0 is the intercept and $N_{s,j(t)}$ is the population density as the off-set. The model of the first-stage only considers fixed factors without spatio-temporal variations.

After fitting the first-stage model using a Bayesian approach, the estimates $\hat{\mu}_{s,t}$ were computed using the posterior means. Continuous-type residuals were calculated as follows:

$$\hat{r}_{s,t} = \log(y_{s,t} + 0.1) - \log(\hat{\mu}_{s,t}),$$

where an extra value of 0.1 is added in the residual calculation because of the zero values of $y_{s,t}$.

In the second-stage, the residuals are modeled to explain the spatio-temporal variations over the first-stage covariates-only model.

$$\begin{aligned} \hat{r}_{s,t} &\sim N(ST_{s,t}, \sigma_r^2), \\ ST_{s,t} &= r_0 + u_s + v_s + \eta_t + \tau_t, \end{aligned} \quad (2)$$

where $r_0 \sim N(0, 100)$ is the intercept. The spatial random component $u_s \sim N(0, \sigma_u^2)$ explains the spatially uncorrelated structures, and v_s explains the spatially-correlated structures with conditional intrinsic auto-regressive (CIAR) distribution from Besag et al. (44), $v_s \sim \text{CIAR}(\sigma_v^2)$. The random components $\eta_t \sim N(0, \sigma_\eta^2)$ and $\tau_t \sim N(\tau_{t-1}, \sigma_\tau^2)$ explain the temporal-uncorrelated and temporal-correlated structures, respectively. After fitting the residual model with a Bayesian approach, the estimated means of the spatio-temporal structures, $\widehat{ST}_{s,t}$, were obtained. We incorporated the estimated spatio-temporal variations into the fixed covariates. The final model is expressed as follows:

$$\log(\lambda_{s,t}) = \beta_0 + \mathbf{X}_{s,j(t)}^T \boldsymbol{\beta} + \mathbf{W}_{s,t}^T \boldsymbol{\alpha} + \log(N_{s,j(t)}) + \widehat{ST}_{s,t} + \epsilon_{s,t}, \quad (3)$$

where $\epsilon_{s,t} \sim N(0, \sigma_\epsilon^2)$ is the uncorrelated space-time random component that is not explained by the estimated space-time structure. Finally, the restricted ZIP regression model was fitted using a Bayesian approach to obtain the final estimates for $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$.

For the parameter estimation, we use non-informative priors, Normal(0, 100), for the coefficient parameters β_0 , γ_0 , $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$, $\boldsymbol{\gamma}$, and $\boldsymbol{\delta}$. The standard deviations σ_u , σ_v , σ_η , and σ_τ are assigned to a uniform distribution, Uniform (0, 10). The NIMBLE package developed by de Valpine et al. (45) in the statistical software R was used to produce posterior samples. After discarding samples as a burn-in, 5,000 posterior samples with thin 50 were collected. Codes for the models can be found at https://github.com/JungsoonChoi/STmodeling_HepA.git.

TABLE 2 Descriptive statistical analysis.

Variable	Mean	SD	Q1	Q3	IQR
Total income per person	33.68	6.25	29.45	36.40	6.95
High education rate	31.12	11.19	21.70	38.12	16.42
Total fertility rate	1.14	0.27	0.95	1.30	0.35
Proportion of males	50.09	1.29	49.26	50.75	1.49
Proportion of people aged 30–49 years	27.66	5.22	23.14	31.77	8.63
Log(number of foreigners)	7.83	1.22	6.88	8.68	1.80
Number of doctors per thousand people	2.69	2.25	1.70	2.80	1.10
Water supply rate	81.05	30.39	78.97	99.90	20.93
Sewage treatment facility rate	84.69	17.25	74.18	99.50	25.32
Average temperature	12.82	9.69	4.32	21.05	16.73
Total precipitation	21.45	38.02	0.39	24.77	24.38
Average humidity	68.46	10.63	61.01	76.66	15.65

3. Results

Table 2 presents a summary of the statistical analysis of all the variables. The proportion of men had a 1.49% interquartile range (IQR = Q3-Q1), and the proportion of people aged 30–49 years had an IQR of 8.63%. Among the socioeconomic factors, these variables showed relatively smaller variations over districts and years. The average temperature and total precipitation had larger standard deviation (SD) values than that of the mean values. Water supply and sewage treatment facility rates have high mean values of 81.05 and 84.69, respectively. Supplementary Figure S1 shows the spatial variation of the average socioeconomic and environmental factors for 2016–2019. Supplementary Table S2 presents the number of HAV cases per 1,000 people divided into five groups for each socioeconomic and environmental factor. For the level of high education, the average number of HAV cases per 1,000 people 0.470 in the G1 districts and 0.615 in the G5 districts, respectively.

3.1. Spatial and temporal distributions of HAV

Figure 1A represents the weekly number of HAV cases from 2016 to 2019. Time series plots of weekly HAV cases in selected districts, Seoul-si Jongno-gu, Busan-si Sasang-gu, Daejeon-si Seo-gu, and Gyeonggi-do Bucheon-si, are shown in Figure 1B. We found that the temporal distribution of weekly HAV cases varied across districts. Figure 1C compares the

temporal variation in the cases from 2016 to 2019. In 2016, the number of HAV cases was high in the spring, but in 2019, it was high in the summer. There was little change in the number of HAV cases in 2018 compared with that of the other years.

Figures 2A–D illustrate the number of cases per 1,000 people in 2016, 2017, 2018, and 2019, respectively. Overall, the central region of Korea, the Seoul metropolitan area, had a higher number of cases than that in the other regions. The Chungcheong Province, which is close to the Seoul metropolitan region, had a greater number of cases per 1,000 people than that in the other regions.

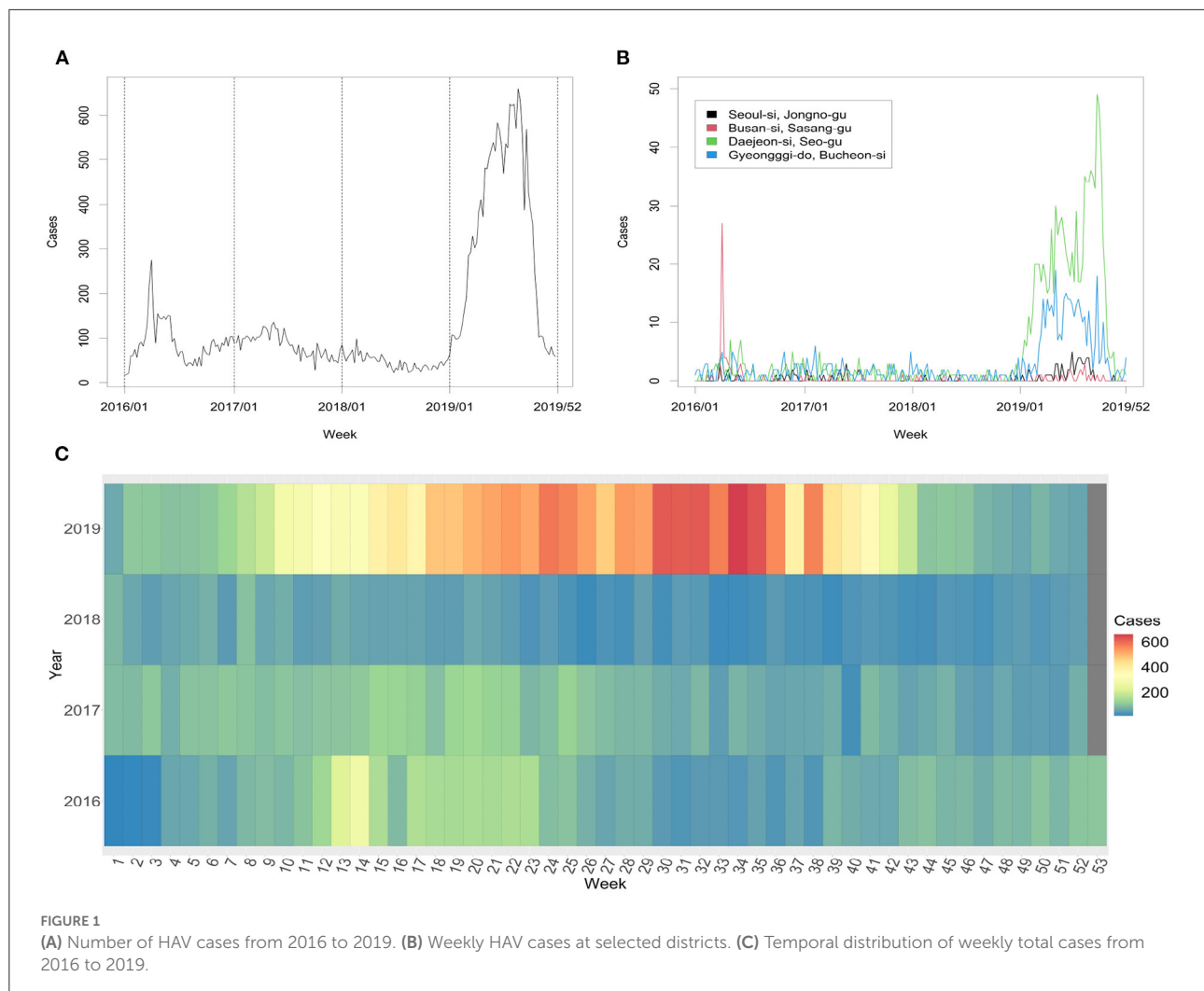
Figure 2 shows the existing connections of observations between the different districts. The more correlated the cases, the closer the districts were to the map. Table 3 presents the global Moran's I values for each year. Here, we calculated the distance-based global Moran's I value and found a positive correlation in space. Figure 3 presents the time series plot of weekly Moran's I and its *P*-values. We also found spatial dependence of HAV cases at most time points.

Figure 4 shows the ACF and PACF for all districts and each selected district. They are important tools in the exploratory data analysis of time series and, in particular, help to understand the correlation between observations at different time points. It is evident that the time series values at Daejeon-si Seo-gu and Gyeonggi-do Bucheon-si are related to their past values. Thus, considering temporal dependencies when modeling HAV cases is required. Figures 3, 4 show that we must not ignore not only temporal dependence but also spatial dependence. Thus, we included both spatial- and temporal-dependent structures in the model for a better fit.

3.2. Bayesian spatio-temporal model

We evaluated the performance of the proposed model, with several competing models as follows: Poisson and ZIP distributions were considered. Type 1 included only fixed factors without a random component. Type 2 included both fixed factors and spatio-temporal random components $ST_{s,t}$ on the right side of (2). Type 3 also includes fixed factors and spatio-temporal random components within the two-stage framework. The proposed model was a ZIP model with type 3.

Table 4 presents the comparison results of the models, with the mean absolute error (MAE), mean squared prediction Error (MSPE), and deviance information criterion (DIC=Dbar+pD) (46). In general, smaller values of the model fit criterion indicate a better model than that of the competitors. The proposed model (ZIP with type 3) had a smaller MAE and MSPE than that of the other models. Overall, the ZIP models provide slightly better performance than that of the Poisson models in terms of MAE, MSPE, and DIC. The ZIP model with type 2 has a slightly smaller DIC than that of the proposed model. However, the parameter estimation result of a ZIP



with the ST model suffered from the spatial confounding bias issue, providing many insignificant coefficients due to the spatio-temporal random components. Thus, we preferred the proposed model in terms of model fit and better interpretation.

All the variables considered in our proposed model were statistically significant at a significance level of 0.05. The parameter estimates are presented in Table 5. Total income per person, high education rate, total fertility rate, the proportion of males, and the number of foreigners were positively associated with the number of HAV cases. However, the proportion of people aged 30–49 years and the number of doctors per 1,000 people were negatively associated with the number of cases. In addition, we found a negative association between environmental factors, including water supply rate, sewage treatment facility rate, and cases. This indicates that the higher the water quality of the environment, the lower the HAV incidence rate. For weather-related factors, the coefficient of average temperature had a positive value, and the coefficient of

precipitation and humidity had a negative value with a small absolute value.

4. Discussion

We investigated the spatio-temporal distribution of the HAV incidence data in Korea from 2016 to 2019 with visualization methods and various statistical methods such as ACF, PACF, and Moran's I. The results showed that the spatial distribution of HAV incidence varied dynamically over the temporal period of interest and that the temporal distribution varied across districts. We also found that the yearly temporal distribution of HAV cases in Korea is quite different. We found a high peak and significant temporal variation in 2019. Son et al. (47) reported that the ingestion of salted clams significantly increased the risk of HAV in Korea in 2019.

Several HAV studies have been limited to frequency analysis, spatial correlation exploration using Moran's I, and comparisons

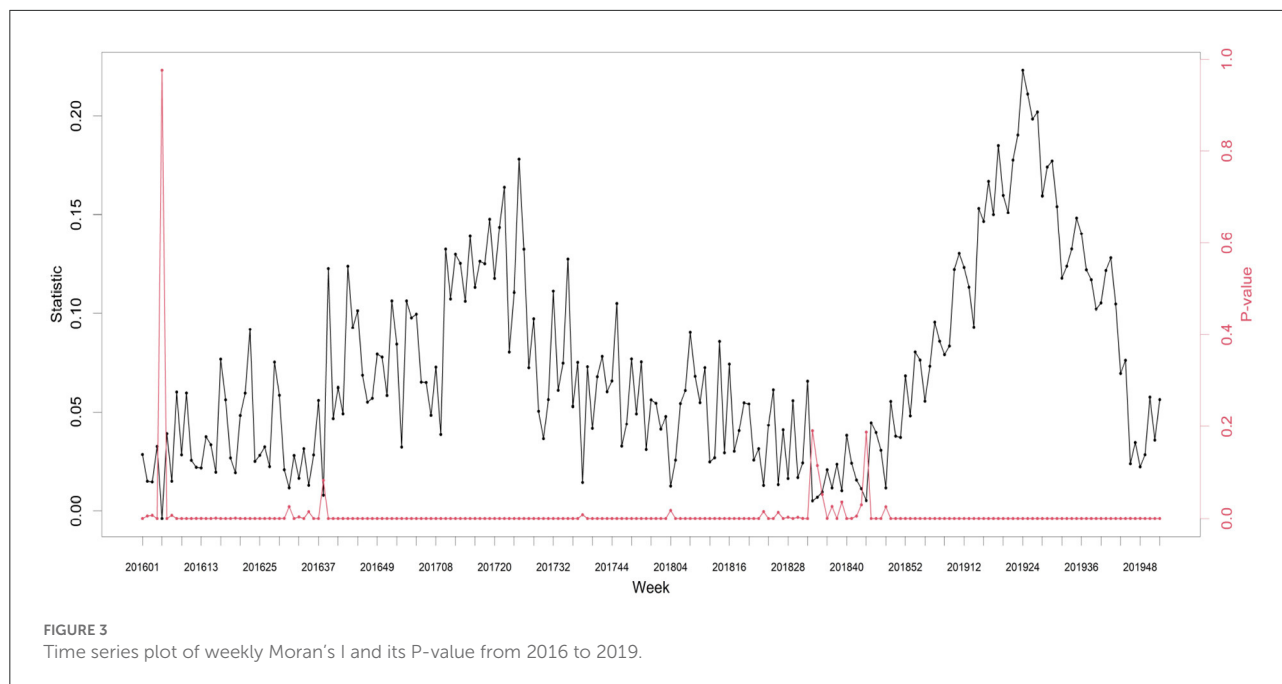


TABLE 3 Distance-based global Moran's I from 2016 to 2019.

	2016	2017	2018	2019
Moran's I	0.138	0.304	0.246	0.186

of SIR and RR. Frequency analysis is useful for finding the frequency of a variable in the entire data, but it is difficult to find a pattern when multiple variables are given conditionally. For example, our frequency analysis in [Supplementary Table S2](#) showed that HAV cases increases up to the 80% quantile of

water supply rates and sewage treatment facility rates, and then decreases in the quantile beyond that. However, such results were inconsistent with previous studies (4, 9), and it was known that these factors were related to other socioeconomic factors and had spatial variations. Therefore, frequency analysis alone has a limitation in investigating the association between environmental factors and HAV. Moran's I index is useful for investigating spatial correlation at a fixed time point, but it has a limitation in that it cannot simultaneously determine spatio-temporal correlation. SIR and RR are mainly used to identify patterns of disease occurrence. By representing the SIR



and RR values on a map, it is easy to identify the regions at high risk for disease. However, it is difficult to reflect the spatio-temporal correlation using SIR and RR simultaneously. For example, Moon et al. (19) investigated the incidence rates in Korea from 2011 to 2013 by year and age groups and represented RR in specific regions by year. They focused more on frequency analysis and could not consider spatio-temporal dependent structures simultaneously. Examining each variable separately, without considering multiple variables simultaneously, may result in a biased conclusion. In this respect, a regression model with multiple variables is better than that of frequency analysis. Furthermore, it is important to consider spatio-temporal association for epidemic data simultaneously.

Our study proposed spatio-temporal modeling of weekly incidence data in Korea using a Bayesian approach to better explain the complicated spatio-temporal dependence structures of HAV cases. The model assessed the effects of socioeconomic, environmental, and weather factors on weekly HAV cases by adjusting the spatio-temporal dynamics.

The contribution of this study is to examine the spatio-temporal distribution of HAV cases using various exploratory data analysis methods and to develop a Bayesian spatio-temporal model for considering simultaneous spatio-temporal dependent structures of the data. In terms of modeling, our contributions are as follows. Because the onset of infectious diseases has spatial and temporal correlations, a regression model that does not reflect these variations may result in a poor model fit. Considering this point, we applied a regression model with

spatio-temporal variations to HAV data in Korea. We attempted to find a model that best reflects spatio-temporal variation. We applied the two-stage framework following (43) to avoid spatial confounding bias issues; thus, we obtained a better model fit than that of the other models. Moreover, the HAV cases are counted and contain many zero values, and we consider the ZIP regression as the base model. Using the ZIP regression model coupled with two-stage and spatio-temporal structures, we demonstrated that spatio-temporal variation could not be neglected in analyzing an epidemic disease.

In the proposed spatio-temporal model, various socioeconomic, environmental, and weather-related factors were statistically significant for HAV occurrence in Korea from 2016 to 2019. The results showed that the higher the level of income and education, the more social activities, and the more frequent contact with people, the higher the possibility of exposure to HAV in Korea. It also showed that the higher the male ratio and number of foreigners residing, the higher the HAV incidence rate. Our findings agree with the studies mentioned earlier in Korea (19, 25, 35). Moreover, the number of medical doctors was negatively associated with the HAV incidence rate, as mentioned by Choi (35). We found that the proportion of people aged 30–49 years and incidence had a negative association after adjusting various socioeconomic and environmental factors. This result is somewhat inconsistent with the previous study of Yoon et al. (6). There might be confounding factors that were not considered in our study. Explanations for the present findings warrant further study on the association of the proportion of people with specific age groups and HAV.

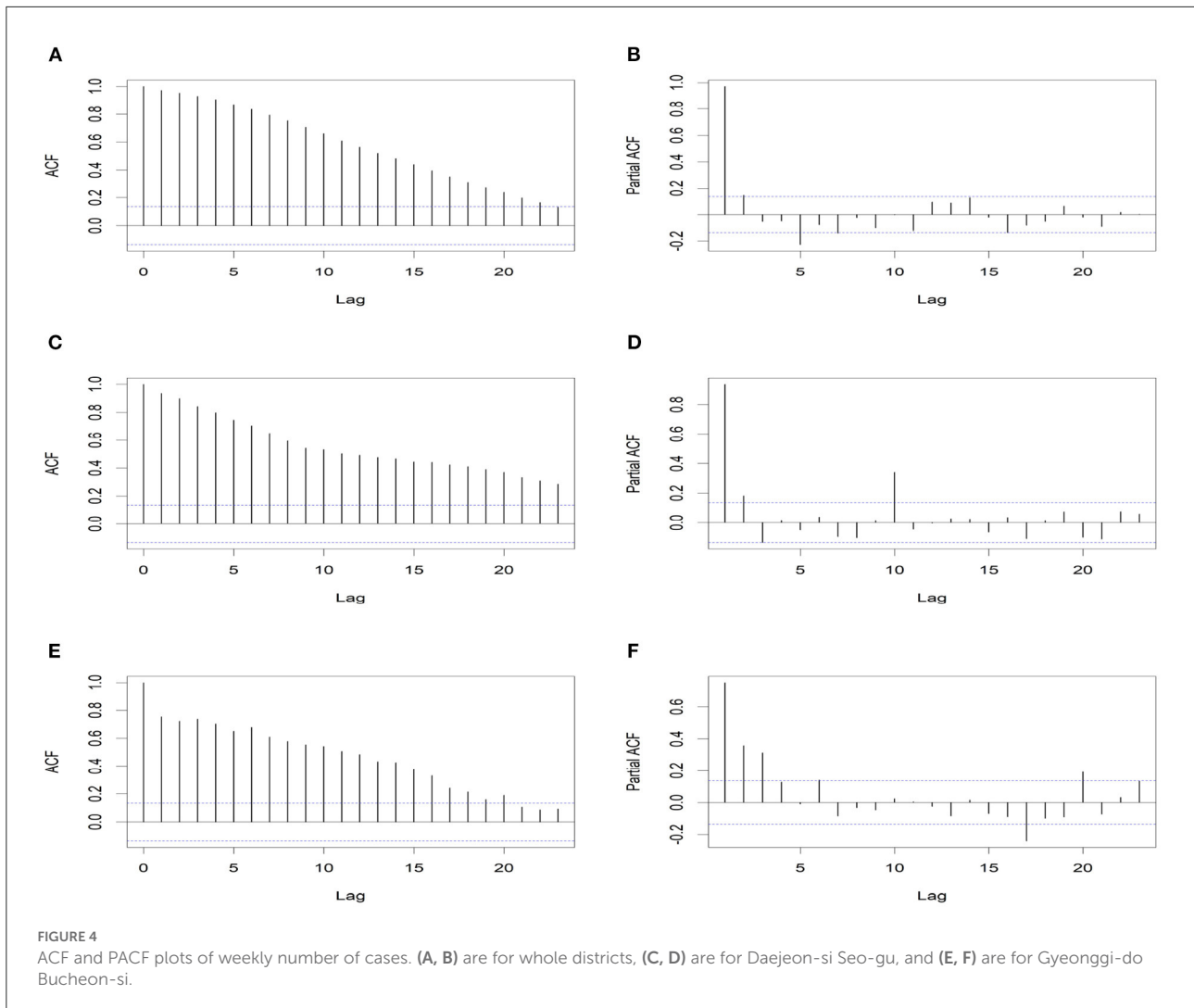


TABLE 4 Model comparison results.

Distribution	Type	MAE	MSPE	Dbar	DIC	pD
Poisson	1	0.745	2.595	128,717	128,727	11
	2	0.479	0.917	76,911	78,239	1,328
	3	0.366	0.329	66,743	73,175	6,432
ZIP	1	0.742	2.583	128,017	128,028	11
	2	0.465	0.858	72,699	73,049	350
	3	0.365	0.328	66,657	73,119	6,462

The bold values indicate the smallest value of MAE, MSPE, and DIC.

We found that the coefficients of the water supply and sewage treatment facility rates were negative, indicating that the higher the water quality and hygiene conditions, the lower the incidence rate of HAV. These results are in line with previous studies (8, 9), even though our exploratory data analysis in [Supplementary Table S2](#) looked like a positive association. Thus,

we again confirmed the importance of spatio-temporal multiple regression modeling to examine the association between factors and HAV simultaneously.

For weather-related factors, the coefficient of average temperature had a positive value, and the coefficient of precipitation and humidity had a negative value with a

TABLE 5 Parameter estimates from the proposed model.

Variable type	Variable	Posterior mean	2.5%	97.5%
Socioeconomic factors	Total income per person	0.042	0.039	0.047
	High education rate	0.025	0.019	0.030
	Total fertility rate	0.437	0.332	0.769
	Proportion of males	0.367	0.286	0.408
	Proportion of people with 30–49 years old	−0.138	−0.150	−0.127
	Log(number of foreigners)	0.517	0.493	0.577
	Number of doctors per thousand people	−0.171	−0.183	−0.161
Environmental factors	Water supply rate	−0.028	−0.034	−0.024
	Sewage treatment facility rate	−0.031	−0.034	−0.019
Weather-related factors	Average temperature	0.038	0.036	0.040
	Total precipitation	−0.001	−0.002	−0.001
	Average humidity	−0.006	−0.008	−0.003

small absolute value. [Supplementary Figure S3](#) showed the positive association between average temperature and HAV cases, although the associations between other weather factors and HAV were not clearly shown. Moreover, [Supplementary Table S3](#) indicated that the number of HAV cases was relatively large in the spring and summer seasons. A clear seasonal variation was observed in 2019. In Korea, the incidence of HAV is relatively high during spring and summer because of increased outdoor activity and ingestion of not clean food handling (19). Thus, we conclude that temperature is more associated with HAV outbreaks than precipitation or humidity in Korea.

As in the existing studies on HAV in Korea (19, 25, 35), we confirmed that there are risk factors for HAV occurrence. The distribution status of HAV varies by region and time. Additionally, differences in socioeconomic variables, such as education level, sex, number of medical doctors, and water quality, affect the number of HAV cases. Environmental and weather-related factors are also important; however, we found that the contribution of socioeconomic factors is more crucial for HAV occurrence. Therefore, we should recognize the different factors in different regions and prepare region-specific control and prevention strategies for HAV infection. Furthermore, Kang et al. (26) mentioned that a particular age group has a low antibody cultivation rate and is more vulnerable to infection. Therefore, we must consider an age-specific strategic vaccine plan.

Association between socioeconomic factors and HAV prevalence may vary from region to region because the different areas have different characteristics. For example, Jacobsen and Koopman (48) described that a higher level of education leads to a sustained decrease in the incidence of HAV, whereas there

was no statistically significant difference in education when examining HAV antibodies between sewage workers in France and the control group in Cadilhac and Roudot-Thoraval (49). While Rachiotis et al. (50) showed that people with higher education levels had higher rates of anti-HAV in stratified analysis among municipal waste collectors, Arvanitidou et al. (51) showed that the prevalence of anti-HAV was significantly higher in less educated persons. Our exploratory data analysis ([Supplementary Table S2](#)) and modeling results provided the positive association between higher education rate and HAV in Korea during 2016–2019. In Korea, people with higher education background tend to have more active social lives and more frequent contact with people so they may have highly exposure to HAV. Seo et al. (25) reported a similar result in Korea. Thus, it is important to consider regional characteristics along with weather-related factors to better understand HAV across Korea.

There was a limitation concerning the data in this study. We focused on regional aggregated data, which could lead to biased results. Thus, ascertaining the direct relationship between factors and outcomes can be limited. If we obtain individual-level HAV case data with individual-level risk factors and conduct spatio-temporal data analyzes, we can find more features that influence the HAV cases and draw clearer pictures of the infection spread problem. Thus, this is one of the future research directions the authors intend to pursue.

Data availability statement

The number of weekly HAV cases can be found in the Korea Disease Control and Prevention Agency database (<https://>

www.kdca.go.kr/index.es?sid=a3). The socioeconomic and environmental datasets can be found in the Korean Statistical Information Service (<https://kosis.kr/eng/>) and Statistics Korea (<https://kostat.go.kr/portal/eng>). The weather-related datasets were obtained from the Korea Meteorological Administration (<http://www.kma.go.kr/eng/index.jsp>).

Author contributions

JJ and JC performed the statistical analyzes. All authors contributed to the conception and design of the study, organized the database, and wrote the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was funded by the research fund of Hanyang University (HY-20200000002693) and by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (NRF-2021R1F1A1049185, NRF-2020R1F1A1A01074157, and NRF-2021R1A2C1010595).

References

- World Health Organization. *Global Hepatitis Report*. (2017). Geneva: World Health Organization (2017).
- Halliday ML, Kang LY, Zhou TK, Hu MD, Pan QC, Fu TY, et al. An epidemic of hepatitis a attributable to the ingestion of raw clams in Shanghai, China. *J. Infect. Dis.* (1991) 164:852–9. doi: 10.1093/infdis/164.5.852
- Ki M, Son H, Choi BY. Causes and countermeasures for repeated outbreaks of hepatitis A among adults in Korea. *Epidemiol Health.* (2019) 41:e2019038. doi: 10.4178/epih.e2019038
- Rakesh P, Sherin D, Sankar H, Shaji M, Subhagan S, Salila S. Investigating a community-wide outbreak of hepatitis a in India. *J Glob Infect Dis.* (2014) 6:59. doi: 10.4103/0974-777X.132040
- Franco E, Meleleo C, Serino L, Sorbara D, Zaratti L. Hepatitis A: Epidemiology and prevention in developing countries. *World J Hepatol.* (2012) 4:68. doi: 10.4254/wjh.v4.i3.68
- Yoon JG, Choi MJ, Yoon JW, Noh JY, Song JY, Cheong HJ, et al. Seroprevalence and disease burden of acute hepatitis A in adult population in South Korea. *PLoS ONE.* (2017) 12:e0186257. doi: 10.1371/journal.pone.0186257
- Tapia-Conyer R, Santos JI, Cavalcanti AM, Urdaneta E, Rivera L, Manterola A, et al. Hepatitis A in Latin America: a changing epidemiologic pattern. *Am J Trop Med Hyg.* (1999) 61:825–829. doi: 10.4269/ajtmh.1999.61.825
- Mantovani SA, Delfino BM, Martins AC, Oliart-Guzmán H, Pereira TM, Branco FL, et al. Socioeconomic inequities and hepatitis A virus infection in Western Brazilian Amazonian children: spatial distribution and associated factors. *BMC Infect Dis.* (2015) 15:1–12. doi: 10.1186/s12879-015-1164-9
- Dogru AO, David RM, Ulugtekin N, Goksel C, Seker DZ, Sözen S. GIS based spatial pattern analysis: Children with Hepatitis A in Turkey. *Environ Res.* (2017) 156:349–57. doi: 10.1016/j.envres.2017.04.001
- Copado-Villagrana ED, Anaya-Covarrubias JY, Viera-Segura O, Trujillo-Ochoa JL, Panduro A, José-Abrego A, et al. Spatial and temporal distribution of hepatitis A virus and hepatitis E virus among children with acute hepatitis in Mexico. *Viral Immunol.* (2021) 34:653–7. doi: 10.1089/vim.2021.0045

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpubh.2022.1085077/full#supplementary-material>

- Zheng B, Wen Z, Pan J, Wang W. Epidemiologic trends of hepatitis A in different age groups and regions of China from 1990 to 2018: observational population-based study. *Epidemiol Infect.* (2021) 149:1552. doi: 10.1017/S0950268821001552
- Shanmugam N, Sathyasekaran M, Rela M. Pediatric liver disease in India. *Clin Liver Dis.* (2021) 18:155. doi: 10.1002/cld.1138
- Cann K, Thomas DR, Salmon R, Wyn-Jones A, Kay D. Extreme water-related weather events and waterborne disease. *Epidemiol Infect.* (2013) 141:671–86. doi: 10.1017/S0950268812001653
- Villar LM, De Paula VS, Gaspar AMC. Seasonal variation of hepatitis A virus infection in the city of Rio de Janeiro, Brazil. *Revista do Instituto de Medicina Tropical de São Paulo.* (2002) 44:289–92. doi: 10.1590/S0036-46652002000500011
- Leal PR, de Paula RJ, Guimarães S, Kampel M. Associations between environmental and sociodemographic data and Hepatitis-A transmission in Pará State (Brazil). *GeoHealth.* (2021) 5:e2020GH000327. doi: 10.1029/2020GH000327
- Adibin, Aisnah, Indrawati, Sumriati, Ramadhan T. Increased risk of hepatitis a due to weather changes: a review. In: *IOP Conference Series: Earth and Environmental Science*. Vol. 755. IOP Publishing (2021). p. 012085. doi: 10.1088/1755-1315/755/1/012085
- Baek K, Choi J, Park JT, Kwak K. Influence of temperature and precipitation on the incidence of hepatitis A in Seoul, Republic of Korea: a time series analysis using distributed lag linear and non-linear model. *Int J Biometeorol.* (2022) 66:1725–36. doi: 10.1007/s00484-022-02313-2
- Fares A. Seasonality of hepatitis: a review update. *J Family Med Primary Care.* (2015) 4:96. doi: 10.4103/2249-4863.152263
- Moon S, Han JH, Bae GR, Cho E, Kim B. Hepatitis A in Korea from 2011 to 2013: current epidemiologic status and regional distribution. *J Korean Med Sci.* (2016) 31:67–72. doi: 10.3346/jkms.2016.31.1.67

20. Jacobsen KH. Globalization and the changing epidemiology of hepatitis A virus. *Cold Spring Harb Perspect Med.* (2018) 8:a031716. doi: 10.1101/cshperspect.a031716
21. Gomez-Barroso D, Varela C, Ramis R, Del Barrio J, Simon F. Space-time pattern of hepatitis A in Spain, 1997-2007. *Epidemiol Infect.* (2012) 140:407-16. doi: 10.1017/S0950268811000811
22. Stoitsova S, Gomez-Barroso D, Vallejo F, Ramis R, Kojouharova M, Kurchatova A. Spatial analysis of hepatitis a infection and risk factors, associated with higher hepatitis a incidence in Bulgaria: 2003-2013. *Compt Rend Acad Bulg Sci.* (2015) 68:1071-8.
23. Leal PR, de Paula RJ, Guimarães S, Kampel M. Sociodemographic and spatiotemporal profiles of hepatitis-A in the state of Pará Brazil, based on reported notified cases. *Geospat Health.* (2021) 16:981. doi: 10.4081/gh.2021.981
24. Kim I, Ryu J, Lee J. Status of construction and operation of large wastewater treatment plants in South Korea. *Water Sci Technol.* (1996) 33:11-18. doi: 10.2166/wst.1996.0292
25. Seo JY, Seo JH, Kim MH, Ki M, Park HS, Choi BY. Pattern of hepatitis a incidence according to area characteristics using national health insurance data. *J Prevent Med Public Health.* (2012) 45:164. doi: 10.3961/jpmph.2012.45.3.164
26. Kang SH, Kim MY, Baik SK. Perspectives on acute hepatitis a control in Korea. *J Korean Med Sci.* (2019) 34:e230. doi: 10.3346/jkms.2019.34.e230
27. Moon H, Noh J, Hur M, Yun Y, Lee CH, Kwon SY. High prevalence of autoantibodies in hepatitis A infection: the impact on laboratory profiles. *J Clin Pathol.* (2009) 62:786-8. doi: 10.1136/jcp.2009.064410
28. Moon HW, Cho JH, Hur M, Yun YM, Choe WH, Kwon SY, et al. Laboratory characteristics of recent hepatitis A in Korea: ongoing epidemiological shift. *World J Gastroenterol.* (2010) 16:1115. doi: 10.3748/wjg.v16.i9.1115
29. Cho HC, Paik SW, Kim YJ, Choi MS, Lee JH, Koh KC, et al. Seroprevalence of anti-HAV among patients with chronic viral liver disease. *World J Gastroenterol.* (2011) 17:236. doi: 10.3748/wjg.v17.i2.236
30. Kim H, Ryu J, Lee YK, Choi MJ, Cho A, Koo JR, et al. Seropositive rate of the anti-hepatitis A immunoglobulin G antibody in maintenance hemodialysis subjects from two hospitals in Korea. *Korean J Intern Med.* (2019) 34:1297. doi: 10.3904/kjim.2017.293
31. Kim YJ, Lee HS. Increasing incidence of hepatitis A in Korean adults. *Intervirology.* (2010) 53:10-14. doi: 10.1159/000252778
32. Lee H, Cho HK, Kim JH, Kim KH. Seroepidemiology of hepatitis A in Korea: changes over the past 30 years. *J Korean Med Sci.* (2011) 26:791-6. doi: 10.3346/jkms.2011.26.6.791
33. Yoon EL, Sinn DH, Lee HW, Kim JH. Current status and strategies for the control of viral hepatitis A in Korea. *Clin Mol Hepatol.* (2017) 23:196. doi: 10.3350/cmh.2017.0034
34. Gigerenzer G, Wegwarth O, Feufel M. Misleading communication of risk. *BMJ.* (2010) 341:c4830. doi: 10.1136/bmj.c4830
35. Choi SY. *Space and Time pattern of Hepatitis A in South Korea, 2004-2012.* [dissertation]. Seoul Korea: Hanyang University (2015).
36. Choi J. Bayesian spatiotemporal modeling in epidemiology: Hepatitis A incidence data in Korea. *Korean J Appl Stat.* (2014) 27:933-45. doi: 10.5351/KJAS.2014.27.6.933
37. Stein ML. *Interpolation of Spatial Data: Some Theory for Kriging.* New York, NY: Springer-Verlag (1999).
38. Brockwell PJ, Davis RA. *Time Series: Theory and Methods.* New York, NY: Springer-Verlag (2009).
39. Moran PA. Notes on continuous stochastic phenomena. *Biometrika.* (1950) 37:17-23. doi: 10.1093/biomet/37.1-2.17
40. Banerjee S, Carlin BP, Gelfand AE. *Hierarchical Modeling and Analysis for Spatial Data.* 2nd ed. Florida: Chapman and Hall/CRC (2014).
41. Paradis E, Schliep K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics.* (2019) 35:526-8. doi: 10.1093/bioinformatics/bty633
42. Nychka D, Furrer R, Paige J, Sain S. fields: Tools for spatial data. *R package version.* (2017).
43. Lawson AB, Choi J, Cai B, Hossain M, Kirby RS, Liu J. Bayesian 2-stage space-time mixture modeling with spatial misalignment of the exposure in small area health data. *J Agric Biol Environ Stat.* (2012) 17:417-41. doi: 10.1007/s13253-012-0100-3
44. Besag J, York J, Mollié A. Bayesian image restoration, with two applications in spatial statistics. *Ann Inst Stat Math.* (1991) 43:1-20. doi: 10.1007/BF00116466
45. de Valpine P, Turek D, Paciorek CJ, Anderson-Bergman C, Lang DT, Bodik R. Programming with models: writing statistical algorithms for general model structures with NIMBLE. *J Comput Graph Stat.* (2017) 26:403-13. doi: 10.1080/10618600.2016.1172487
46. Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A. Bayesian measures of model complexity and fit. *J R Stat Soc B.* (2002) 64:583-639. doi: 10.1111/1467-9868.00353
47. Son H, Lee M, Eun Y, Park W, Park K, Kwon S, et al. An outbreak of hepatitis A associated with salted clams in Busan, Korea. *Epidemiol Health.* (2022) 44:e2022003. doi: 10.4178/epih.e2022003
48. Jacobsen KH, Koopman JS. Declining hepatitis A seroprevalence: a global review and analysis. *Epidemiol Infect.* (2004) 132:1005-22. doi: 10.1017/S0950268804002857
49. Cadilhac P, Roudot-Thoraval F. Seroprevalence of hepatitis A virus infection among sewage workers in the Parisian area, France. *Eur J Epidemiol.* (1996) 12:237-40. doi: 10.1007/BF00145411
50. Rachiotis G, Papagiannis D, Thanasis E, Dounias G, Hadjichristodoulou C. Hepatitis A virus infection and the waste handling industry: a seroprevalence study. *Int J Environ Res Public Health.* (2012) 9:4498-503. doi: 10.3390/ijerph9124498
51. Arvanitidou M, Mamassi P, Vayona A. Epidemiological evidence for vaccinating wastewater treatment plant workers against hepatitis A and hepatitis B virus. *Eur J Epidemiol.* (2004) 19:259-62. doi: 10.1023/B:EJEP.0000020444.64546.3b