Advances in Continuous and Discrete Models

RESEARCH Open Access

Estimating real-time reproduction number for HPV infection in Xinjiang, China

Shayidan Abuduwaili[1], Lei Wang[2,3], Zhidong Teng[2,3], Abidan Ailawaer[1] and Ramziya Rifhat[2,3*]

*Correspondence:
ramziyarifhat@163.com
[2]College of Medical Engineering and Technology, Xinjiang Medical University, Urumqi, 830017, China
[3]Institute of Medical Engineering Interdisciplinary Research, Xinjiang Medical University, Urumqi, 830017, China
Full list of author information is available at the end of the article

Abstract

Objective: This study aimed to investigate the prevalence of human papillomavirus (HPV) infection in Xinjiang, estimate the real-time reproduction number (R_t) of HPV infection, evaluate the effectiveness of local prevention and control policies, and provide a theoretical basis for cervical cancer (CC) control strategies in the region. Methods: Using the R0 package, the R software was employed to estimate the dynamic changes in the real-time reproduction number (R_t) of HPV infection in Xinjiang from January 2015 to December 2019. Four methods—Maximum Likelihood Estimation (ML), Exponential Growth (EG), Sequential Bayesian (SB), and Time-Dependent (TD)—were used to calculate the R_t values. The study compared the results from these methods to determine the most reliable approach for estimating HPV transmission. Results: Among the four methods, the TD method provided the best fit for observed and predicted cases. The estimated overall R_t was 1.1381 (95% CI: 1.0762–1.1972). For the periods 2015–2017 and 2017–2019, the R_t values were 1.2827 and 1.0544, respectively. Notably, the Rt value showed a downward trend after 2017 but remained above 1, indicating continued transmission of HPV. Conclusion: Although HPV infection rates in Xinjiang appear to be decreasing, the persistent R_t values suggest that HPV transmission has not been fully controlled. This highlights the urgent need for strengthened prevention and control measures to reduce HPV-related cervical cancer risk in the region.

Keywords: Human papillomavirus; Epidemiology; Real-time basic reproduction number; Trend analysis

1 Introduction

Human papillomavirus (HPV) is ubiquitously present in nature and can induce specific infections in the skin of the human body, with certain strains persisting in the mucosal epithelium. Primarily transmitted through sexual contact, HPV infection poses a significant health concern, affecting the vagina, cervix, and anal canal, representing some of the most prevalent sexually transmitted viral diseases. Cervical cancer (CC), a malignant neoplasm, poses a substantial threat to women's well-being and may severely impact their quality of life [1]. Globally, 99.7% of cervical squamous cell carcinoma cases are associated with persistent infection of high-risk HPV [2]. Although HPV infection is highly prevalent, not all cases progress to malignant tumors. The persistent of the infection emerges

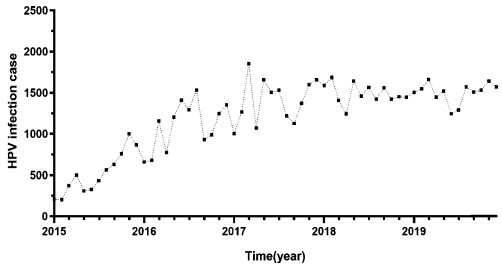

Figure 1 Number of HPV infection cases in Xinjiang from 2015–2019

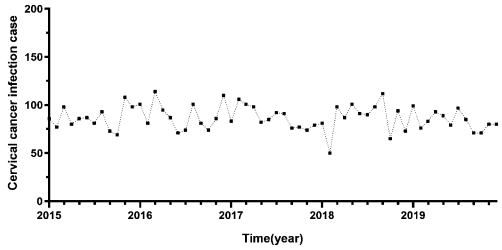

Figure 2 Number of CC cases in Xinjiang from 2015–2019

as a critical factor in the development of CC. Typically, HPV infection manifests without overt clinical symptoms and is transient in nature. Within two years, the human immune system can eliminate or eliminate 90%of HPV infection, with only a minority of individuals experiencing persistent infection, leading eventually to cervical intraepithelial lesions and, potentially, CC [3].

According to the 2022 report on global cancer mortality, CC cases worldwide rose from 604,000 in 2020 to 662,000 in 2022, with deaths increasing from 342,000 in 2020 to 349,000 in 2022 [4, 5]. Both the incidence and mortality rates of CC are escalating globally. The situation in China is particularly concerning. In 2022, China reported 15.07 million new cases and 5.57 million deaths from CC, representing 13.83% and 4.54% of global cases and deaths, respectively [6]. Xinjiang, situated in northwest of China, stands out as one of the regions with high incidence rates of CC in China [7–9]. Analysis of the epidemic sequence of HPV infection and CC cases in Xinjiang from 2015 to 2019 underscores the severity of HPV infection in the region. The incidence of infection cases rose from 2015 to 2017 and has since remained at a consistently high level, while the number of CC cases has remained elevated (Fig. 1 and 2).

The elevated incidence of CC and precancerous lesions in Xinjiang is attributed to the joint effect of various factors like regional environment, lifestyle habits, level of education, medical conditions, medical advancements, and assistance measures. Therefore, exploring the distribution of HPV infection and CC in different age groups, and regions in Xinjiang, as well as understanding the relationship and patterns with the aforementioned factors, is extremely crucial for preventing, handling with, and managing HPV infection and CC in the future. In terms of age distribution, the peak ages for HPV infection in Xinjiang women are 30–40 years and 40–50 years, as demonstrated in Fig. 3. The average age for CC to be found firstly in Xinjiang women is 51 years, and the risk is the highest for women during 40–59 years old, as explicated in Fig. 4. These results may have been influenced by

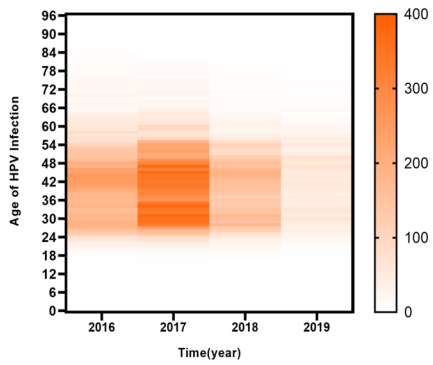

Figure 3 Age distribution of HPV infection cases in Xinjiang, 2015–2019

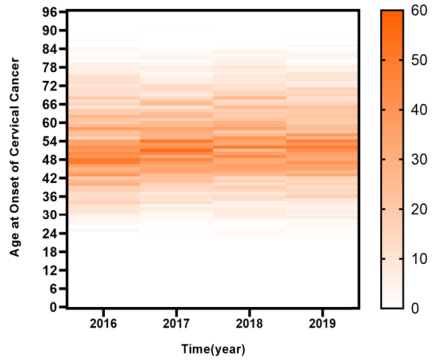

Figure 4 Age distribution of CC cases in Xinjiang, 2015–2019

several factors. First, HPV, as a sexually transmitted disease, tends to occur in individuals with sexual experience, making them more likely to be in contact with partners infected with HPV, hence increasing the risk of infection. Second, with aging and weakening of the immune system, the body becomes less resistant to persistent HPV infection, increasing the possibility of CC. Regarding regional distribution, the incidence of CC in southern Xinjiang is significantly higher than the average level of the province, especially compared to the northern and eastern regions, as explicated in Fig. 5. The higher incidence might be due to the rising HPV infection rates in the southern region and the uneven distribution of healthcare services, causing a relative lag in prevention, early detection, and intervention of CC. Additionally, this phenomenon could be influenced by local culture, socioeconomic disparities, and uneven distribution of medical resources. Therefore, further research can be developed to fully understand its underlying causes.

In summary, addressing the prevention and control of HPV infection in Xinjiang to subsequently mitigate the incidence of CC represents a pressing medical and societal challenge requiring immediate attention.

Numerous studies have delved into HPV infection using statistical models, primarily focusing on factors associated with HPV prevalence and outbreaks. These factors encompass sexual behavior [10, 11], age [12, 13], gender [14, 15], immune system status [16, 17], health knowledge and behavior, geographical environment and social economy [18, 19].

There are few studies focusing on dynamic models of HPV transmission, with the majority being theoretical in nature. Considering the epidemiological characteristics of HPV infection, most proposed dynamic models are centered on global dynamic analyses to obtain strategies for HPV control [20–22]. The basic reproduction number (RPN)(R_0) serves

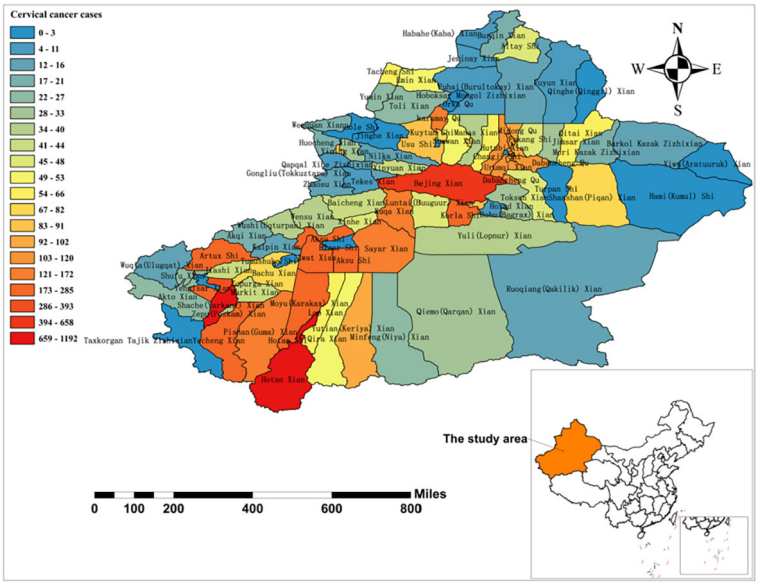

Figure 5 Regional distribution of CC incidence in Xinjiang from 2015 to 2019

as a key epidemiology indicator for assessing virus spread, representing the average new infection number induced by infectious individuals in a wholly susceptible population. When $R_0 < 1$, the infection will dwindle and eventually disappear from the population. Conversely, if $R_0 \geq 1$, the disease will persist and propagate within the population, potentially leading to an enduring epidemic (with infection rates continuing to rise). Ramziya Rifhat et al. [23] developed a dynamic model of HPV infection and secondary CC transmission regarding the transmission characteristics specific to Xinjiang, which incorporates a two-stage recovery process for HPV infection and precancerous lesions. Parameter values were optimized using the Markov Chain Monte Carlo (MCMC) algorithm, leveraging data from CC cases in Xinjiang Province, China. The estimated basic RPN was $1.1490(95\%CI : 0.6778 \sim 1.9084)$, underscoring the significant impact of HPV infection on CC incidence in Xinjiang. Furthermore, the study delves into the sensitivity of key parameters to the basic RPN through numerical simulation. Ultimately, practical control strategies were proposed to mitigate the further spread of HPV infection and CC within the region. Gao et al. [24] constructed a bisexual model to evaluate HPV vaccine allocation strategies, focusing on Guangxi Province, China. They investigated the optimal distribution of vaccines between genders given a fixed vaccine dose. After numerical analysis, it was found that first vaccinating the vaccine to the gender with a low input rate, followed by distributing the remaining vaccines among other genders, can minimize the basic RPN $R_{min} = 0.9987$, ensuring it remains below the critical threshold of 1.

The real-time RPN (R_t) quantifies the anticipated number of secondary cases stemming from primary infections at time t. A value of $R_t > 1$ indicates ongoing disease transmission, whereas a decline in R_t indicates a gradual slowdown in disease spread. Monitoring changes in R_t estimates over time facilitates a precise assessment of the scale of epidemic progression, delineating peak disease incidence and inflection points more clearly. This dynamic metric furnishes a specific basis for evaluating the effect of prevention and control measures [25]. While basic RPN denotes an average value, real-time RPN offers

varying estimates at different time points, enhancing our ability to discern shifts in the rate of infectious disease transmission. Some researchers employed dynamic basic RPN methodologies to assess the epidemic trend of COVID-19 and H1N1 influenza and the effect of prevention and control measures [26–28]. Utilizing epidemic data of COVID-19 from Beijing, Shanghai, Guangzhou, and Shenzhen, Han Ke et al. [28] estimated the real-time RPN's fluctuations at different times by using mature models to evaluate the trend of COVID-19 infectivity over time within the context of prevailing prevention and control measures. Thomas Obadia et al. [29] utilized R statistical software to establish a tailored package, employing exponential growth (EG), maximum likelihood (ML), sequential Bayesian (SB), and time-dependent (TD) methods to estimate the real-time RPN of H1N1 influenza in 2009. It aimed to evaluate the influenza epidemic trend and the efficiency of prevailing prevention and control measures. Building upon the package developed by Thomas Obadia et al. within the R statistical software, Wang Ying et al. [30] estimated the real-time RPN of daily reported COVID-19 confirmed cases in Hubei Province from January 17 to February 8, 2020. Employing the EG, ML, SB, and TD, they compared the fitting effects of the four methods against actual observations, with the EG method demonstrating the best fitting effect. The estimated R_t value is 3.49, with a reduced R_t value of 2.95 during the implementation of closure control measures, indicating the effective reduction of virus transmission rates. Roya Nikbakht et al. [31] used ML, EG, TD, and AR to estimate the R_t value using Canadian influenza cumulative case data. Fitting results revealed that the TD method exhibited the smallest mean square error (MSE). The TD method estimated the real-time RPN of 2018 data to be 1.52 ($95\%CI : 1.11 \sim 1.94$), highlighting the persistence of the epidemic and the necessity for continued measures to control, mitigate, and prevent the spread of the epidemic.

Analyzing pertinent data regarding HPV infection and estimating the real-time RPN (R_t) holds significant importance for understanding the transmission ability of the virus and assessing the efficiency of subsequent prevention and control measures. Currently, few studies have been conducted to analyze the HPV virus infection trends and prevention and control in Xinjiang using real-time RPN. This research investigated an innovative methodology for estimating the dynamic change in R_t, the basic RPN, at different time points to evaluate the transmission capacity of the virus and the effect of subsequent prevention and control measures.

2 Materials and methods

We review and outline the application of techniques for estimating both the serial interval distribution and RPNs during epidemics. Additionally, we introduce tools for assessing the impact of assumptions on estimation sensitivity.

2.1 Data sources

As a comprehensive large-scale hospital, Xinjiang Medical University affiliated hospitals possess advanced medical equipment and technological capabilities to provide support for the diagnosis and treatment of diseases such as HPV infection and cervical cancer. With a wide range of patient sources, including patients from different cities and counties in Xinjiang, it can better represent the demographic characteristics and disease spectrum of the region. Furthermore, these hospitals have the capability for remote consultations, enabling data sharing and collaboration with hospitals in other regions of Xinjiang to further

enhance the overall understanding of HPV infection and cervical cancer incidence rates. Therefore, this study aims to collect HPV infection and cervical cancer case data from Xinjiang Medical University-affiliated hospitals to reflect the overall situation of HPV infection and cervical cancer incidence rates in Xinjiang.

We selected data from recent years to analyze the trends in HPV infection and cervical cancer incidence in Xinjiang. This decision was informed by several factors. First, we sought to ensure the contemporaneity of the data by excluding earlier periods. Second, the impact of the HPV screening program initiated in 2009 requires a certain amount of time to manifest in the overall infection trends, so we avoided overly early data. Moreover, the COVID-19 pandemic in 2019 shifted public attention away from HPV infection, resulting in decreased awareness. Consequently, this study does not consider data on HPV infection and cervical cancer incidence beyond 2019.

2.2 Defining a generation time (GT) distribution

The GT refers to the duration between infection in a primary and a secondary case. Typically, its distribution is deduced from the temporal lag between all pairs of infectors and infectees [32]. Given its inherent unobservability, it is often approximated by the serial interval distribution, which characterizes the elapsed time between the onset of symptoms. Our software features a'generation.time()' function tailored to depict a discretized GT distribution. Discretization is conducted over defined intervals such as $[0, 0.5)$, $[0.5, 1.5)$, and $[1.5, 2.5)$, with users able to select their preferred time unit (hour, day, week...). Our software supports multiple descriptions, including "empirical", which necessitates a comprehensive distribution specification, and parametric distributions such as "gamma", "lognormal", or "weibull". In the latter scenario, users are required to provide the mean and standard deviation in their desired time units. Additionally, we offer a function'est.GT()' for estimating the serial interval distribution through exponential growth and maximum likelihood, utilizing a sample of observed time intervals between symptom onsets in primary and secondary cases.

2.3 Exponential growth method (EG)

As delineated by Wallinga and Lipsitch [33], the EG rate in the initial outbreak phase is closely tied to the initial reproduction ratio. This growth rate e quantifies the per capita change in new case numbers over time. Due to the discrete nature of incidence data, Poisson regression is recommended for estimating this parameter [29, 34], rather than linear regression on the logged incidence. The RPN is computed as

$$R = \frac{1}{M(-e)},$$

where M denotes the moment-generating function of the discretized GT distribution. Selecting a period on the epidemic curve characterized by exponential growth is imperative. The deviance-based R-squared statistic is employed to aid in this selection process. No assumption is made regarding population mixing.

2.4 Maximum likelihood estimation (ML)

This model, introduced by White and Pagano [35], operates under the assumption that the secondary case number resulting from an index case follows a Poisson distributed with an

expected value R. By observing $(N_0, N_1, \ldots, N_T)$ incident cases across consecutive time intervals, and with knowledge of the GT distribution ω, R is estimated by maximizing the log-likelihood

$$LL(R) = \Sigma_{t=1}^{T} \log\left(\frac{e^{-\mu_t}\mu_t^{N_t}}{N_t!}\right),$$

where

$$\mu_t = R\Sigma_{m=1}^{t} N_{t-m}\omega_m.$$

Once more, it is imperative to compute the likelihood over a timeframe characterized by exponential growth, with the possibility of employing the deviance R-squared measure to identify the optimal period. Notably, no assumptions are made regarding population mixing within the community.

2.5 Sequential Bayesian method (SB)

While introduced by its authors as "real-time Bayesian", this method facilitates more precise sequential estimation of the initial RPN. It takes an approximation to SIR model, wherein the incidence at time $m+1$ noted as N_{m+1}, is approximately as Poisson distributed with a mean of $N_m e^{\gamma(R-1)}$ [36], where $\frac{1}{\gamma}$ represents the reciprocal of the infectious period average duration. The proposed algorithm, outlined within a Bayesian framework, commences with a non-informative prior distribution on the RPN R. Subsequently, this distribution is updated as fresh data becomes available:

$$P(R|N_0, N_1, \ldots, N_{m+1}) = \frac{P(N_{m+1}|R, N_0, N_1, \ldots, N_m)P(R|N_0, N_1, \ldots, N_m)}{P(N_0, N_1, \ldots, N_{m+1})}$$

In essence, the prior distribution for R employed on each subsequent day corresponds to the posterior distribution from the preceding day. At each time point, the posterior mode can be calculated, the highest probability density interval can be determined. Similar to previous iterations, this methodology mandates that the epidemic is experiencing exponential growth, neglecting susceptible depletion. It implicitly assumes an exponential distribution for the GT and posits random mixing within the population.

2.6 Time-dependent method (TD)

The methodology proposed by Wallinga and Teunis [37] for computing RPNs is TD. It involves averaging all the compatible transmission networks observed. Specifically, the probability P_{ij}, denoting the transmission from case j with onset at time t_i to case i with onset at time t_j, is determined as follows:

$$P_{ij} = \frac{N_i w(t_i - t_j)}{\sum_{i \neq k} N_i w(t_i - t_k)}.$$

The effective RPN for case j is thus denoted as $R_j = \sum_i p_{ij}$ and is averaged as $R_t = \frac{1}{N_t}\sum_{t_j=t} R_j$ over all cases with the same date of onset. CI for R_t is derived through simulation. Correction for real-time estimation, accounting for unobserved secondary cases,

are feasible [38]. Additionally, it is viable to accommodate imported cases throughout the epidemic.

In this study, statistical analysis was performed using R 4.1.3. The real-time RPN R_t was estimated utilizing the'estimate.R()' function in the R0 software package. The estimated R_t values for the SB and TD methods were derived from averaging across all time points in the infection dataset. Given that the EG and ML methods necessitate an exponential growth phase, we additionally performed sensitivity analysis on the exponential growth phase through the'Sensitivity.analysis()' function. This approach aids in identifying the optimal time window for estimating the basic RPN, thereby enhancing the study's credibility.

3 Results

3.1 Estimating real-time RPNs

The dataset for this study comprises monthly confirmed HPV infections reported by hospitals affiliated with Xinjiang Medical University. Data collection spanned from January 2015 to December 2019. In line with the incubation period of HPV infection, a Gamma distribution with an average of 2.6 months and a standard deviation of 1 month was employed to model the time distribution of HPV infection. To construct the HPV epidemic curve, we employed ML, EG, SB, and TD to estimate R_t (Table 1).

The R_t values estimated using case data from January 2015 to December 2019 are as follows: The estimated R_t for the ML method was $1.0574(95\%CI : 1.0465 \sim 1.0684)$, for the EG method was $1.0534(95\%CI : 1.0520 \sim 1.0549)$, for the SB method was $1.1625(95\%CI : 1.1120 \sim 1.2115)$, and for the TD method was $1.1381(95\%CI : 1.0762 \sim 1.1972)$. Surprisingly, despite utilizing the same data, the estimated range exhibited a considerable variation of up to 10% (from 1.0534 to 1.1625). In addition, the confidence intervals did not consistently overlap (Fig. 6). However, the fitting of each model to the data remained relatively similar across all cases, except for the SB method, which displayed poorer fitting (Fig. 7).

No substantial disparity existed among the real-time RPNs, as estimated by various methods. However, the R_t estimates derived from each method surpass the critical threshold of 1, which indicates that HPV transmission in Xinjiang remains uncontained and continues to spread. Therefore, it remains imperative to prioritize HPV virus transmission in

Table 1 Real-time RPN and 95%CI estimated by ML, EG, SB, and TD method

Method	Initial Estimate R_t [95%CI]	Optimal time range estimate R_t [95%CI]
ML (optimal time window: 12:20)	1.0574 [1.0465 , 1.0684]	1.3300 [1.2577 , 1.4079]
EG (optimal time window: 1:20)	1.0534 [1.0520 , 1.0549]	1.3134 [1.3016 , 1.3254]
SB	1.1625 [1.1120 , 1.2115]	1.3205 [1.2179 , 1.4200]
TD	1.1381 [1.0762 , 1.1972]	1.2916 [1.2208 , 1.3633]

All estimates were generated using the first 60 months of data (default column)or the optimal fitting time window ("optimal" column). For the SB method, the optimal reported estimate was obtained on month 20, as this date best fits the end of the exponential growth period. For the TD method, the optimal estimates were averaged over the first 27 months.

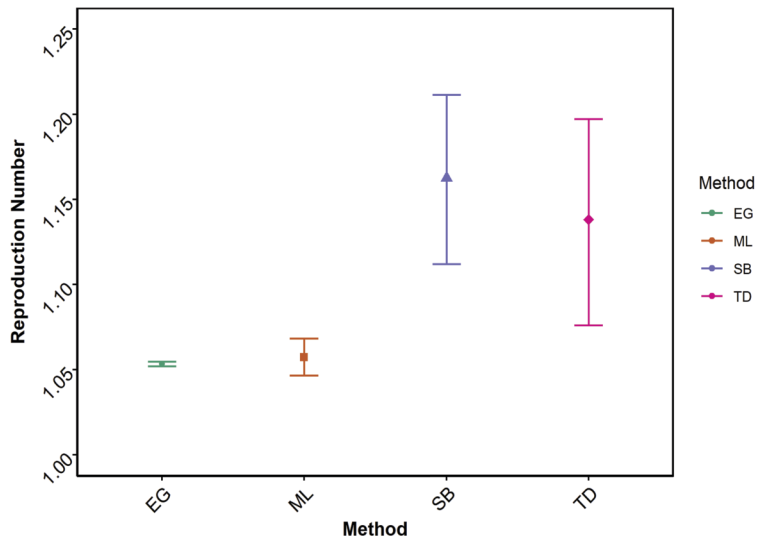

Figure 6 R_t value and its 95% CI of HPV infection in Xinjiang from 2015 to 2019 estimated by EG, ML, SB, and TD methods

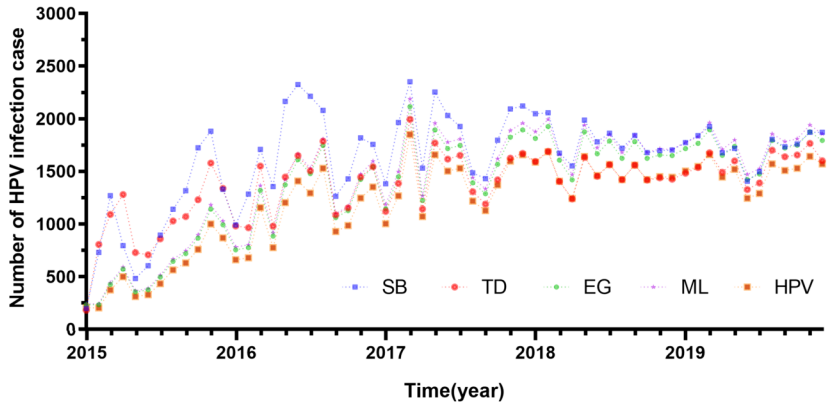

Figure 7 Fitting effect diagram of the four methods ML, EG, SB, and TD and actual observations

Xinjiang and bolster the implementation of prevention and control measures to curb the incidence of CC among women in Xinjiang and advance towards the goal of eliminating CC promptly.

3.2 Sensitivity analysis

The EG and ML methods necessitate the user to designate the time frame during which exponential growth occurs. By default, this period spans from the first case to the date of maximum incidence. Nevertheless, a more optimal selection can be made by leveraging the deviance R-squared statistic across a spectrum of potential time periods. The highest R-squared value indicates the period in which the analytical model best fits the data, and this interval is chosen to provide estimates. To identify this time span, the'Sensitivity.analysis()' function systematically computes the deviance R-squared statistic across user-defined time intervals. A plot illustrating the highest R-squared value over progres-

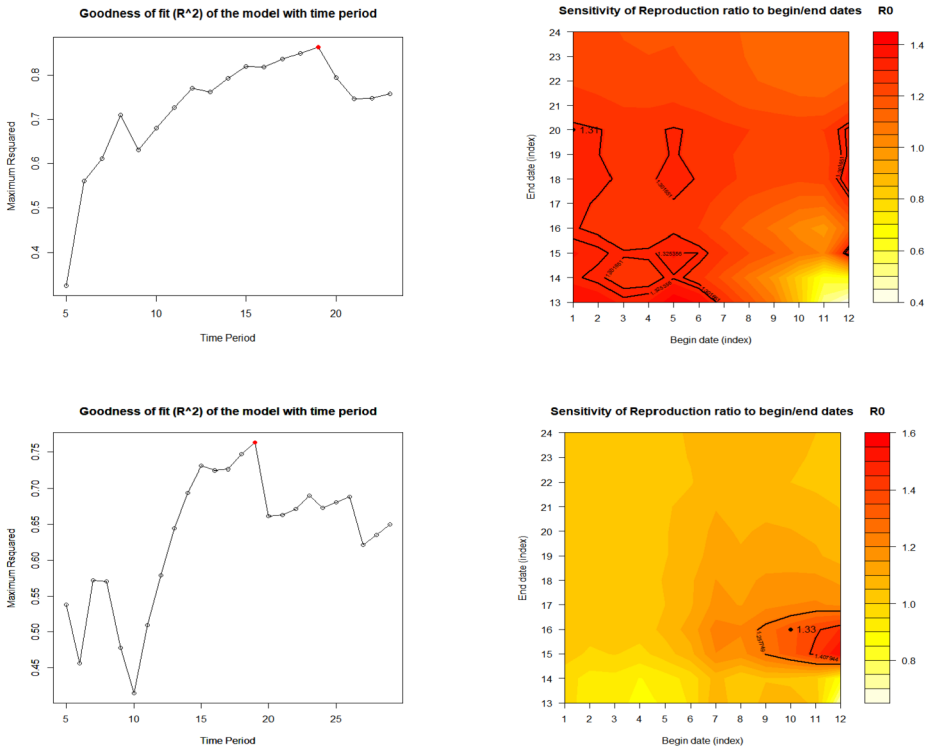

Figure 8 Sensitivity of the reproduction ratio to the choice of the time period for estimation. (a), (c) The maximum deviance R-squared statistic across varying time periods, with the red dot indicating the optimal value. (b), (d) Estimates of the reproduction ratio for various start and end dates, with the dot denoting the best fit and solid black lines delineating the corresponding 95%CI. Estimates falling within the 95%CI of the best-fit value can be obtained by employing a diverse range of start and end dates, as indicated by values between the solid black lines

sively longer time periods can be generated (Fig. 8 (a)(c)), while the corresponding estimates are displayed based on the selected time window (Fig. 8 (b)(d)).

During the exponential growth period analyzed in this study (January 2015–March 2017), the optimal exponential growth interval was determined using the deviation R-squared statistic. The analysis indicated that the most appropriate section of the epidemic curve for exponential growth in the EG and ML methods spanned 19 months. Specifically, from the first month to the 20th month, the estimated RPN for the EG method was $1.3134(95\%CI : 1.3016 \sim 1.3254)$ (Fig. 8 (b)). The RPN estimate for the ML method was $1.3300(95\%CI : 1.2577 \sim 1.4079)$ (Fig. 8 (d)).

Across a wide range of time windows, the RPN estimates remained within the 95%CI of the optimal fit, implying robustness to variations in the exponential growth period. The estimates derived from the best-fitting time window for each method are reported in Table 1. Notably, employing the "best fitting" time period for EG and ML methods resulted in reduced variability between estimates.

According to the research by Thomas Obadia et al. [29], the SB method is deemed more suitable for estimating the real-time RPN during the latter portion of the initial exponential growth phase. Analysis of the data in this study revealed a brief, sharp decline following an increase in cases from January 2015 to August 2016. Therefore, for the SB method, the period spanning from the first month to the 20th month was identified as

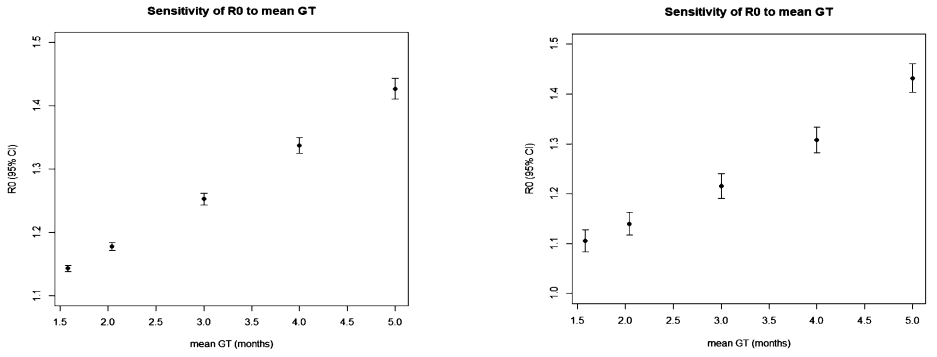

Figure 9 Sensitivity of the RPN to the choice of the GT distribution. Reproduction ratio estimates were computed using various mean GTs. (a) The variation in the estimated real-time RPN using the EG method across different generation times. (b) The variation in the estimated real-time RPN using the ML method across different generation times. CIs are shown as vertical bars

the most appropriate time period for estimating the real-time RPN, yielding an estimate of $1.3205(95\%CI : 1.2179 \sim 1.4200)$. The TD method is best applied during the period when the infection rate peaks. Therefore, the TD methodology was utilized to estimate the real-time RPN using incidence data from January 2015 to March 2017, generating an estimated value of $1.2916(95\%CI : 1.2208 \sim 1.3633)$ (Table 1). Since 2014, the HPV detection pilot project has been officially initiated in Xinjiang. Secondly, domestic vaccines have also been available on the market since 2016. In addition, the widespread use of mobile phone applications such as TikTok has heightened the influence of digital networks on individuals, leading to an increased awareness of HPV among women in Xinjiang. These factors collectively contribute to enhancing women's attention to HPV in Xinjiang, which further promotes the enthusiasm of women in Xinjiang for HPV detection and increases the detection rate of HPV.

Another pertinent consideration is how estimates vary based on the determined GT distribution. We varied the GT distribution for the EG and ML methodologies (Fig. 9 (a)(b)): as anticipated, estimates elevated with the mean GT [**s41??**]. Employing the same epidemic curve, reported reproduction ratios ranged from 1.1 to 1.5 as the mean GT varied from 1.5 to 5 months. <ref:??>

3.3 The fitting of the four methods with the actual observations

In Fig. 7, the real-time RPNs estimated by the EG, ML, SB, and TD methods were utilized to forecast the number of HPV infection cases, and the predicted values were fitted with the actual observations. TD method exhibited the smallest deviation and displayed superior fitting effectiveness.

Figure 10 illustrates that the R_t value estimated by the TD method, exhibiting optimal fitting, displays a discernible downward trend over the years, with a gradual convergence toward 1 since 2018. Observing the HPV infection epidemic curve in Xinjiang from 2015 to 2019 (Fig. 1), it is evident that the number of HPV infection cases continued to rise from 2015, peaking in March 2017, before stabilizing into a relatively flat trend thereafter. The estimated R_t value was 1.2827 during 2015–2017, decreasing to 1.0544 during 2017–2019, indicative of a significant reduction in the estimated RPN. This decline can be attributed to several factors: heightened awareness of HPV infection due to increased internet penetra-

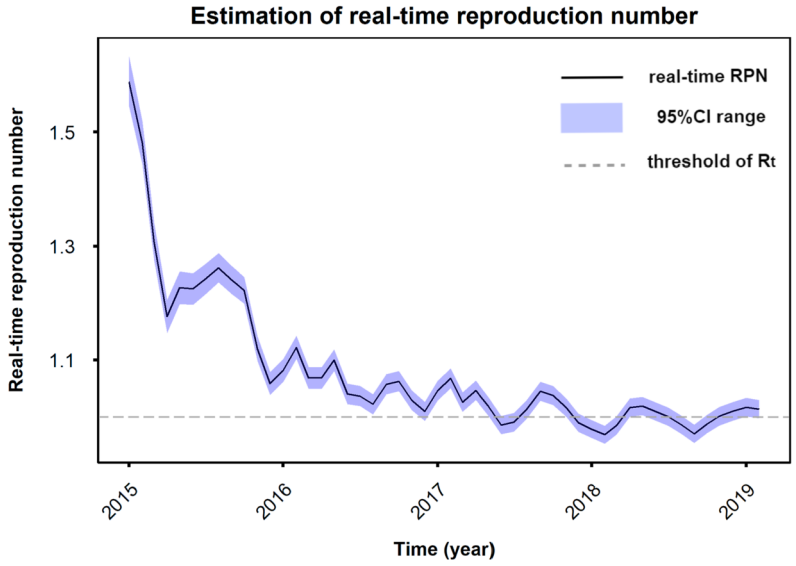

Figure 10 Estimating the trend of the real-time RPN R_t over time using the TD methodology. (The light blue shadow is the 95% CI range, the blue solid line is the real-time RPN, the gray solid line is the average of the real-time RPN from 2015 to 2019, and the black dotted line is the threshold 1)

tion, improvements in women's education leading to enhanced cultural standards, and reduced instances of early marriage, early sexual activity, and early childbirth among women in Xinjiang. Additionally, HPV vaccination has also played a significant role in preventing infection. Since the end of 2016, China has successively launched HPV domestic vaccines and implemented vaccine policies, gradually expanding vaccination coverage [39].

If the R_t estimation value remained above 1, it signifies the necessity for the implementation of effective prevention and control strategies to curb the spread of HPV infection and mitigate the incidence of CC in Xinjiang.

4 Conclusion

In this study, data encompassing the monthly count of confirmed HPV infections reported by Xinjiang Medical University-affiliated hospitals from January 2015 to December 2019 was collected. EG, ML, SB, and TD were employed to estimate the real-time RPN according to the epidemic curve. Subsequently, these methods were compared to determine the one yielding the highest accuracy and optimal fitting. The real-time RPN was estimated utilizing the function'estimate.R()' within the R0 package of the statistical software R4.1.3 to evaluate the transmission rate of HPV, providing a scientific basis for disease prevention and control.

R_t serves as a critical parameter in assessing the prevalence of infectious diseases. The higher the R_t value, the faster the spread of such infectious diseases. Through the fitting of observed and predicted case counts using the four methods, it was determined that the TD method exhibited the optimal fitting effect, yielding an estimated R_t value of $1.1381(95\%CI : 1.0762 \sim 1.1972)$. Obviously, the R_t value for HPV infection remains above 1, signaling the persistence of a severe trend in Xinjiang. Consequently, intensified control measures are imperative to curtail further spread of the infection within the region.

In this study, the R_t values for HPV infection in Xinjiang were calculated using different methods, facilitating an assessment of the transmission potential of the HPV virus. Comparisons were made between R_t values before and after the implementation of control measures such as sex education, regular screening, and vaccination. Initially, R_t values suitable for the TD method were estimated during the early epidemic phase. Subsequently, following the implementation of these control measures, a significant decrease in R_t values was observed compared to earlier periods (with estimated values of 1.2827 for 2015–2017 and 1.0544 for 2017–2019), indicating the efficacy of these control measures in reducing the transmission rate of HPV virus. This study provides vital parameters for further analysis of HPV infection, including the evaluation of control measure implementation effectiveness and the prediction of future infection trends, and also provides a scientific basis for adjusting control measures.

With the rapid advancement of the national economy and the enhancement of national policies, Xinjiang has witnessed the formulation of numerous policies aimed at bolstering various facets of its society, including its economy, security, education, medical care, and other aspects. By strengthening the publicity of HPV infection-related knowledge, women in Xinjiang have paid more attention to HPV infection and related diseases. Through tailored sex education suitable for age and education level, Xinjiang women have exhibited increased vigilance towards inappropriate sexual behaviors. By vigorously promoting the role of HPV vaccination, the HPV vaccination rate of women in Xinjiang has been elevated, with vaccination initiation occurring at increasingly younger ages. Consequently, it is anticipated that the HPV infection epidemic in Xinjiang will be effectively controlled in the foreseeable future, bringing the complete elimination of CC within reach.

Author contributions
SA: numerical analysis and writing. LW: Modelling, review and editing. ZT: Data collection and numerical analysis. AA: Data collectiong. RR: Numerical simulation, mathematical analysis, writing, review and editing. All authors read and approved the final manuscript.

Funding
This research is supported by the National Natural Science Foundation of China(Grant Nos. 12101529).

Data availability
Data sharing is not applicable because the current study fails to generate or analyze datasets.

Declarations

Competing interests
The authors declare that they have no competing interests.

Author details
[1]School of Public Health, Xinjiang Medical University, Urumqi, 830017, China. [2]College of Medical Engineering and Technology, Xinjiang Medical University, Urumqi, 830017, China. [3]Institute of Medical Engineering Interdisciplinary Research, Xinjiang Medical University, Urumqi, 830017, China.

Received: 18 October 2024 Accepted: 14 January 2025 Published online: 04 February 2025

References

1. Zhang, X.: Relationship between HPV and cervical cancer. Gems. Health **25**, 7 (2019). (in Chinese)
2. Dong, A.R.: Study on mechanism of cervical cancer induced by HPV. China Mod. Dr. **50**(34), 33–34 (2012). CNKI:SUN:ZDYS.0.2012-34-019. (in Chinese)
3. Bai, H., Cai, J.Y.: Pathological study on the relationship between HPV infectionin female lower reproductive tract and cervical intraepithelial neoplasia and cervical cancer. Mod. Oncol. **25**(4), 617–619 (2017). https://doi.org/10.3969/j.issn.1672-4992.2017.04.031. (in Chinese)
4. Long, C.H., Yuan, W.G.: Global and domestic cervical cancer epidemic status and prevention strategy. Chin. Clin. Oncol. **28**(1), 90–93 (2023). https://doi.org/10.3969/j.issn.1009-0460.2023.01.013. (in Chinese)
5. International Agency for Research on Cancer: https://gco.iarc.who.int/

6. Zheng, R.S., Chen, R., Han, B., et al.: Cancer incidence and mortality in China, 2022. J. Natl. Cancer Cent. **46**(3), 221–231 (2024). https://doi.org/10.3760/cma.j.cn112152-20240119-00035
7. Wang, X.L., Li, F., Li, H.N., et al.: Analysis of cervical cancer screening results ofwomen in poor areas of Xinjiang from 2013 to 2016. Chin. J. Women Child Health Res. **33**(3), 62–66 (2022). https://doi.org/10.3969/j.issn.1673-5293.2022.03.010. (in Chinese)
8. Suzuk, L., Peng, Y.H., Zhou, K., et al.: Analysis of the incidence trend of cervical cancer in different ethnic groups in Xinjiang. J. Xinjiang Med. Univ. **2006**(07), 569–571 (2006). (in Chinese)
9. Zhang, J., Gao, C.: Analysis of current situation of female HPV infection in Xinjiang. Lab. Med. Clin. **18**(14), 2093–2096 (2021). https://doi.org/10.3969/j.issn.1672-9455.2021.14.029. (in Chinese)
10. Vaccarella, S., Franceschi, S., Herrero, R., et al.: Sexual behavior, condom use, and human papillomavirus: pooled analysis of the IARC human papillomavirus prevalence surveys. Cancer Epidemiol. Biomark. Prev. **15**(2), 326–333 (2006). https://doi.org/10.1158/1055-9965.EPI-05-0577
11. Castle, P.E., Shields, T., Kirnbauer, R., et al.: Sexual behavior, human papillomavirus type16 (HPV 16) infection, and HPV 16 seropositivity. Sex. Transm. Dis. **29**(3), 182–187 (2002). https://doi.org/10.1097/00007435-200203000-00009
12. Giannella, L., Giorgi Rossi, P., Delli Carpini, G., et al.: Age-related distribution of uncommon HPV genotypes in cervical intraepithelial neoplasia grade 3. J. Gynecol. Oncol. **161**(3), 741–747 (2021). https://doi.org/10.1016/j.ygyno.2021.03.025
13. Zhao, X.Q., Song, H.Y., Sun, Y.X., et al.: Analysis of HPV infection status of female in some areas of Qingdao City. Precis. Clin. Med. **36**(5), 428–432 (2021). https://doi.org/10.13362/j.jpmed.202105013. (in Chinese)
14. Zeng, C.L., Zhu, J.F., Zhu, G.N., et al.: Comparative study of HPV genotypes in 8876 patients with condyloma acuminatum of different genders. China J. Lepr. Skin Dis. **36**(5), 272–274 (2020). https://doi.org/10.12144/zgmfskin202005272. (in Chinese)
15. Sun, L.: Analysis of sex difference of HPV infection in Zhengzhou. J. Henan Med. Coll. **34**(5), 571–575 (2022). https://doi.org/10.3969/j.issn.1008-9276.2022.05.012. (in Chinese)
16. Zhang, X.L.: Study on the correlation between immune cells and tumor markers and HPV levels in high-risk HPV infected patients with cervical cancer. PLA J. Prev. Med. **37**(12), 163–165 (2019). https://doi.org/10.13704/j.cnki.jyyx.2019.12.067. (in Chinese)
17. Gonalves, M.A., Donadi, E.A.: Immune cellular response to HPV: current concept. Braz. J. Infect. Dis. **8**(1), 1–9 (2004). https://doi.org/10.1590/S1413-86702004000100001
18. Liu, Y., Chen, R., Wang, W.N.: Analysis of screening and prevention knowledge of cervical cancer in urban women in Urumqi from 2014 to 2020. Cap. J. Public Health **16**(6), 367–371 (2022). https://doi.org/10.16760/j.cnki.sdggws.2022.06.008. (in Chinese)
19. Hoang, H.T., Ishizaki, A., Nguyen, C.H., Tran, V.T., et al.: Infection with high-risk HPV types among female sex workers in northern Vietnam. J. Med. Virol. **85**(2), 288–294 (2013). https://doi.org/10.1002/jmv.23456
20. Sari, E.R., Adi-Kusumo, F., Aryati, L.: Mathematical analysis of a SIPC age-structured model of cervical cancer. Math. Biosci. Eng. **19**(6), 6013–6039 (2022). https://doi.org/10.3934/mbe.2022281
21. Phan, T.A., Sarower, F., Duan, J., Tian, J.P.: Stochastic dynamics of human papillomavirus delineates cervical cancer progression. J. Math. Biol. **87**(6), 85 (2023). https://doi.org/10.1007/s00285-023-02018-z
22. Ziyadi, N.: A male-female mathematical model of human papillomavirus (HPV) in African American population. Math. Biosci. Eng. **14**(1), 339–358 (2017). https://doi.org/10.3934/mbe.2017022
23. Rifhat, R., Teng, Z.D., Wang, L., et al.: Mathematical modeling analysis and simulation of human papillomavirus infection and cervical cancer in Xinjiang, China. Math. Methods Appl. Sci. **46**(18), 18651–18671 (2023). https://doi.org/10.1002/mma.9584
24. Gao, S., Martcheva, M., Miao, H., Rong, L.: A two-sex model of human papillomavirus infection: vaccination strategies and a case study. J. Theor. Biol. **536**, 111006 (2022). https://doi.org/10.1016/j.jtbi.2022.111006
25. Zhang, X.B., Yan, D.Y., Chen, C., et al.: Research progress of basic reproduction number and effective reproduction number index in epidemiology of infectious diseases. Chin. J. Dis. Control **25**(07), 753–757+790 (2021)
26. Cauchemez, S., Boelle, P.Y., Donnelly, C.A., et al.: Real-time estimates in early detection of SARS. Emerg. Infect. Dis. **12**(1), 110–113 (2006). https://doi.org/10.3201/eid1201.050593
27. Bolle, P.Y., Bernillon, P., Desenclos, J.C.: A preliminary estimation of the reproduction ratio for new influenza A(H1N1) from the outbreak in Mexico, March-April 2009. Euro Surveill. **14**(19), 19205 (2009). https://doi.org/10.2807/ese.14.19.19205-en
28. Han, K., Jia, W.P., Cao, W.Z., et al.: Estimation of basic regeneration number of novelcoronavirus pneumonia in first-tier cities and assessment of epidemic status. Acad. J. Chin. Pla Med. Sch. **41**(4), 421 (2020). https://doi.org/10.3969/j.issn.2095-5227.2020.04.023. (in Chinese)
29. Obadia, T., Haneef, R., Bolle, P.Y.: The R0 package: a toolbox to estimate reproduction numbers for epidemic outbreaks. BMC Med. Inform. Decis. Mak. **12**, 147 (2012). https://doi.org/10.1186/1472-6947-12-147
30. Wang, Y., You, X.Y., Wang, Y.J., et al.: Assessment of the basic reproduction number of the novel coronavirus pneumonia epidemic in China. Chin. J. Epidemiol. **41**(04), 476–479 (2020). (in Chinese)
31. Nikbakht, R., Baneshi, M.R., Bahrampour, A., Hosseinnataj, A.: Comparison of methods to estimate basic reproduction number (R_0) of influenza, using Canada 2009 and 2017-18 A (H1N1) data. J. Res. Med. Sci. **24**, 67 (2019). https://doi.org/10.4103/jrms.JRMS-888-18
32. Svensson, A.: A note on generation times in epidemic models. Math. Biosci. **208**, 300–311 (2007). https://doi.org/10.1016/j.mbs.2006.10.010
33. Wallinga, J., Lipsitch, M.: How generation intervals shape the relationship between growth rates and reproductive numbers. Proc. R. Soc. B, Biol. Sci. **274**(1609), 599–604 (2007). https://doi.org/10.1098/rspb.2006.3754
34. Hens, N., Van Ranst, M., Aerts, M., et al.: Estimating the effective reproduction number for pandemic influenza from notification data made publicly available in real time: a multi country analysis for influenza A/H1N1v 2009. Vaccine **29**, 896–904 (2011). https://doi.org/10.1016/j.vaccine.2010.05.010
35. White, L.F., Pagano, M.: A likelihood-based method for real-time estimation of the serial interval and reproductive number of an epidemic. Stat. Med. **27**(16), 2999–3016 (2008). https://doi.org/10.1002/sim.3136
36. Bettencourt, L.M.A., Ribeiro, R.M.: Real time Bayesian estimation of the epidemic potential of emerging infectious diseases. PLoS ONE **3**, e2185 (2008). https://doi.org/10.1371/journal.pone.0002185

37. Wallinga, J., Teunis, P.: Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. Am. J. Epidemiol. **160**, 509 (2004). https://doi.org/10.1093/aje/kwh255
38. Cauchemez, S., Boëlle, P.Y., Donnelly, C.A.: Real-time estimates in early detection of SARS. Emerging infectious diseases **12**(1), 110 (2006)
39. Deng, L.Y.: Public awareness and acceptance of HPV vaccine in China. Nanjing University (2018). (In Chinese)

Publisher's Note