





**Emerging Viral Infections in Vulnerable  
Populations: Epidemiology and mathematical  
modeling**

**Yunpeng Ji**

**冀云鹏 著**

The studies presented in this thesis were performed at the Laboratory of Gastroenterology and Hepatology, Erasmus MC-University Medical Center Rotterdam, the Netherlands.

© Copyright by Yunpeng Ji. All rights reserved. Seed855@163.com  
No part of the thesis may be reproduced or transmitted, in any form, by any means, without express written permission of the author.

Cover design: the author of this thesis  
Layout design: the author of this thesis.

Printed by: Ridderprint, [www.ridderprint.nl](http://www.ridderprint.nl)

ISBN: 978-94-6458-864-4

# Emerging Viral Infections in Vulnerable Populations: Epidemiology and mathematical modeling

Opkomende virale infecties in kwetsbare groepen:  
Epidemiologie en wiskundige modelering

## Thesis

to obtain the degree of Doctor from the  
Erasmus University Rotterdam  
by command of the  
rector magnificus

Prof. dr. A. L. Bredenoord

and in accordance with the decision of the Doctorate Board

The public defense shall be held on

Wednesday, 1 February 2023 at 13:00 hrs

by

**Yunpeng Ji**

born in Hohhot, Inner Mongolia, China

## **Doctoral Committee**

### **Promotor:**

Prof. dr. M.P. Peppelenbosch

### **Inner Committee:**

Prof. dr. R.A. de Man

Prof. dr. M.K. Ikram

Dr. S. van den Hof

### **Copromotor:**

Dr. Q. Pan

# TABLE OF CONTENTS

Chapter 1 .....	1
General introduction and outline of the thesis	
<b><i>PART I: HPV infection in women</i></b>	
Chapter 2 .....	17
Prevalence of human papillomavirus infection in women in the Autonomous Region of Inner Mongolia: A population-based study of a Chinese ethnic minority	
	Journal of Medical Virology. 2018, 90:148-156
Chapter 3 .....	35
The Burden of Human Papillomavirus and <i>Chlamydia trachomatis</i> Coinfection in Women: A Large Cohort Study in Inner Mongolia, China	
	The Journal of Infectious Diseases. 2019, 219(2):206-214
Chapter 4 .....	55
Optimizing Cervical Cancer Screening and Prevention in Rural Areas of China: A Modeling Analysis up to 2030	
	embargoed
<b><i>PART II: HEV cross-species transmission and infection in pregnant women</i></b>	
Chapter 5 .....	59
Estimating the global prevalence of hepatitis E virus in swine and pork products	
	One Health. 2021, 14:100362
Chapter 6 .....	73
Estimating the burden and modeling mitigation strategies of pork-related hepatitis E virus foodborne transmission in representative European countries	
	One Health. 2021, 13:100350
Chapter 7 .....	93
Prevalence and clinical features of hepatitis E virus infection in pregnant women: A large cohort study in Inner Mongolia, China	
	Clinics and Research in Hepatology and Gastroenterology. 2021, 45(4):101536

***PART III: COVID-19 spread and control measures***

Chapter 8.....	113
Potential association between COVID-19 mortality and health-care resource availability	
	Lancet Global Health. 2020, 8, e480
Chapter 9.....	119
A multi-regional, hierarchical-tier mathematical model of the spread and control of COVID-19 epidemics from epicentre to adjacent regions	
	Transboundary and Emerging Diseases. 2021, 00:1-10
Chapter 10 .....	149
Distinct effectiveness in containing COVID-19 epidemic: comparative analysis of two cities in China by mathematical modeling	
	PLOS Global Public Health. 2021.1(11): e0000043
Chapter 11 .....	171
Summary and Discussion	
Chapter 12 .....	183
Dutch Summary	
Appendix .....	189
Acknowledgements	
Publications	
PhD Portfolio	
Curriculum Vitae	



# **Chapter 1**

**General introduction and outline of the thesis**



Viruses are nanometer-size organic particles and propagate by invading living cells. The genetic material of viruses is encoded in either DNA or RNA molecules, packed by viral protein to form an intact biological particle. Some viruses are surrounded by an additional envelope that greatly increases the biodiversity of viruses. The life cycle of viruses majorly includes six steps, namely, attachment, penetration, uncoating, gene expression and replication, particle assembly, and release. The hosts of viruses are nearly all forms of life on earth, including bacteria, plants, insects, reptiles, fishes, avian, and mammals. As far as the most concerned, viral infections in humans can cause severe diseases. Viruses can be transmitted by a regular human-to-human route, or more subtly, transmitted by infecting livestock and poultries in priority then follow various routines to reach human hosts indirectly.

The human body is gifted with powerful defense mechanisms against pathogen invasion. However, dysregulation of these mechanisms may cause failure in clearing the infection and trigger pathogenesis to the body. Further, a small group of highly pathogenic viruses can cause super-severe infections regardless of the status of the host. These include the few most deadly viruses, such as rabies, Marburg and Ebola viruses, which have fatality rates over 50% and some can even reach 100% in specific populations. Besides, the human population has been constantly threatened by emerging and re-emerging pathogenic viruses. From the 1918 influenza pandemic to the ongoing coronavirus pandemic, viruses never stop. Many elements are likely to contribute to the emergence of new viral diseases, including virus evolution, environmental changes, population mitigations, and factors relevant to the diversity of human demography, cultures and behaviors<sup>[1]</sup>.

For most of the pathogenic viruses, the disease burden and clinical outcome are highly context- and host-dependent. The risk and burden of a new viral disease often displays spatial and temporal differences. Thus, in this thesis, I focus on the study of viral infections that are affecting specific vulnerable populations. These viruses include human papillomaviruses (HPVs), hepatitis E virus (HEV) and SARS-CoV-2.

## **Human papillomaviruses**

HPVs are a group of small non-enveloped double-stranded DNA viruses. More than 150 HPV genotypes have been identified so far, of which 12 genotypes (16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, and 59) are classified as *carcinogenic to humans* by the International Agency for Research on Cancer. HPV transmission generally occurs via sexual exposure, especially the newly sexually active adolescent or young adult women who are at the highest risk of virus contact<sup>[2]</sup>. The majority (about 90%) of newly occurred HPV infections are undetectable within 1-2 years, which is attributed to immune clearance, or viral latent<sup>[3,4]</sup>. Besides, serum antibody response can be HPV-genotype specific, and will not confer full immunological protection

against re-infection<sup>[5,6]</sup>. Therefore, individuals will remain at risk for HPV-related complications even following apparently successful clearing of an early infection. A minority of HPV infections can be persistent beyond 12 months, which increase the risk of carcinogenic development of cervical pre-malignancy (namely, high grade squamous intraepithelial lesions (HSIL) or CIN grade 2 or 3 (CIN2/3)) and cancer, if the lesions remain undetected and untreated<sup>[7]</sup>.

### **Burdens and prevention of cervical cancer**

Cervical cancer is the fourth most common cancer among women globally, with an estimated 604,127 new cases and 341,831 deaths in 2020<sup>[8]</sup>. In developed countries, programs of HPV screening and vaccination started relatively earlier. In the Netherlands, the screening for cervical cancer has been common since the 1970's. In 2007, Australia launched a HPV vaccination program, which made it the first western country to introduce HPV vaccine to the female citizens. A mathematical modeling estimated that, by 2020, Australia would reduce the incidence of cervical cancer to fewer than six new cases, and, by 2028, to fewer than four new cases per 100,000 women<sup>[9]</sup>. The numbers actually observed (913 new cases for a population of 27 million persons in 2021) are in line with the model. Subsequently, a routine HPV vaccination program in Scotland was launched in 2008<sup>[10]</sup>, and a similar program in the Netherlands followed in 2009. As the improving of biotechnology, multi-valent vaccines have been designed and applied. In England, in comparison with the continuing two-valent vaccination program (against HPV 16 and 18), the introduction of a nine-valent vaccine in 2019 is estimated to further reduce 36% of cervical cancers, and decrease incidence further from 9.5 to 6.1 per 100,000 women by 2036-40<sup>[11]</sup>.

However, in developing countries, cervical cancer accounts for the second mostly diagnosed cancer second to breast cancer, and is the third leading cause of cancer-related death after breast and lung cancers<sup>[8]</sup>. Moreover, more than 90% of global cervical cancer deaths occur in these regions, especially in the areas of India, which represents one quarter. Regions, with a higher incidence of HIV, like sub-Saharan Africa, bear more disease burdens<sup>[12]</sup>. Pap testing and HPV screening have been proven to be effective, but their use in low-income countries is limited due to economical and logistical obstacles in health systems and infrastructure<sup>[13,14]</sup>. It is estimated that there are still two to three generations of elder women who are beyond the age range of vaccination, and some of whom have already acquired the infection that shall not be vaccinated<sup>[15, 16]</sup>. Although cervical cancer is considered<sup>[15, 16]</sup> preventable due to the slow disease progression and the availability of preventing methods<sup>[17]</sup>, the elimination of cervical cancer requires tremendous efforts from multiple stakeholders, and is still a challenging task in these areas. Apart from cervical cancer, HPV is thought to be the leading cause of anal

cancer (estimates range up to 90 %), of vaginal cancers (65 %), vulvar cancers (50 %), and oropharyngeal cancers (estimates range from 45 % to 90 %)<sup>[18]</sup>.

### **HPV transmission**

Although sexual transmission is best documented in HPV infection, numerous studies suggest non-sexual routes as well. The horizontal transmission mechanisms of HPV include vomiting as well as tactile, oral and dermal contact (other than sexual). HPVs are very sturdy viruses, resistant to heat and lyophilization, showing still 30% infectivity after seven days of dehydration<sup>[19]</sup>. Hence, they have the capacity to remain contagious on surfaces, clothing and gynecological equipment and fomites. A main concern is that most individuals are not aware of being infected because they are asymptomatic. Two different studies, one conducted in a university and one in a hospital, have revealed an abundance of HPV DNA on the fingers of infected persons<sup>[18, 20]</sup>. These evidences collectively show that the route of HPV transmission is primarily through skin-to-skin or skin-to-mucosa pattern, with or without sexual contact.

### **Hepatitis E Virus**

HEV is a nonenveloped virus as well with an icosahedral capsid and a positive-sense, single-stranded RNA genome. It includes four main infectious genotypes, namely, HEV 1, 2, 3, and 4. HEV1 and HEV2 transmit majorly via contaminated water or fecal shedding from infected persons, while HEV3 and HEV4 are zoonotic. Animal reservoirs are mainly domestic pigs, wild boar, deer, and rabbits<sup>[21]</sup>. HEV3 is thought of widely circulating in Europe. Between 2005 and 2015, 21,018 hepatitis E cases had been reported from 22 European countries. By screening blood donors, the incidence of HEV in Europe was observed ranging from 1/4771 to 1/762<sup>[22]</sup>. In general, although HEV infection is self-limiting in healthy individuals, HEV acquisition in immunocompromised patients, in particular organ transplant recipients who require immunosuppressive medication, bear higher risk of developing chronic infection, which may rapidly progress into fibrosis and cirrhosis<sup>[23]</sup>. Blood transfusion may contribute as well to chronic HEV cases in organ transplant patients<sup>[24]</sup>. However, viral transmission via contacting animal reservoirs or consuming contaminated livestock products, thought as an important part of HEV 3 undercurrent in Europe, is still quantitatively undetermined.

### **HEV zoonotic transmission**

Extensive evidence has proved the HEV cross-species transmission between human and animal reservoirs<sup>[25]</sup>. HEV strains isolated from animals in pig farms, slaughter houses or nearby environment are genetically close to viruses identified in HEV-infected patients. Case reports documented the HEV strains isolated from patient blood, and comparatively found genetic similarity to those sampled from animal food the patient recently ingested<sup>[26]</sup>. Pig

farms are thought of one of the viral sources as HEV transmission within the herds has been observed. In Europe, the virologic prevalence of HEV ranges from 10% to 100% at the farm-scale<sup>[27,28]</sup>, while the individual virologic incidence within the farm ranges from 1% to 89%. Importantly, the HEV incidence in a given farm or area seldom changes over a period, implying HEV is constantly circulating<sup>[27]</sup>. In pork food industry, the fraction of contaminated food fluctuates from 1% to 50%<sup>[29,30]</sup>. Zoonotic HEV strains are frequently encountered in raw pig livers or pork-derived food such as meat and sausage. Thus, it is necessary to investigate HEV contamination at each section of modern pork food industry, including animal breeding, slaughtering, product processing and so on<sup>[31]</sup>. Due to the lack of rigorous surveillance of HEV and the huge and complex network of food industry, it is still a difficult job to do so.

### **HEV infection in pregnancy**

Another heavy clinical burden of HEV infection is observed in pregnant women, which is associated with HEV 1 and HEV 2. Globally, in 2015, the number of symptomatic cases of hepatitis E and HEV-relevant deaths were estimated to be 3.4 million and 70,000, respectively, of which a significant fraction occurred in pregnant women<sup>[32]</sup>. Severe HEV cases in pregnancy are mostly observed in developing countries, although the nongeographic disparities existed. For example, studies from Egypt reported that the progression and severity of HEV infection are close between pregnant and non-pregnant women<sup>[33]</sup>. To the contrary, evidence from India on severe HEV infection and substantial maternal mortality rates has consistently accumulated<sup>[34,35]</sup>. Acute hepatitis is an important cause of death among pregnant women, and mostly occurs in the third trimester. Of note, vertical transmission has been observed in HEV infected mothers as well, and can reach up to 46%<sup>[36,37]</sup>. The common clinical presentations of neonatal HEV infection include jaundice, hepatosplenomegaly, respiratory distress syndrome and sepsis<sup>[36, 38]</sup>. Some cases may result in stillbirth or even neonatal death<sup>[36,38]</sup>. Unfortunately, supportive care is the only available approach for pregnant and neonatal HEV infection, and a huge lack of public concern on this vulnerable population needs to be changed.

### **SARS-CoV-2**

SARS-CoV-2 is a new member of coronaviruses. It invades bronchial epithelial cells, pneumocytes and upper respiratory tract cells, causing mild to severe, even life-threatening respiratory diseases. At the end of 2019, SARS-CoV-2 firstly emerged in Wuhan, China, leading to an outbreak of viral pneumonia, which is now designated as coronavirus disease 2019 (COVID-19). This pandemic has infiltrated every corner of world, resulting in, around December 2022, over 648 million confirmed infections and over 6 million deaths (<https://www.worldometers.info/coronavirus/>). The death fatality of SARS-CoV-2 could reach

up to 2% in some vulnerable populations, constituting of elderly people and patients with pre-existing conditions such as diabetes, obesity, respiratory illness and cardiovascular diseases.

To control the spread of COVID-19, control measures have been implemented such as social distancing, travel restriction, city block-down and so on. The resultant effectiveness of the controls may vary dramatically among different areas, settings and time periods<sup>[39]</sup>. Fortunately, targeting vaccines have been expeditiously developed and applied to aid public health globally. The accessibility and affordability of these vaccines remain major issues for many resource-limited regions. Furthermore, the emergence of new variants of concern, notably the lineages alpha, beta, gamma and delta, is associated with potential immune escape, compromised vaccine effect and forces a re-adjustment of vaccine contents<sup>[40]</sup>. Thus, the eventual outcome of the COVID-19 pandemic remains uncertain, and joint efforts are essential for coping with this public crisis.

### **Mathematical modeling**

Model simulation offers an important tool for understanding the dynamics of virus transmission and the virus-associated clinical burdens. It is used to estimate the effect of interventions as well by providing a quantitative framework for integrating and interpreting disease relevant data into policy-relevant outcomes<sup>[41,42]</sup>. In simulation, the entire process of viral epidemic will be rewritten in mathematical languages, and important concepts like transmission rate, sizes of susceptible populations and the eventual epidemic outcome will be estimated. The modeling-based estimate is able to help policy makers have an effective and rapid response to viral outbreak and pandemic.

In an entirely susceptible population, the expected number of offspring of an infectious case is called the effective reproduction number ( $R_t$ ). It is named the basic reproduction number ( $R_0$ ) as well when the viral transmission just starts. The size of  $R_0$  or  $R_t$  often decides how fast an infectious disease can transmit among the susceptible population, and a value of  $R_t$  smaller than one indicates that the disease is being eliminated or can hardly lead to an epidemic. Thus, the estimate of  $R_0$  or  $R_t$  is essential for understanding the dynamics of an infectious disease and preparing appropriate measures to reduce its influences.

Model simulation includes numerous steps including model framework building, parameter estimation (or data fitting), model calibration and model sensitivity analysis. The Susceptible-(Exposed)-Infectious-Recovered framework (SIR/SEIR) is a classic ordinary-differential-based model for simulating infectious disease. SIR models can be deterministic, in which a set of parameters will produce a set of data outcome, or stochastic, dealing with randomness and generate results with minor or obvious differences. To capture the transmission dynamics of a disease in real world, models should be subjected to uncertainty. A common strategy of

managing uncertainty is to run the model numerous times with a range of different parameters to generate sets of simulations. Parameters will be given in the form of a range of numbers or a type of distribution. Data fitting is essential for finding the most appropriate parameters to have model be consistent with the observed epidemic. This can be performed by least square, a simple method to reduce the differences between observed data and model outcomes, or by maximum likelihood estimation, an approach of parameter estimation by computing the likelihood of observed data on the basis of existing model and priori parameters<sup>[43]</sup>. Another important concept, sensitivity analysis, focuses on to what a degree the changing of certain parameters will influence the outcomes of model simulations. It assesses the stability of a model, and estimates which parameter may influence the model simulating greatly.

### **Scope of this thesis**

The aim of this thesis is to understand the viral infection and transmission from numerous perspectives. Briefly, elements with epidemiological features, natural history, risk factors and clinical burdens were analyzed in the investigation of HPV, HEV and SARS-CoV-2 in a specific population or area. I prioritized these three pathogenic viruses that have distinct model-of-transmission for comprehensive understanding of epidemic spread and control. Agents involved in co-infection were studied as well to deepen the understanding of pathogenesis interaction. Furthermore, mathematical modeling was used to estimate several key features of an epidemic, such as the dynamic of viral transmission, potential burdens and the effectiveness of controls.

### **Outline of this thesis**

Emerging and re-emerging viral infections are continuously threatening the society, and multidimensional research is needed to understand and cope with these threats. In **Chapter 2**, I investigated the prevalence of HPV and its association with cervical carcinogenesis in women in the Inner Mongolia of China. The studied region is located in Northwest of China and contains a multi-ethnic population. Mongolians, as the major ethnic population in this region, have distinct cultural traditions such as sheltering, hunting, and costume use. Urbanization has reshaped the traditions of Mongolians, although a small fraction remains live in pastoral areas with a mixed half-modern, and half-grazing lifestyle. Thus, this region is ideal for studying how life style, genetics, and culture impact on HPV transmission and prevalence. In **Chapter 3**, I studied the part of *Chlamydia Trachomatis* (*C. Trachomatis*) in HPV persistent infection and the potential mechanism involved. Generally, *C. Trachomatis* infection is asymptomatic and not prevalent, however, its role in persistent progression of HPV infection is undetermined. I found *C. Trachomatis* co-infection appears to accelerate the progression of



HPV-related cervical carcinogenesis, and therefore it raises another question that whether *C. Trachomatis* screening is necessary in the prevention and control of cervical cancer mainly caused by HPV. In **Chapter 4**, I employed a modified SIR model to predict the trend of cervical cancer in a rural population of China by the year of 2030 with various scenarios, including strategies on the basis of HPV screening, cytological screening and vaccination, respectively. The influence of scenario with no intervention was estimated as well.

Next, I studied HEV and its transmission in different subjects. In **Chapter 5**, I studied the global prevalence of HEV in swine and related food products by meta-analysis. It shows that nearly 60% of domestic pigs and 27% of wild boars have ever encountered with HEV infection. Nearly 13% of domestic and 9.5% of wild swine are actively infected by finding detectable HEV RNA. Of note, about 10% of commercial pork products are HEV RNA positive. In **Chapter 6**, I simulated the dynamic of foodborne transmission of HEV in four European areas. The research target is to estimate the effect of hypothetically implementation of HEV vaccine for prevention of such transmission from swine to human. The model shows vaccination in animals is highly effective as it modified the dynamics of HEV transmission in pig herds. To maximize the effect of HEV vaccine, I modeled a limited vaccination scenario targeting at a sub-population that consumes pork food most frequently. The simulation shows this scenario could be effective as well and is capable of substituting the strategy tendentious to the general population. In **Chapter 7**, I investigated the epidemiological and clinical features of HEV infection in pregnant women in Inner Mongolia of China. I observed 6.0% of participants were anti-HEV IgG antibody positive and 0.6% were anti-HEV IgM antibody positive. However, HEV viral RNA was not detected. Pregnant women with recent/ongoing HEV infection, as indicated by anti-HEV IgM positivity, have slightly higher ALT levels, and HEV-infected expecting mothers were at higher risk for developing hyperlipidemia, preterm delivery and neonatal jaundice.

Next, I studied SARS-CoV-2. In **Chapter 8**, I investigated the fatality rate of COVID-19 at the early stage by comparing the epidemical data collected in Wuhan and other parts of China. I found a significantly higher mortality rate of COVID-19 in Wuhan, which is partially attributed to the local crisis of less accessible to the healthcare resources. Unfortunately, before implementing city-lockdown, the virus has already spread out of Wuhan. In **Chapter 9**, I build a multi-regional, hierarchical-tier model, SLIHR, for better understanding the complexity and heterogeneity of COVID-19 spreading to multiple areas in China. By fitting the epidemiological and population flow data, the model revealed important insight into how COVID-19 spread to different regions by various transportations. It showed that the viral transmission is highly dependent on two elements, the size of mitigating populations, and how fast the populations mitigate. Another note is that lessons can be learned from the control of early COVID-19 in different parts of China. In **Chapter 10**, after COVID-19 reached and encountered resistance in most major cities in China, I comparatively simulated the spreading and control of COVID-

19 in Wuhan and another virus raging city, Wenzhou. I found the joint of rapid hospital admission and stringent quarantine within-population is effective to prevent COVID-19 spreading in regions where epidemic is majorly caused by viral importation. Therefore, the study offers important insights for improving preparedness and early response to future emerging epidemic/pandemic.

## References

1. Choi YK. Emerging and re-emerging fatal viral diseases. *Exp Mol Med*. 2021 May;53(5):711-712. doi: 10.1038/s12276-021-00608-9. Epub 2021 May 6. PMID: 33953321; PMCID: PMC8099700.
2. Burchell AN, Winer RL, de Sanjosé S, Franco EL. Chapter 6: Epidemiology and transmission dynamics of genital HPV infection. *Vaccine*. 2006 Aug 31;24 Suppl 3: S3/52-61. doi: 10.1016/j.vaccine.2006.05.031. Epub 2006 Jun 2. PMID: 16950018.
3. Fu TC, Carter JJ, Hughes JP, Feng Q, Hawes SE, Schwartz SM, Xi LF, Lasof T, Stern JE, Galloway DA, Koutsky LA, Winer RL. Re-detection vs. new acquisition of high-risk human papillomavirus in mid-adult women. *Int J Cancer*. 2016 Nov 15;139(10):2201-12. doi: 10.1002/ijc.30283. Epub 2016 Aug 4. PMID: 27448488; PMCID: PMC5222584.
4. Liu SH, Cummings DA, Zenilman JM, Gravitt PE, Brotman RM. Characterizing the temporal dynamics of human papillomavirus DNA detectability using short-interval sampling. *Cancer Epidemiol Biomarkers Prev*. 2014 Jan;23(1):200-8. doi: 10.1158/1055-9965.EPI-13-0666. Epub 2013 Oct 15. PMID: 24130223; PMCID: PMC3947138.
5. Carter JJ, Koutsky LA, Hughes JP, Lee SK, Kuypers J, Kiviat N, Galloway DA. Comparison of human papillomavirus types 16, 18, and 6 capsid antibody responses following incident infection. *J Infect Dis*. 2000 Jun;181(6):1911-9. doi: 10.1086/315498. Epub 2000 May 31. PMID: 10837170.
6. Gravitt PE. Evidence and impact of human papillomavirus latency. *Open Virol J*. 2012; 6: 198-203. doi: 10.2174/1874357901206010198. Epub 2012 Dec 28. PMID: 23341855; PMCID: PMC3547385.
7. Schiffman M, Castle PE, Jeronimo J, Rodriguez AC, Wacholder S. Human papillomavirus and cervical cancer. *Lancet*. 2007 Sep 8;370(9590):890-907. doi: 10.1016/S0140-6736(07)61416-0. PMID: 17826171.
8. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2018 Nov;68(6):394-424. doi: 10.3322/caac.21492. Epub 2018 Sep 12. Erratum in: *CA Cancer J Clin*. 2020 Jul;70(4):313. PMID: 30207593.
9. Hall MT, Simms KT, Lew JB, Smith MA, Brotherton JM, Saville M, Frazer IH, Canfell K. The projected timeframe until cervical cancer elimination in Australia: a modelling study. *Lancet Public Health*. 2019 Jan;4(1): e19-e27. doi: 10.1016/S2468-2667(18)30183-X. Epub 2018 Oct 2. PMID: 30291040.
10. Kavanagh K, Pollock KG, Cuschieri K, Palmer T, Cameron RL, Watt C, Bhatia R, Moore C, Cubie H, Cruickshank M, Robertson C. Changes in the prevalence of human papillomavirus following a national bivalent human papillomavirus vaccination programme in Scotland: a 7-year cross-sectional study. *Lancet Infect Dis*. 2017 Dec;17(12):1293-1302. doi: 10.1016/S1473-3099(17)30468-1. Epub 2017 Sep 28. PMID: 28965955.
11. Johnson HC, Lafferty EI, Eggo RM, Louie K, Soldan K, Waller J, Edmunds WJ. Effect of HPV vaccination and cervical cancer screening in England by ethnicity: a modelling study. *Lancet Public Health*. 2018 Jan;3(1): e44-e51. doi: 10.1016/S2468-2667(17)30238-4. Epub 2017 Dec 19. PMID: 29307388; PMCID: PMC5765530.
12. De Vuyst H, Alemany L, Lacey C, Chibwasha CJ, Sahasrabudhe V, Banura C, Denny L, Parham GP. The burden of human papillomavirus infections and related diseases in sub-saharan Africa.

- Vaccine. 2013 Dec 29;31 Suppl 5(0 5): F32-46. doi: 10.1016/j.vaccine.2012.07.092. PMID: 24331746; PMCID: PMC4144870.
13. Sankaranarayanan R. Screening for cancer in low- and middle-income countries. *Ann Glob Health*. 2014 Sep-Oct;80(5):412-7. doi: 10.1016/j.aogh.2014.09.014. PMID: 25512156.
  14. Alliance for Cervical Cancer Prevention. Pap smears: an important but imperfect screening method. Lyon, France: International Agency for Research on Cancer; 2002. Available from: [http://screening.iarc.fr/doc/RH\\_pap\\_smears.pdf](http://screening.iarc.fr/doc/RH_pap_smears.pdf).
  15. World Health Organization. Human papillomavirus vaccines: WHO position paper, October 2014. *Wkly Epidemiol Rec* 2014; 89: 465–91.
  16. Campos NG, Castle PE, Wright TC Jr, Kim JJ. Cervical cancer screening in low-resource settings: A cost-effectiveness framework for valuing tradeoffs between test performance and program coverage. *Int J Cancer*. 2015 Nov 1;137(9):2208-19. doi: 10.1002/ijc.29594. Epub 2015 May 21. PMID: 25943074; PMCID: PMC4910518.
  17. Committee on Adenocarcinoma Health Care Immunization Expert Work Group. Committee opinion no. 641: human papillomavirus vaccination. *Obstet Gynecol* 2015; 126: e38–43.
  18. Boda D, Neagu M, Constantin C, Voinescu RN, Caruntu C, Zurac S, Spandidos DA, Drakoulis N, Tsoukalas D, Tsatsakis AM. HPV strain distribution in patients with genital warts in a female population sample. *Oncol Lett*. 2016 Sep;12(3):1779-1782. doi: 10.3892/ol.2016.4903. Epub 2016 Jul 22. PMID: 27602111; PMCID: PMC4998207.
  19. Casalegno JS, Le Bail Carval K, Eibach D, Valdeyron ML, Lamblin G, Jacquemoud H, Mellier G, Lina B, Gaucherand P, Mathevet P, Mekki Y. High risk HPV contamination of endocavity vaginal ultrasound probes: an underestimated route of nosocomial infection? *PLoS One*. 2012;7(10): e48137. doi: 10.1371/journal.pone.0048137. Epub 2012 Oct 24. PMID: 23110191; PMCID: PMC3480505.
  20. Ryndock EJ, Meyers C. A risk for non-sexual transmission of human papillomavirus? *Expert Rev Anti Infect Ther*. 2014 Oct;12(10):1165-70. doi: 10.1586/14787210.2014.959497. PMID: 25199987.
  21. Rose N, Lunazzi A, Dorenlor V, Merbah T, Eono F, Eloit M, Madec F, Pavio N. High prevalence of Hepatitis E virus in French domestic pigs. *Comp Immunol Microbiol Infect Dis*. 2011 Sep;34(5):419-27. doi: 10.1016/j.cimid.2011.07.003. Epub 2011 Aug 27. PMID: 21872929.
  22. Hartl J, Otto B, Madden RG, Webb G, Woolson KL, Kriston L, Vettorazzi E, Lohse AW, Dalton HR, Pischke S. Hepatitis E Seroprevalence in Europe: A Meta-Analysis. *Viruses*. 2016 Aug 6;8(8):211. doi: 10.3390/v8080211. PMID: 27509518; PMCID: PMC4997573.
  23. Zhou X, de Man RA, de Knecht RJ, Metselaar HJ, Peppelenbosch MP, Pan Q. Epidemiology and management of chronic hepatitis E infection in solid organ transplantation: a comprehensive literature review. *Rev Med Virol*. 2013 Sep;23(5):295-304. doi: 10.1002/rmv.1751. Epub 2013 Jul 1. PMID: 23813631.
  24. Kamar N, Izopet J, Pavio N, Aggarwal R, Labrique A, Wedemeyer H, Dalton HR. Hepatitis E virus infection. *Nat Rev Dis Primers*. 2017 Nov 16; 3: 17086. doi: 10.1038/nrdp.2017.86. PMID: 29154369.
  25. Meng XJ, Halbur PG, Shapiro MS, Govindarajan S, Bruna JD, Mushahwar IK, Purcell RH, Emerson SU. Genetic and experimental evidence for cross-species infection by swine hepatitis E virus. *J Virol*. 1998 Dec;72(12):9714-21. doi: 10.1128/JVI.72.12.9714-9721.1998. PMID: 9811705; PMCID: PMC110481.
  26. EFSA Panel on Biological Hazards (BIOHAZ), Ricci A, Allende A, Bolton D, Chemaly M, Davies R, Fernandez Escamez PS, Herman L, Koutsoumanis K, Lindqvist R, Nørrung B, Robertson L, Ru G, Sanaa M, Simmons M, Skandamis P, Snary E, Speybroeck N, Ter Kuile B, Threlfall J, Wahlström H, Di Bartolo I, Johne R, Pavio N, Rutjes S, van der Poel W, Vasickova P, Hempen M, Messens W, Rizzi

- V, Latronico F, Girones R. Public health risks associated with hepatitis E virus (HEV) as a food-borne pathogen. *EFSA J.* 2017 Jul 11;15(7): e04886. doi: 10.2903/j.efsa.2017.4886. PMID: 32625551; PMCID: PMC7010180.
27. Salines M, Andraud M, Rose N. From the epidemiology of hepatitis E virus (HEV) within the swine reservoir to public health risk mitigation strategies: a comprehensive review. *Vet Res.* 2017 May 25;48(1):31. doi: 10.1186/s13567-017-0436-3. PMID: 28545558; PMCID: PMC5445439.
  28. Casas M, Pujols J, Rosell R, de Deus N, Peralta B, Pina S, Casal J, Martín M. Retrospective serological study on hepatitis E infection in pigs from 1985 to 1997 in Spain. *Vet Microbiol.* 2009 Mar 30;135(3-4):248-52. doi: 10.1016/j.vetmic.2008.09.075. Epub 2008 Oct 1. PMID: 18996653.
  29. Colson P, Borentain P, Queyriaux B, Kaba M, Moal V, Gallian P, Heyries L, Raoult D, Gerolami R. Pig liver sausage as a source of hepatitis E virus transmission to humans. *J Infect Dis.* 2010 Sep 15;202(6):825-34. doi: 10.1086/655898. PMID: 20695796.
  30. Pavio N, Merbah T, Thébault A. Frequent hepatitis E virus contamination in food containing raw pork liver, France. *Emerg Infect Dis.* 2014 Nov;20(11):1925-7. doi: 10.3201/eid2011.140891. PMID: 25340373; PMCID: PMC4214317.
  31. Nantel-Fortier N, Letellier A, Lachapelle V, Fravallo P, L'Homme Y, Brassard J. Detection and Phylogenetic Analysis of the Hepatitis E Virus in a Canadian Swine Production Network. *Food Environ Virol.* 2016 Dec;8(4):296-304. doi: 10.1007/s12560-016-9252-6. Epub 2016 Jul 15. PMID: 27422131.
  32. World Health Organisation (WHO). Hepatitis E Fact Sheet. World Heal Organ 2020. <https://www.who.int/news-room/fact-sheets/detail/hepatitis-e> (accessed September 20, 2020).
  33. Navaneethan U, Al Mohajer M, Shata MT. Hepatitis E and pregnancy: understanding the pathogenesis. *Liver Int.* 2008 Nov;28(9):1190-9. doi: 10.1111/j.1478-3231.2008.01840.x. Epub 2008 Jul 25. PMID: 18662274; PMCID: PMC2575020.
  34. Jin H, Zhao Y, Zhang X, Wang B, Liu P. Case-fatality risk of pregnant women with acute viral hepatitis type E: a systematic review and meta-analysis. *Epidemiol Infect.* 2016 Jul;144(10):2098-106. doi: 10.1017/S0950268816000418. Epub 2016 Mar 4. PMID: 26939626.
  35. Seifoleslami M. An update of the incidence of fulminant hepatitis due to viral agents during pregnancy. *Interv Med Appl Sci.* 2018 Dec;10(4):210-212. doi: 10.1556/1646.10.2018.40. PMID: 30792915; PMCID: PMC6376349.
  36. Khuroo MS, Khuroo MS. Hepatitis E: an emerging global disease - from discovery towards control and cure. *J Viral Hepat.* 2016 Feb;23(2):68-79. doi: 10.1111/jvh.12445. Epub 2015 Sep 6. PMID: 26344932.
  37. Aggarwal R, Krawczynski K. Hepatitis E: an overview and recent advances in clinical and laboratory research. *J Gastroenterol Hepatol.* 2000 Jan;15(1):9-20. doi: 10.1046/j.1440-1746.2000.02006.x. PMID: 10719741.
  38. El Sayed Zaki M, El Aal AA, Badawy A, El-Deeb DR, El-Kheir NY. Clinicolaboratory study of mother-to-neonate transmission of hepatitis E virus in Egypt. *Am J Clin Pathol.* 2013 Nov;140(5):721-6. doi: 10.1309/AJCPT55TDMJNPLL. PMID: 24124153.
  39. Leung K, Wu JT, Leung GM. Effects of adjusting public health, travel, and social measures during the roll-out of COVID-19 vaccination: a modelling study. *Lancet Public Health.* 2021 Sep;6(9):e674-e682. doi: 10.1016/S2468-2667(21)00167-5. Epub 2021 Aug 11. PMID: 34388389; PMCID: PMC8354806.
  40. Sonabend R, Whittles LK, Imai N, Perez-Guzman PN, Knock ES, Rawson T, Gaythorpe KAM, Djaafara BA, Hinsley W, FitzJohn RG, Lees JA, Kanapram DT, Volz EM, Ghani AC, Ferguson NM, Baguelin M, Cori A. Non-pharmaceutical interventions, vaccination, and the SARS-CoV-2 delta variant in England: a mathematical modelling study. *Lancet.* 2021 Nov 13;398(10313):1825-1835. doi: 10.1016/S0140-6736(21)02276-5. Epub 2021 Oct 28. PMID: 34717829; PMCID: PMC8550916.

41. Basu S, Andrews J. Complexity in mathematical models of public health policies: a guide for consumers of models. *PLoS Med.* 2013 Oct;10(10): e1001540. doi: 10.1371/journal.pmed.1001540. Epub 2013 Oct 29. PMID: 24204214; PMCID: PMC3812083.
42. Knight GM, Dharan NJ, Fox GJ, Stennis N, Zwerling A, Khurana R, Dowdy DW. Bridging the gap between evidence and policy for infectious diseases: How models can aid public health decision-making. *Int J Infect Dis.* 2016 Jan; 42: 17-23. doi: 10.1016/j.ijid.2015.10.024. Epub 2015 Nov 3. PMID: 26546234; PMCID: PMC4996966.
43. Fojo AT, Kendall EA, Kasaie P, Shrestha S, Louis TA, Dowdy DW. Mathematical Modeling of "Chronic" Infectious Diseases: Unpacking the Black Box. *Open Forum Infect Dis.* 2017 Aug 14;4(4): ofx172. doi: 10.1093/ofid/ofx172. Erratum in: *Open Forum Infect Dis.* 2018 Jan 25;5(1): ofx206. PMID: 29226167; PMCID: PMC5716064.



# **PART I**

## ***HPV infection in women***





# Chapter 2

## **Prevalence of human papillomavirus infection in women in the Autonomous Region of Inner Mongolia: A population-based study of a Chinese ethnic minority**

Xiaohua Wang<sup>#</sup>, Yunpeng Ji<sup>#</sup>, Juan Li, Hong Dong, Bo Zhu, Yan Zhou, Jie Wang, Xueyuan Zhou, Yan Wang, Maikel P. Peppelenbosch, Qiuwei Pan, Xiaoping Ji, Dongjun Liu  
(#: contributed equally)

Journal of Medical Virology. 2018, 90:148-156



## Abstract

Human papillomavirus (HPV) is one of the most common sexually transmitted infectious pathogens. Persistent infection has been linked to cancer development, in particular to cervical cancer. This study aims to investigate the epidemiology of HPV infection in women in Inner Mongolia of China and to dissect the disparities between the Han and Mongolian ethnic populations. Cervical cell samples from 5655 women (17-68 years old) were collected during routine gynecologic examination. HPV infection was established using the HPV GenoArray kit detecting 21 HPV genotypes. The overall HPV prevalence was 14.5%. HPV16 (5.0%), HPV58 (2.2%), and HPV52 (1.5%) are the most common genotypes. Of the 21 genotypes investigated, high-risk HPV genotypes dominate in all age groups. HPV16 and HPV58 are the most common genotypes in patients with cervical lesions. HPV prevalence among Han women is 11.5% and the most common genotypes are HPV16 (4%) and HPV58 (2.1%). HPV prevalence is significantly higher in Mongolian women (32.6%), with the most common genotypes being HPV16 (10.7%), HPV31 (7.1%), and HPV52 (4.3%). The multiple infection rate in Mongolian participants (14.9%) is also higher than that of Han participants (4.3%). Urbanization, the number of sex partners, and PAP history appear as risk factors for HPV infection in Han, but not in Mongolian participants. HPV infection is highly prevalent in women in Inner Mongolia, China. HPV16 remains the most common genotype in this area. However, there are clear ethnical disparities in respect to the HPV epidemiology between the Han and Mongolian population.

**Keywords:** Han, human papillomavirus (HPV), Inner Mongolia, Mongolian, prevalence

## **Introduction**

Persistent infection with human papillomaviruses (HPVs) is a major cause of the development of cervical intraepithelial neoplasia and invasive cervical cancer, which is the fourth most common cancer in women around the world.<sup>1-3</sup> Globally, for 2008, it is estimated that there were at least 529 000 new cervical cancer patients and 274 000 attributed deaths.<sup>4</sup> Over 100 HPV genotypes have been identified and 40 are known to infect the genital tract.<sup>5</sup> Genotypes are classified as either low-risk HPV (LR) genotype or high-risk (HR) according to their oncogenic potential. Geographical differences exist in the prevalence of cervical cancer, but more than 85% of the cases are seen in developing countries. Parts of Africa, South Central Asia, and South America are considered as high-risk areas. China has been considered as an area with a relatively low risk for cervical cancer, but the HPV prevalence and genotype vary across the country.<sup>6</sup> Disease burden also varies according to the age, ethnicity, and residence area of women in China.<sup>7</sup>

The Inner Mongolia Minority Autonomous Region (Inner Mongolia), in northern China, founded in 1947, which was part of the ancient Silk Road region, is multi-ethnic of nature with a relatively large population. Since the launch of the Reform and Opening-up policy in China, now three decades ago, this area has experienced rapid industrialization and urbanization. By the year 2014, its population was in excess of 24 million people. Its top two ethnic constituents are the Han (82%) and Mongolian (17%), the latter being the main Mongolian population living in China. Other ethnic groups including Manchu, Hui, Daur, and Ewenki are also present in this region, although they only contribute relatively minor to the total population. One can easily envision that factors associated to ethnic identity, like migratory history and sexual etiquette, may contribute to the prevalence and genotype distribution of HPV infection, but this has not been conclusively established yet. Hence, in the present study we aimed to investigate the epidemiology of HPV infection in women in Inner Mongolia, China, with a focus on the Han and Mongolian ethnics.

## **Materials and methods**

### **Study population**

A survey was conducted by the Medical Molecular Genetics Laboratory of Inner Mongolia Maternal and Child Care Hospital from April 2014 to December 2015. A total of 5687 women (age range from 17 to 68 years) who visited the hospital for physical examination were invited to participate in this program and submitted written consent. Participants were excluded from the study according to following criteria: (1) under treatment for gynecological diseases or a history of vaccination against HPV or (2) unwilling to undergo HPV testing. The study was supervised by the local ethics committee of this hospital. Every participant was interrogated

using a standardized questionnaire to collect information about geographical location, education status, profession, marital status, the presence of multiple sex partners, and PAP history. The Hospital Ethics Review Board approved the study.

### **Specimen collection**

During the period of sample collection, which lasted from April 2014 to December 2015, about 250-320 participants were recruited in this screening program every month, while approximately 60-80 samples were tested every week. For most of the Mongolian participants (about 690 cases), samples and information were obtained from July 2014 to September 2015, with 40-54 cases every month. Samples from different ethnical backgrounds were evenly distributed each week. The interval time between sample collection and HPV testing was about 1-3 days.

Collection of HPV DNA specimens was carried out using plastic cervical swabs soaked with specimen transport medium (physiological saline) (Chaozhou HybriBio Biotechnology Corp., Chaozhou, China), as previously described.<sup>8</sup> After cervical material was collected, swabs were transferred to transport medium and stored at 4°C, if HPV testing was not available immediately. The cervical sample can be kept in physiological saline for 2 weeks at 4°C and 6 months at -20°C, according to the manufacturer's manual.

### **HPV DNA extraction**

After removal of the swabs and supernatant, cervical cells were isolated through centrifugation for 5 min employing a relative centrifugal force of 9660g. The DNA of the samples was then extracted using an alkali lysis and routine DNA extraction kits (Chaozhou HybriBio Biotechnology Corp).

### **HPV genotyping**

Following DNA isolation, HPV presence and genotype was determined using the HPV GenoArray test kit (PCR+film chip blot) (Chaozhou HybriBio Biotechnology Corp.). The assay detects 15 HR-HPV genotypes (HPV genotype 16, 18, 31, 33, 35, 39, 45, 51, 52, 53, 56, 58, 59, 66, and 68) and six LR-HPV genotypes (HPV genotype 6, 11, 42, 43, 44, and 81), using a MY09/11 primer detection system.<sup>9</sup> Briefly, HPV DNA template was extracted using a commercial Alkali lysis kit. PCR reactions involved 25 µL of fluid, including 24 µL of PCR Mix and 1 µL of DNA template. Thermo cycler conditions included an initial denaturation at 95°C for 5 min followed by 40 cycles of 95°C for 20 s, 55°C for 30 s, and 72°C for 30 s. The reaction was ended by an elongation of 72°C for 5 min. The size of the resulting PCR products is about 441 bp.<sup>9</sup>

The resulting 25 µL of PCR products was further examined by flow-through hybridization and gene chip blotting. The final results were directly visualized on a nylon membrane on which genotype-specific oligonucleotides were immobilized.

## **Pathological diagnosis**

Participants were referred to colposcopy if HPV testing indicated the presence of HR-risk HPV. Patients with HR-HPV infection would be advised to consent to histological evaluation of the biopsy samples obtained at colposcopy. The time interval between HPV testing and colposcopy was less than 3 months. Participants with normal or inflamed cervixes were classified as negative; while those with cervical intraepithelial neoplasia (CIN) were classified as CIN1, CIN2, CIN3, and cervical carcinoma as appropriate.

## **Derivation of age quintiles**

Age quintiles were computed from the whole population data. Specifically the age for each quintile for the whole population was as follows: quintile 1 = 17-28 years of age, quintile 2 = 28-33 years of age, quintile 3 = 33-39 years of age, quintile 4 = 39-45 years of age, and quintile 5 = 45-68 years of age.

## **Statistical analysis**

The presence of statistically significant differences in prevalence of HPV infections with HPV genotypes among general, Han and Mongolian groups was calculated using a Chi-squared test. The 95% confidence interval (CI) was established using a binomial distribution analysis. Odds ratios (ORs) for possible risk factors in different ethnic groups and corresponding 95% CIs were calculated by means of univariate and multivariate logistic regression. A P-value <0.05 was considered as statistically significant.

## **Results**

### **Overall prevalence of HPV infection and genotype distribution**

Of the 5687 women who provided cervical cell samples, five had mismatched questionnaire information and 27 were not analyzable because of the presence of blood cells or cervical secretes, leaving 5655 women with valid HPV results. The percentage of unanalyzable women is thus 0.56% (32/5687). No contamination during the HPV testing of cervical samples occurred.

Among the 5655 women evaluated, 820 participants (14.5%) were found to be HPV DNA positive, of which 244 showed multiple-genotype (4.3%) infections (Table 1). Of these, 785 women were designated as suffering from HR-HPV infection. Fifteen HR-HPV genotypes and six LR-HPV genotypes were all found during the course of this study, HPV16 being the most prevalent genotype (5.0%), followed by HPV58 (2.2%), HPV52 (1.5%), HPV18 (1.3%), and HPV31 (1.2%). HPV81 (CP8304) was the most common LR-HPV genotype (0.8%). Of note, LR-HPV42 and LR-HPV43 were not present as single-genotype infection.

### **HPV genotype distribution associated with cervical lesions**

The distribution of prevalent HPV genotypes according to histological diagnosis is presented in Table 2. From a total of 224 patients presenting with cervical lesions, HPV16 (38.8%) was the most common genotype in these patients, followed by HPV58 (14.3%), HPV18 (10.3%), HPV31 (10.3%), and HPV52 (8.5%). HPV16 was also the most prevalent genotype in patients with CIN1 (34.7%), followed by HPV58 (13.2%) and HPV31 (9.7%). For patients with CIN 2/3, HPV16 (44.6%) was the most common genotype, followed by HPV58 (17.6%) and HPV18 (16.2%). For patients with carcinoma, HPV16 (66.7%) was the most common genotype. LR-HPV42, LR-HPV43, and LR-HPV44 were not detected in patients with cervical lesions.

### **Prevalence of HPV infection and genotype distribution by age quintile**

The 5655 subjects of this study were divided into five groups by age quintile. The 820 HPV DNA-positive women were also classified based on age quintile. Figure 1 displays HPV prevalence by age quintiles. The HPV prevalence in age quintile 4 between 39 and 45 years (16.3%) was higher than that observed in other groups. Each age group was further divided into HPV16 or 18, other HR-HPV genotypes, and LR-HPV genotypes only. The prevalence of HPV16/18 was low (3.4%) among women in age quintile 1 between 17 and 28 years and progressively increased with age to 6.8% and 6.7% in age quintile 2 and quintile 5, respectively; while other HR-HPV genotypes were more equally distributed between the age groups, although still higher in age quintile 4 (9.8%). In LR-HPV infections, most LR-HPV genotypes (11 cases) were detected in age quintile 1.

### **Ethnic differences in prevalence of HPV infection and genotype distribution**

Most of the subjects were recruited from two major ethnic groups, including 4838 Han (85.6%) with an average age of  $39.8 \pm 2.0$  and 751 Mongolians (13.3%) with an average age of  $38.2 \pm 3.9$ . HPV prevalence in Mongolian women (32.6%) was significantly higher than that in Han women (11.5%) ( $P < 0.001$ ) (Table 1). The multiple-genotype HPV infection rate among Mongolian women (14.9%) was also significantly higher than that observed among Han women (2.7%) ( $P < 0.001$ ).

TABLE 1 HPV prevalence and genotype distribution among different ethnic groups

Variables	Total (n=5655)		Han (n=4838)		Mongolian (n=751)		P-value
	Cases	P% (95%CI)	Cases	P% (95%CI)	Cases	P% (95% CI)	
Positive	820	14.5 (13.58-15.42)	557	11.5 (10.60-12.40)	245	32.6 (29.25-35.95)	<0.001
Single infection	576	10.2 (9.41-10.99)	428	8.8 (8.00-9.60)	133	17.7 (14.97-20.43)	<0.001
Multiple Infection	244	4.3 (3.77-4.83)	129	2.7 (2.24-3.16)	112	14.9 (12.35-17.45)	<0.001
HR-HPV							
16	282	5 (4.43--5.57)	194	4 (3.45-4.55)	80	10.7 (8.49-12.91)	<0.001
18	72	1.3 (1.01--1.60)	54	1.1 (0.81-1.39)	16	2.1 (1.07--3.13)	0.022
31	66	1.2 (0.92-1.48)	13	0.3	53	7.1 (5.26-8.94)	<0.001
33	52	0.9	40	0.8	12	1.6 (0.70-2.5)	0.036
35	8	0.1	2	0.04	6	0.8	<0.001
39	64	1.1 (0.83-1.37)	41	0.8	23	3.1 (1.86-4.34)	<0.001
45	12	0.2	2	0.04	10	1.3 (0.49-2.11)	<0.001
51	64	1.1 (0.83-1.37)	56	1.2 (0.89-1.51)	6	0.8	0.687
52	82	1.5 (1.18-1.82)	46	1 (0.72-1.28)	32	4.3 (2.85-5.75)	<0.001
53	70	1.2 (0.83-1.37)	60	1.2 (0.89-1.51)	8	1.1 (0.35-1.85)	0.918
56	38	0.6	24	0.5	14	1.9 (0.56-2.24)	<0.001
58	125	2.2 (1.82-2.58)	102	2.1 (1.70-2.50)	21	2.8 (1.62-3.98)	0.243
59	27	0.5	16	0.3	9	1.2 (0.42-1.98)	0.005
66	51	0.9	35	0.7	16	2.1 (1.07-3.13)	<0.001
68	39	0.6	30	0.6	9	1.2 (0.42-1.98)	0.212
LR-HPV							
6	29	0.5	13	0.3	16	2.1 (1.07-3.13)	<0.001
11	25	0.4	21	0.4	4	0.5	0.708
42	4	0.07	3	0.06	1	0.1	0.498
43	2	0.04	2	0.04	-	-	
44	2	0.04	2	0.04	-	-	
81	44	0.8	18	0.4	26	3.5 (2.19-4.81)	<0.001

Bold type indicates statistically significant values. n, number of cases; P%, prevalence percentage. 95%CI: 95% confidence interval was obtained by using binomial distribution analysis model.

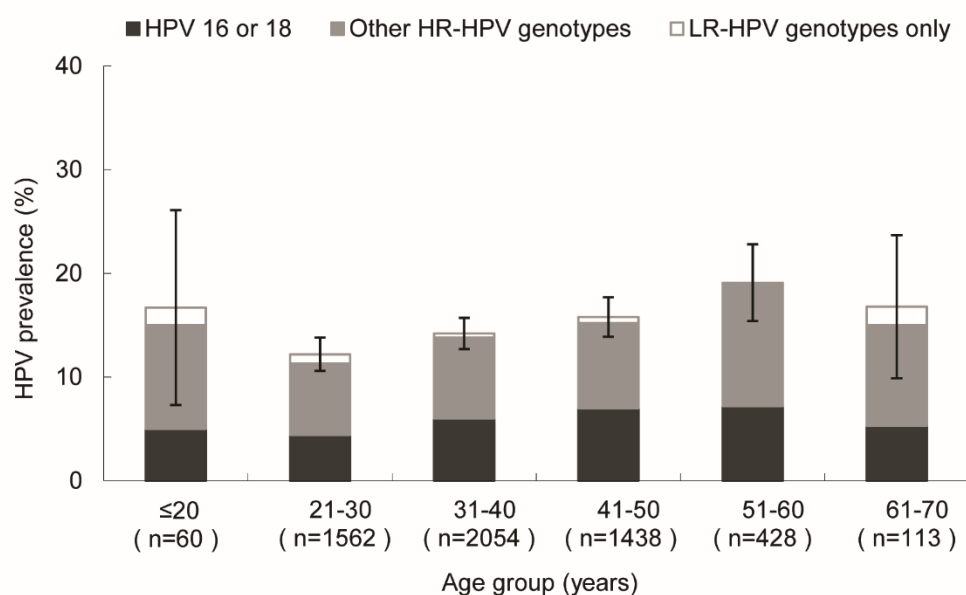
P-value: The HPV prevalence between Han and Mongolian groups was compared using Chi-squared test.



**TABLE 2 Prevalence of HPV genotypes with cervical intraepithelial neoplasias**

Genotype	Total (%) (n=224)		CIN I (n=144)		CIN 2/3 (n=74)		Carcinoma (n=6)	
	Single	Multiple	Single	Multiple	Single	Multiple	Single	Multiple
6	4 (1.8)	3	-	-	-	-	-	-
11	5 (2.2)	1	-	1	1 (1.4)	-	-	-
16	87 (38.8)	38	23	10	33 (44.6)	3	1	4 (66.7)
18	23 (10.3)	2	4	8	12 (16.2)	1	-	-
31	23 (10.3)	8	4	3	7 (9.5)	-	2	2 (33.3)
33	15 (6.7)	4	1	3	4 (5.4)	-	1	1 (16.7)
35	6 (2.7)	1	-	2	2 (2.7)	-	-	-
39	15 (6.7)	4	3	3	6 (8.1)	-	-	-
45	8 (3.6)	4	2	1	3 (4.1)	-	-	-
51	14 (6.3)	7	1	3	4 (5.4)	-	-	-
52	19 (8.5)	7	2	4	6 (8.1)	-	1	1 (16.7)
53	14 (6.3)	3	1	3	4 (5.4)	-	-	-
56	17 (7.6)	5	2	3	5 (6.7)	-	-	-
58	32 (14.3)	11	4	9	13 (17.6)	-	-	-
59	4 (1.8)	-	1	2	3 (4.1)	-	-	-
66	15 (6.7)	7	1	3	4 (5.4)	-	-	-
68	7 (3.1)	1	1	2	3 (4.1)	-	-	-
81 (CP8304)	12 (5.4)	2	1	1	2 (2.71)	-	-	-

n, number of cases.



**Fig. 1 Age-specific prevalence of HPV genotypes and corresponding 95% CIs.** 5655 participants were divided into six groups according to their age, namely younger than 20 years (60), 21-30 (1562), 31-40 (2054), 41-50 (1438), 51-60 (428) and 61-70 (113). 820 HPV DNA positive women were also classified based on their age. Each age group was further divided into HPV 16 or 18, other HR-HPV genotypes and LR-HPV genotypes.

A total of 21 genotypes were detected in Han women, listed in descending order, as HPV16, 58, 53, 51, 18, 52, 39, 33, 66, 68, 56, 11, 81, 59, 31, 6, 42, 35, 43, 35, 43, 44, and 45 (Table 1). Nineteen genotypes were identified among Mongolian women, listed in descending order, as HPV16, 31, 52, 81, 39, 58, 18, 66, 6, 56, 33, 45, 59, 68, 53, 35, 51, 11, and 42. The prevalence of HPV16 among Mongolian women was significantly (10.7%) higher than that observed among Han women (4.0%) ( $P < 0.001$ ). Moreover, HPV31 was also significantly more prevalent among Mongolian women (7.1%) than among Han women (0.3%) ( $P < 0.001$ ).

This study also recruited 38 Manchu women, 14 Hui women, 7 Daur women, and 7 Ewenki women, with an apparent HPV prevalence of 31.6% (12/38), 21.4% (3/14), 14.3% (1/7), and 28.6% (2/7), respectively.

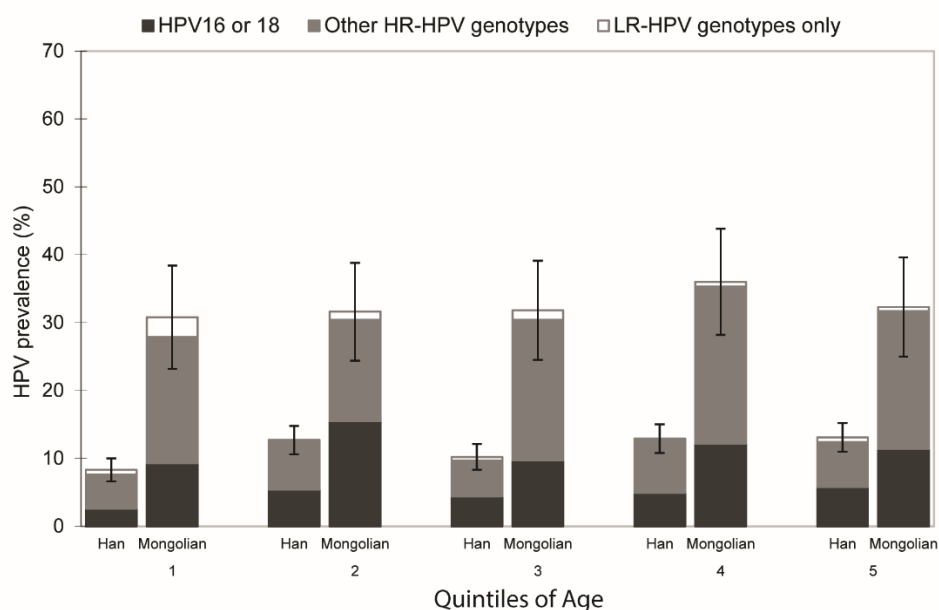
### Prevalence of HPV according to ethnic groups by age quintile

HPV prevalence among ethnic groups across age quintiles is displayed in Table 3 and Fig. 2. Among Han/Mongolian populations, there were 975/140, 952/161, 984/145, 971/147, and 956/158 participants for age quintile 1, 2, 3, 4, and 5, respectively. HPV prevalence in the Mongolian was significantly higher than that seen in the Han for all age quintiles (all  $P < 0.001$ ). In addition, prevalence of specific HPV genotypes was analyzed for ethnical origin as well. Statistically significant relationships were found between ethnical identity and prevalence for HPV16/18 (all  $P < 0.05$ ) and HR-HPV genotypes (all  $P < 0.001$ ) in each age quintile.

TABLE 3 Ethnic differences in HPV prevalence by age quintile

Groups	Han		Mongolian		Han vs Mongolian P-value
	Total	Positive (P%)	Total	Positive (P%)	
Age 1 (17-28)	1131	81 (8.3)	140	43 (30.7)	<b>&lt;0.001</b>
HPV16 or 18		25(2.6)		13 (9.3)	<b>&lt;0.001</b>
Other HR-HPV genotypes		49 (5.0)		26 (18.6)	<b>&lt;0.001</b>
LR-HPV genotypes		7 (0.7)		4 (2.9)	<b>0.019</b>
Age 2 (28-33)	1131	121(12.7)	161	51 (31.7)	<b>&lt;0.001</b>
HPV16 or 18		51 (5.4)		25 (15.5)	<b>&lt;0.001</b>
Other HR-HPV genotypes		67 (7.0)		24 (14.9)	<b>&lt;0.001</b>
LR-HPV genotypes		3 (0.3)		2 (1.2)	0.106
Age 3 (33-39)	1131	100 (10.2)	145	46 (31.7)	<b>&lt;0.001</b>
HPV16 or 18		43 (4.4)		14 (9.7)	<b>0.011</b>
Other HR-HPV genotypes		52 (5.3)		30 (20.7)	<b>&lt;0.001</b>
LR-HPV genotypes		5 (0.5)		2 (1.4)	0.216
Age 4 (39-45)	1131	126 (12.9)	147	53 (36.1)	<b>&lt;0.001</b>
HPV16 or 18		48 (4.9)		18 (12.2)	<b>0.001</b>
Other HR-HPV genotypes		75 (7.7)		34 (23.1)	<b>&lt;0.001</b>
LR-HPV genotypes		3 (0.3)		1 (0.7)	0.484
Age 5 (45-68)	1131	125 (13.1)	158	51 (32.3)	<b>&lt;0.001</b>
HPV16 or 18		55 (5.8)		18 (11.4)	<b>0.015</b>
Other HR-HPV genotypes		63 (6.6)		32 (20.3)	<b>&lt;0.001</b>
LR-HPV genotypes		7 (0.7)		1 (0.6)	0.892

Bold type indicates statistically significant values. n, number of cases; P%, prevalence percentage.



**Fig. 2** Ethnic differences in HPV prevalence by quintiles of ages and corresponding 95% CIs. 5655 participants were divided according to quintiles of age, and within each quintile the ethnic differences in HPV prevalence were analyzed. Age quintiles 1-5 are 17-28, 28-33, 33-39, 39-45 and 45-68 years old, respectively. Each age quintile was further divided into HPV 16 or 18, other HR-HPV genotypes and LR-HPV genotypes. The numbers of Han participants according to quintile (1-5) were: 975, 952, 984, 971 and 956. The numbers of Mongolian participants according to quintile were: 140, 161, 145, 147 and 158.

## Potential risk factors

Table 4 shows the association between HPV positivity and urbanization status, educational background, profession, marital status, the presence of multiple sex partners, and PAP history, after adjustment for age. HPV prevalence was significantly higher among women who had multiple sex partners (16.2%, 145/896) than those who had a single one (10.5%, 413/3942) (Table 4). Reporting of multiple lifetime sexual partners for Han women (16.2%, 145/896; OR = 1.5) was significantly associated with HPV positivity ( $P < 0.001$ ). This trend was also observed in Mongolian women, but did not reach statistical significance. HPV prevalence among Han women (17.7%, 197/1112) who had PAP history showed significantly higher prevalence than those who did not have such test (9.7%, 360/3726). Among Han women, those who had lived in rural areas (15.8%, 85/539) had a higher risk of being infected with HPV than those who had a lifetime history of living in urbanized areas (11.0%, 472/4299). Education, profession, and marital status were not significantly associated with HPV infection.

**TABLE 4 HPV Infection in Han and Mongolian groups according to residence areas, education, profession, marital status, the presence of multiple sex partners, and PAP history**

Variables	Han			Mongolian			Han vs Mongolian P-value
	Total	Positive	OR (95%CI)	Total	Positive	OR (95%CI)	
Areas			<b>0.004</b>			0.076	
Urban	4299	472	1	664	206	1	<0.001
Rural	539	85	1.4(1.12-1.84)	87	39	1.4(0.96-2.175)	<0.001
Education			0.36			0.662	
Junior high school or less	3120	351	1	484	151	1	<0.001
Senior high school	887	97	1(0.77-1.23)	129	43	1.1(0.723-1.579)	<0.001
College and higher	831	109	1.2(0.93-1.47)	138	51	1.2(0.82-1.71)	<0.001
Professions			0.147			0.485	
Job	4415	498	1	712	235	1	<0.001
No job	423	59	1.2(0.93-1.65)	39	10	0.8(0.38-1.58)	0.105
Marital status			0.072			0.155	
Married	3996	412	1	680	212	1	<0.001
Single	800	142	1.7(1.40-2.11)	70	33	1.5(0.97-2.35)	<0.001
Divorced or widowed	42	3	0.7(0.21-2.25)	1	0	0	-
Multiple sex partners			<0.001			0.694	
No	3942	413	1	528	169	1	<0.001
Yes	896	145	1.5(1.26-1.89)	223	76	1.1(0.78-1.46)	<0.001
pap history			<0.001			0.06	
No	3726	360	1	595	180	1	<0.001
Yes	1112	197	1.8(1.52-2.20)	156	65	1.4(0.986-1.923)	<0.001

Bold type indicates statistically significant values.

OR: odds ratios was obtained using multivariate logistic regression analysis model.

P-value: The HPV prevalence between Han and Mongolian groups was compared using Chi-squared test.

## Discussion

In this study, 5655 women from Inner Mongolia were recruited in a HPV screening program, and 820 (14.5%) participants were established to be HPV positive. The observed prevalence is higher than that reported for Beijing (6.7%) but lower than that reported in Gansu (19.9%) and Xinjiang (19.7%),<sup>10–12</sup> conforming a tendency for HPV infection to increase from the Northeast to the Northwest of China. The frequency of multiple-HPV-strain infection displays an opposite trend. Multiple infection accounted for 29.8% (244/820) of all infections in the present study, which is higher than that in Gansu (18.18%) but lower than that in Beijing (38.3%) and Shenyang (31.3%)<sup>10–13</sup>. We have observed in our population three major HPV genotypes, HPV16 (5.0%), HPV58 (2.2%), and HPV52 (1.5%). Other prevalent genotypes were HPV18 (1.3%), HPV53 (1.2%), and HPV31 (1.2%). As previous studies reported,<sup>13–15</sup> HPV16, HPV58, and HPV52 are the three most common HPV genotypes in China but their relative order is HPV52, HPV16, and HPV58. In Hong Kong,<sup>16</sup> the same order as in our study is observed, but the reported frequencies differ. HPV18 is prevalent in our study. It is also the most common genotype in Korea and the second most common genotype in Western countries.<sup>14,17</sup>

Relating these results to histological diagnosis, HPV16 (38.8%) emerges as the most common genotype in CIN1, CIN2, CIN3, and carcinoma. In general, its prevalence increases with the severity of the cervical lesions (Table 1). The second most prevalent in patients with CIN1 is HPV58 (13.19%) and CIN2/3 (17.56%), and HPV31 (33.33%) in carcinoma. It thus seems that HPV52 and HPV58 are highly associated to high degree CIN in North China, which is in apparent disagreement with an earlier report showing that HPV58 and HPV52 are only prevalent in the South and Southwest China.<sup>18</sup> Although HPV58 was not detected in patients with carcinoma, we speculate that this may be due to the limited number of cervical cancer samples encountered. A previous worldwide study reported that, in Asia, the prevalence of HPV58 is 18.1% in CIN2, 18.0% in CIN3, and 7.9% in invasive cervical cancer, values which were all significantly higher than those reported in other continents.<sup>19</sup> It has been reported that specific variants on E7 of HPV58 were more common in Asia and more tightly associated with high-grade cervical intraepithelial neoplasia.<sup>20</sup> These reports strengthened the potential role of HPV58 as a causative agent in cervical cancer. Therefore, we recommend that the prophylactic HR-HPV vaccines that are currently based on neutralizing only HPV16/18 genotypes should be modified to further include HPV58 when used in Asian populations.

The lowest HPV prevalence occurred among participants in age quintile 1 in this study (Fig. 1). No decrease was seen with increasing age, and the highest HPV prevalence is among women in age quintile 4. This tendency was similar to that reported in recent studies in areas of China and in rural India.<sup>21,22</sup> It is suggested that middle-aged women infected with HR-HPV infection are more susceptible to high-degree CIN or worse, as young women are more capable of clearing HPV as a consequence of much stronger immunity. This notion is supported by the observation of a peak in apparent HPV16 infection rates in age quintile 4, which is also the age at which the incidence of high-grade cervical lesions and carcinoma starts to rise.

Han and Mongolian are the two major ethnic groups in Inner Mongolia, which borders the south of Mongolia proper. The HPV prevalence in Han women (11.5%) is slightly higher than that reported for Beijing and is similar to the global infection rate (11-12%).<sup>23</sup> Surprisingly, HPV prevalence is particularly high among Mongolian women (32.6%) in this region, but comparable to the prevalence in Mongolian women (35.0%) in Ulaanbaatar,<sup>24</sup> the capital of Mongolia. Furthermore, HPV58 and 52 are the most common genotypes in Northern China, but among Mongolian women, HPV31 is the second most common genotype and its prevalence in this ethnic group is significantly higher than that among Han women. Consistently, this genotype is also highly prevalent among women in Ulaanbaatar. Another similarity that we observe is the multiple-genotype infection is higher among Mongolian women and amounts to 14.9%, which is comparable to the rate reported for Ulaanbaatar (14.8%). These results show a clear association between ethnicity with HPV prevalence.

Age is a variable of great importance when assessing ethnic differences in HPV prevalence in this area. Interestingly, the age-specific prevalence in Mongolian presents a similar trend as that of the whole population (Figs. 1 and 2), although the prevalence rates differ greatly. In addition to age, other risk factors that significantly increase the odds of HPV infection among Han women include the area of living, the presence of sexual partners, and PAP history. A pooled analysis of 17 population-based studies in China also demonstrated a significantly different prevalence of HR-HPV in women from rural (18.0%) and urban (15.2%) areas<sup>25</sup>. Similar to previous studies,<sup>26,27</sup> we have demonstrated that the presence of multiple sexual partners of the participants is an important factor for HPV transmission. The association between HPV infection and a history of a previous PAP smear suggests that women requesting such a test consider themselves more at risk for the presence of a sexually transmitted disease and suffer more HPV infection. Because most of the participants did not provide sufficient information regarding the sexual behavior of their husbands, the role of male sexual behavior was possibly underestimated in our study. In contrast, these potential risk factors do not show a significant impact on the Mongolian women (may be due to a smaller sample size), although similar patterns were observed as seen in Han women. This may further strengthen the importance of ethnic identity for the presence of HPV infection.

Inner Mongolia, geologically located at the central part of Asia, bordering with Mongolia and Russia, is one of the five Autonomous Regions of China. It serves as the bridge for cultural communication between the eastern and western world. As a second-largest ethnic group, Mongolians experienced their heydays during the Yuan Dynasty, but today the separate cultural identity of this group is in danger, in particular through miscegenation of the people with different ethnicities (including Han, Manchu, Hui, Daur, Ewenki, and Korean) in this region. This may lead to frequent cultural and genetic recombination or exchange. HPV genotype distribution varies largely across the globe, which can at least partially be attributed to differences in genetics and lifestyles. Interestingly, the DRB\*1501 allele of the HLA class II gene has been reported to be associated with high susceptibility to HPV infection in the Han population of Inner Mongolia, but strikingly has also been related to protection from cervical

cancer in the Han of Xinjiang<sup>28,29</sup>. As Xinjiang is like Inner Mongolia an important autonomous region but mainly composed Uygur, Kazak, and Hui ethnic groups, we postulate that both living environment and genetics are important factors with respect to the prevalence and genotype distribution of HPV infection.

In summary, this study has characterized the prevalence and genotype distribution of HPV infection in women from Inner Mongolia, China. HPV16, HPV58, and HPV52 are the most prevalent genotypes in this area. Importantly, we have observed ethnic disparity of HPV infection. Compared to Han population, HPV31 and multiple-genotype infection are more common among Mongolian women. These epidemiological data will help to establish effective strategies for prevention and management of HPV infection in this area.

## References

1. Bosch FX, Lorincz A, Munoz N, Meijer CJ, Shah KV. The causal relation between human papillomavirus and cervical cancer. *Clin Pathol.* 2002; 55: 244–265.
2. Cuzick J, Szarewski A, Cubie H, et al. Management of women who test positive for high-risk types of human papillomavirus: the HART study. *Lancet.* 2003; 362: 1871–1876.
3. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA Cancer J Clin.* 2015; 65: 87–108.
4. Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int J Cancer.* 2010; 127: 2893–2917.
5. Munoz N, Bosch FX, de Sanjose S, et al. Epidemiologic classification of human papillomavirus types associated with cervical cancer. *N Engl J Med.* 2003; 348: 518–527.
6. Yang L, Huangpu XM, Zhang SW, et al. Changes of mortality rate for cervical cancer during 1970's and 1990's periods in China. *Acta Acad Med Sin.* 2003; 25: 386–390.
7. Li J, Kang LN, Qiao YL. Review of the cervical cancer disease burden in mainland China. *Asian Pac J Cancer Prev.* 2011; 12: 1149–1153.
8. Lin M, Yang LY, Li LJ, Wu JR, Peng YP, Luo ZY. Genital human papillomavirus screening by gene chip in Chinese women of Guangdong province. *Aust N Z J Obstet Gynaecol.* 2008; 48: 189–194.
9. Bauer HM, Greer CE, Manos MM. Determination of genital human papillomavirus infection using consensus PCR. In: Herrington CS, McGee JOD, eds. *Diagnostic Molecular Pathology: A Practical Approach.* Oxford, United Kingdom: Oxford University Press; 1992:132–152.
10. Zhao R, Zhang WY, Wu MH, et al. Human papillomavirus infection in Beijing, People's Republic of China: a population-based study. *Br J Cancer.* 2009; 101: 1635–1640.
11. Du H, Suo LC, Liu HX. Current situation of cervical HPV infection in females in Gansu Province. *Pract Prev Med.* 2015; 22: 17–19.
12. Chen ZF, Ma HFZ, Ding Y. The investigation of reproductive tract human papillomavirus infection genotypes and distribution in Hans, Uygurs and Hazakhs. *Chin J Pract Gynecol Obstetrics.* 2011; 27: 602–605.
13. Li LK, Dai M, Clifford GM, et al. Human papillomavirus infection in Shenyang City, People's Republic of China: a population-based study. *Br J Cancer.* 2006; 95: 1593–1597.
14. Clifford GM, Smith JS, Plummer M, Munoz N, Franceschi S. Human papillomavirus types in invasive cervical cancer worldwide: a meta-analysis. *Br J Cancer.* 2003; 88: 63–73.
15. Lu S, Cong X, Li M, Chang FX, Ma L, Cao YT. Distribution of high-risk human papillomavirus genotypes in HPV-infected women in Beijing, China. *J Med Virol.* 2015; 87: 504–507.
16. Chan PK, Li WH, Chan MY, Ma WL, Cheung JL, Cheng AF. High prevalence of human papillomavirus type 58 in Chinese women with cervical cancer and precancerous lesions. *J Med Virol.* 1999; 59: 232–238.



17. Quek SC, Lim BK, Domingo E, et al. Human papillomavirus type distribution in invasive cervical cancer and high-grade cervical intraepithelial neoplasia across five countries in Asia. *Int J Gynecol Cancer*. 2013; 23: 148–156.
18. Chen W, Zhang X, Molijn A, et al. Human papillomavirus type-distribution in cervical cancer in China: the importance of HPV 16 and 18. *Cancer Causes Control*. 2009; 20: 1705–1713.
19. Chan PK, Ho WC, Chan MC, et al. Meta-analysis on prevalence and attribution of human papillomavirus types 52 and 58 in cervical neoplasia worldwide. *PLoS ONE*. 2014;9: e107573.
20. Chan PK, Zhang C, Park JS, et al. Geographical distribution and oncogenic risk association of human papillomavirus type 58 E6 and E7 sequence variations. *Int J Cancer*. 2013; 132: 2528–2536.
21. Dai M, Bao YP, Li N, et al. Human papillomavirus infection in Shanxi Province, People's Republic of China: a population-based study. *Br J Cancer*. 2006; 95: 96–101.
22. Franceschi S, Rajkumar R, Snijders PJF, et al. Papillomavirus infection in rural women in southern India. *Br J Cancer*. 2005; 92: 601–606.
23. Bouvard V, Baan R, Straif K, et al. A review of human carcinogens—Part B: biological agents. *Lancet Oncol*. 2009; 10: 321–322.
24. Dondog B, Clifford G M, Vaccarella S, et al. Human papillomavirus infection in Ulaanbaatar, Mongolia: a population-based study. *Cancer Epidemiol Biomarkers Prev*. 2008; 17: 1731–1738.
25. Zhao FH, Lewkowitz AK, Hu SY, et al. Prevalence of human papillomavirus and cervical intraepithelial neoplasia in China: a pooled analysis of 17 population-based studies. *Int J Cancer*. 2012; 131: 2929–2938.
26. Krishnakumar V, Santhanam S, Das BC, Kalimuthusamy N. Prevalence and risk factors of HPV infection among women from various provinces of the world. *Arch Gynecol Obstet*. 2012; 285: 771–777.
27. Li J, Huang R, Schmidt JE, Qiao YL. Epidemiological features of human papillomavirus (HPV) infection among women living in Mainland China. *Asian Pac J Cancer Prev*. 2013; 14: 4015–4023.
28. Xie J, Wang L, Wan G, Ma X, Qiao H. Study on the association between HPV16-positive cervical cancer and polymorphism of HLA-DRB1, HLA-DQB1 in Inner Mongolian women. *J Report Med*. 2007; 16: 91–95.
29. Sun Q, Qi C, Yang A, et al. A study on the correlations of human leucocyte antigens-DRB1\* 1501, DQB1\* 0602 alleles with HPV16 infection and invasive squamous cell carcinoma of cervix in Xinjiang Uigur and Han women. *J Shihezi Univ*. 2009; 27: 133–138.



# Chapter 3

## **The Burden of Human Papillomavirus and *Chlamydia* *Trachomatis* Coinfection in Women: A Large Cohort Study in Inner Mongolia, China**

Yunpeng Ji, Xiaoxia Ma, Zhaocai Li, Maikel P. Peppelenbosch, Zhongren Ma, Qiuwei Pan

The Journal of Infectious Diseases. 2019, 219(2):206-214



## Abstract

**Background.** *Chlamydia trachomatis* may coinfect with human papillomavirus (HPV) and complicate the cervical pathogenesis. This study aimed to evaluate the prevalence, risk factors, and clinical outcomes of HPV/*C. trachomatis* coinfection in women from Inner Mongolia, China. **Methods.** We performed a polymerase chain reaction (PCR)-based HPV/*C. trachomatis* screening and cervical samples were analyzed by thinprep cytologic test. Statistical analysis was used to assess the association between demographic factors and coinfection. **Results.** Among the 2345 women recruited, the prevalences of HPV, *C. trachomatis*, and HPV/*C. trachomatis* coinfection were 36.0%, 14.3%, and 4.8%, respectively. The rate of multiple HPV genotypes was higher in coinfecting women. HPV66 was the most frequently identified genotype in coinfecting participants. The HPV DNA load was significantly higher in HPV mono infected cases. In contrast, the DNA load of *C. trachomatis* was significantly higher in the coinfection group. Risk factors, including single women (odds ratio [OR] = 6.0, 95% confidence interval [CI], 4.044–8.782) and women with multiple sex partners (OR = 1.9, 95% CI, 1.324–2.824), were associated with coinfection. Importantly, coinfection was associated with increased risk for high-grade squamous intraepithelial lesions. **Conclusions.** HPV and *C. trachomatis* coinfection is an important risk factor for the progression of cervical lesions.

**Keywords:** human papillomavirus, *Chlamydia trachomatis*, coinfection, risk factors, cytological findings, DNA load

## Introduction

Cervical cancer is the fourth most common cancer in women around the world<sup>[1]</sup>. It is estimated that there were at least 528 000 new cervical cancer patients and 266 000 attributed deaths in 2012<sup>[2]</sup>. Geographical differences exist in the prevalence of cervical cancer and more than 85% of the cases occur in developing countries. Certain parts of Africa, South Central Asia, and South America are considered to be high-risk areas. In China, screening for cervical cancer is an important public health program launched by the government to meet the gap in health condition between its west and east, where large differences in cervical cancer burdens are present<sup>[3]</sup>.

Infection with human papillomaviruses (HPV) is a major cause of the development of cervical cancer. The medical burden of HPV infection varies according to several demographic factors, such as age, ethnicity, education background, and lifestyle<sup>[4]</sup>. Mechanistically, particular genotypes of HPV express oncoproteins E6 and E7, which may cause cell immortalization and transformation. Furthermore, other factors are involved in HPV-related carcinogenesis, including distribution of HPV genotypes, duration of the infection, viral load, and the presence of other coagents<sup>[5]</sup>.

Several sexually transmitted infectious agents, including *Chlamydia trachomatis*, *Gardnerella vaginalis*, *Trichomonas vaginalis*, and *Mollicutes*, are known to cause local inflammation and contribute to HPV-related cervical lesion progression<sup>[6]</sup>. *C. trachomatis* is a widespread sexually transmitted bacterium and can cause diseases including cervicitis, salpingitis, urethritis, endometritis, pelvic inflammatory disease, tubal factor infertility, and ectopic pregnancy. Although the infection is mainly asymptomatic in most patients<sup>[7,8]</sup>, active pathogen-host interactions, including cell entry and disruption of host immune and metabolic elements, may affect cellular processes involved in carcinogenesis<sup>[8]</sup>.

The Inner Mongolian Autonomous Region is 1 of the 5 multiethnic population areas in China. It is a less-developed area with the main ethnic groups including Han, Mongolian, Manchu, Hui, Daur, and Ewenki. Migration and intermarriage of ethnic groups has increased the abundance of the local genetic pool and reduced lifestyle differences. Therefore, this region is ideal for studying diseases influenced by multiple socioeconomic and genetic factors. The aim of this study was to investigate the prevalence, risk factors, and clinical outcomes of HPV/*C. trachomatis* coinfection in a hospital-based large cohort in this area.

## Methods

### Study Design

A study of coinfection between HPV and *C. trachomatis* was conducted in Inner Mongolia Maternal and Child Care Hospital from January 2015 to September 2017. A population of 2359 women (age range from 19 to 83 years) who visited the department of gynecology of the hospital was invited to participate in this program and submit written consent. A detailed questionnaire was administered to all participants to document their lifestyle pattern and demographic information, including education background, number of lifetime sex partners, residence areas, age, marital status, ethnicity, and profession. Besides detection of pathogens, cervical samples of all participants were also analyzed by thinprep cytologic test (TCT). Participants under medical treatment or with a history of vaccination against HPV were excluded from the study.

### Sample Collection

Cervical samples were taken by swab and cytobrush for detection of pathogen DNA and cytological tests, respectively. Swabs were transferred to a tube containing a transport medium (physiological saline; Chaozhou HybriBio Biotechnology Corp, China). The interval between sampling and testing was within 3 days (cervical samples can be kept in physiological saline for 2 weeks at 4°C according to the manufacturer's manual).

### DNA Extraction

After remove of the swab, cervical cells were collected by centrifugation for 5 minutes with relative centrifugal force 9660g. DNA samples were extracted with alkali lysis using a DNA extraction kit (Chaozhou HybriBio Limited Corp, Chaozhou, China). DNA extracted was quantified by Thermo Scientific NanoDrop 2000 for use in analysis of HPV genotypes and quantification of genomic DNA of the 2 pathogens.

### HPV Genotyping

The method of HPV genotyping was described in our previous study<sup>[9]</sup>. Briefly, a HPV GenoArray test kit (polymerase chain reaction [PCR] plus film chip blot) (Chaozhou HybriBio Limited Corp, Chaozhou, China) was used to detect the 15 high-risk (HR)-HPV genotypes (HPV16, 18, 31, 33, 35, 39, 45, 51, 52, 53, 56, 58, 59, 66, and 68) and 6 low-risk (LR)-HPV genotypes (HPV6, 11, 42, 43, 44, and 81) with the MY09/11 primer detection system<sup>[10]</sup>. The PCR reaction system was 25 µL, comprising 24 µL PCR mix and 1 µL DNA template (50 ng/µL). Thermocycler conditions included an initial denaturation at 95°C for 5 minutes followed by 40

cycles of 95°C for 20 seconds, 55°C for 30 seconds, and 72°C for 30 seconds. The reaction was ended with an elongation of 72°C for 5 minutes. The size of the PCR products was about 441bp<sup>[10]</sup>.

The 25- $\mu$ L PCR products were examined with flow-through hybridization and gene chip blotting. The final results were directly visualized on a nylon membrane on which genotype-specific oligonucleotides were immobilized.

### qRT-PCR Analysis of Pathogen Genomic DNA

We detected and quantified genomic DNA of the 2 pathogens by a quantitative real-time PCR-based (qRT-PCR) method. A commercial multiplex real-time 12+2 HPV test kit (Chaozhou HybriBio Limited Corp, Chaozhou, China) was used for the quantitative detection of HPV DNA, which specifically targets 14 HR-HPV genotypes (HPV16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, and 68). Ten-fold serial dilutions (10<sup>-7</sup> copies/mL) of the DNA standard curve were prepared and analyzed. The qRT-PCR test was performed in an Applied Biosystems 7500 real-time PCR system (Applied Biosystems, Foster City, CA), according to the manufacturer's instruction. Briefly, 14 sets of primers and fluorescent probes provided by the manufacturer in a PCR mix reagent were used to target the E region of each HR-HPV genotype and detect the amplification products (size: 70–200 bp)<sup>[11]</sup>. Probes and primers were classified into 4 groups, namely, group 1 for general detection of the 12 non-HPV16/HPV18 HR-HPV genotypes (HPV31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, and 68), group 2 for HPV 16, group 3 for HPV 18, and group 4 for a  $\beta$ -globin internal control. Probes were labeled with fluorophores: hexachlorofluorescein for HPV16, carboxy-X-rhodamine for HPV18, 6-carboxy-fluorescein (FAM) for 12 non-HPV16/HPV18 HR-HPV genotypes, and Cy5 for  $\beta$ -globin. The PCR reaction system was 20  $\mu$ L, comprising 18  $\mu$ L PCR mix and 2  $\mu$ L DNA template (50 ng/ $\mu$ L). The reaction conditions for PCR were 95°C for 10 minutes, followed by 45 cycles of 95°C for 10 seconds and 60°C for 60 seconds. The fluorescence signal was detected at 60°C.

A *C. trachomatis* real-time PCR kit (Shanghai Liferiver Limited Corp, Shanghai, China) was used for detection of *C. trachomatis* genomic DNA. The PCR reaction system was 40  $\mu$ L, comprising 36  $\mu$ L PCR mix and 4  $\mu$ L DNA templates (50 ng/ $\mu$ L). Primers and probe, designed according to the sequences of the cryptic plasmid of 5 different *C. trachomatis* strains (serotypes A, B, D, L1, and L2) and stored in the PCR mix, were provided by the manufacturer<sup>[12]</sup>. This included a forward primer (5'-CATGAAAACCTCGTTCCGAAAT-AGAA-3'), a reverse primer (5'-TCAGAGCTTTACCTAACAACGCATA-3') (which amplifies a 71-bp DNA segment of *C. trachomatis*), and a minor-groove binder probe labeled with 5' FAM (5'-TCGCATGCAAGATATCGA-3'). The PCR test was also performed in the Applied Biosystems 7500 real-time PCR system and the cycling conditions were 2 minutes at 37°C, 2 minutes at 94°C,



followed by 40 cycles of 15 seconds at 93°C and 1 minute at 60°C. Signal detection was at 60°C. In order to perform quantification of *C. trachomatis* genomic DNA, a standard curve was generated with serial 10-fold dilutions of known quantities (10–7 copy/mL) of a commercial positive control (Bio-Rad Amplichek CT/GC Controls).

## **Cytology**

Slides for liquid-based cytology were prepared and stained according to manufacturer's instructions at the pathology laboratory of the hospital. The results were reported according to the Bethesda System 2001<sup>[13]</sup>. The following cytological findings were reported: negative for intraepithelial lesion or malignancy (NILM); atypical glandular cells (AGC); squamous intraepithelial lesions (SIL) of low (LSIL) or high (HSIL) grade; atypical squamous cells (ASC) of undetermined significance (ASC-US) or not possible exclude HSIL (ASC-H); and SCC (squamous cell carcinoma).

## **Statistics Analysis**

Statistical analyses were performed with SPSS version 13.0 (Chicago, IL); *t* test and Mann-Whitney U test was used to compare pathogen DNA copy numbers between single and coinfection groups. The difference in odds ratios (ORs) for possible risk factors associated with pathogenic infection and corresponding 95% confidence intervals (CIs) were calculated by means of univariate and multivariate logistic regression. Significance was set at  $P < .05$ . Age quintiles were computed from the whole population data.

## **Ethical Consideration**

The study was approved and supervised by the local ethics committee of Inner Mongolia Maternal and Child Care Hospital.

## **Results**

### **Overall Prevalence and Genotype Distribution of HPV Infection**

Of the 2359 women who provided cervical cell samples, 2 had mismatched questionnaire information and 12 samples had too many blood cells or excessive cervical secretion that affected DNA extraction, leaving 2345 women with valid testing for pathogen DNA. The percentage of invalid participants for testing was 0.59% (14/2359). No contamination occurred during testing of the samples.

Analysis by HPV genotyping and HPV real-time PCR gave 839 and 851 positive results, respectively. Thirty-two samples showed discordant results with the 2 different approaches and were further evaluated by sequencing. The sequencing results confirmed that 22 samples were positive (including 8 cases of HPV51/53, 6 cases of HPV16/52, 2 cases of HPV16/33/31, 2 case of HPV16, 2 case of HPV31, 1 case of HPV52, and 1 case of HPV68); the other 10 samples were negative. Thus the number of HPV-positive results was 844 after sequencing confirmation.

Therefore, among the 2345 women, 844 participants (36.0%) were found to be HPV DNA positive, of which 301 showed multiple-genotype (12.8%) infections (Table 1). Fifteen HR-HPV genotypes and 6 LR-HPV genotypes were detected. HR-HPV infection was confirmed in 824 women. Specifically, HPV16 was the most prevalent genotype (10.4%), followed by HPV58 (6.4%), HPV52 (6.2%), HPV51 (4.2%), and HPV39 (3.3%), while HPV81 (CP8304) was the most common LR-HPV genotype (2.7%).

**Table 1. The distribution of HPV genotypes and C. trachomatis prevalence by cytology status**

HPV and C. trachomatis	Cytological findings			≥ASC-us OR(95%CI)	P-value	Cytological findings with ≥ASC-US							HSIL	P-value
	Overall (n=2345) n (%)	NILM (n=1755) n (%)	≥ASC-us (n=590) n (%)			AGC (n=6) n (%)	ASC-US (n=256) n (%)	ASC-H (n=19) n (%)	LSIL (n=170) n (%)	HSIL (n=136) n (%)	SCC (n=3) n (%)	OR(95%CI)		
C. trachomatis	335 (14.3)	166 (7.1)	169 (7.2)	0 (0)	77 (3.3)	1 (0.04)	52 (2.2)	36 (1.7)	3 (0.1)	2.8(1.88-4.18)	<0.001			
HPV DNA	844 (36.0)	413 (17.6)	431 (18.4)	2 (0.1)	140 (6.0)	15 (0.6)	143 (6.1)	128 (5.5)	3 (0.1)	4.0(3.07-5.21)	<0.001			
Multiple infection	301 (12.8)	125 (5.3)	176 (7.5)	-	70 (3.0)	5 (0.2)	64 (2.7)	36 (1.5)	1 (0.04)	3.7(2.47-5.60)	<0.001			
Single infection	543 (23.2)	289 (12.3)	254 (10.8)	2 (0.1)	70 (3.0)	9 (0.4)	79 (3.4)	92 (3.9)	2 (0.09)	4.1(3.07-5.50)	<0.001			
HR-HPV														
HPV-16	245 (10.4)	100 (4.3)	145 (6.2)	-	38 (1.6)	6 (0.3)	30 (1.3)	69 (2.9)	2 (0.09)	8.9(6.26-12.7)	<0.001			
HPV-18	52 (2.2)	17 (0.7)	35 (1.5)	-	12 (0.5)	-	14 (0.6)	9 (0.4)	-	6.8(3.0-15.6)	<0.001			
HPV-31	64 (2.7)	31 (1.3)	33 (1.4)	-	13 (0.6)	1 (0.04)	10 (0.4)	9 (0.4)	-	3.7(1.75-8.03)	<0.001			
HPV-33	57 (2.4)	19 (0.8)	38 (1.6)	-	9 (0.4)	2 (0.09)	12 (0.5)	15 (0.6)	-	10.2(5.06-20.5)	<0.001			
HPV-35	21 (0.9)	6 (0.3)	15 (0.6)	-	8 (0.3)	1 (0.04)	3 (0.1)	3 (0.1)	-	6.5(1.60-26.1)	0.003			
HPV-39	77 (3.3)	34 (1.4)	41 (1.7)	-	12 (0.5)	2 (0.08)	17 (0.7)	10 (0.4)	-	3.8(1.84-7.85)	<0.001			
HPV-45	13 (0.6)	3 (0.1)	10 (0.4)	-	4 (0.2)	-	4 (0.2)	2 (0.09)	-	8.6(1.43-51.9)	0.005			
HPV-51	99 (4.2)	62 (2.6)	37 (1.6)	-	14 (0.6)	1 (0.04)	17 (0.7)	5 (0.2)	-	1.0(0.41-2.63)	0.933			
HPV-52	147 (6.3)	72 (3.1)	75 (3.2)	-	25 (1.0)	5 (0.2)	27 (1.2)	18 (0.8)	-	3.2(1.87-5.57)	<0.001			
HPV-53	59 (2.5)	33 (1.4)	26 (1.1)	-	1 (0.04)	-	16 (0.7)	8 (0.3)	1 (0.04)	3.1(1.42-6.91)	0.003			
HPV-56	56 (2.4)	27 (1.2)	29 (1.2)	-	8 (0.3)	-	15 (0.6)	6 (0.3)	-	2.9(1.16-7.07)	0.017			
HPV-58	150 (6.4)	95 (4.1)	65 (2.8)	1 (0.04)	19 (0.8)	2 (0.09)	25 (1.1)	18 (0.8)	-	2.4(1.44-4.17)	0.001			
HPV-59	28 (1.2)	25 (1.1)	3 (0.1)	-	0 (0)	-	3 (0.1)	-	-	-	-			
HPV-66	61 (2.6)	26 (1.1)	35 (1.5)	-	8 (0.3)	-	19 (0.8)	8 (0.3)	-	4.0(1.76-8.94)	<0.001			
HPV-68	46 (2.0)	21 (0.9)	25 (1.1)	-	3 (0.1)	1 (0.04)	7 (0.3)	4 (0.2)	-	2.5(0.83-7.26)	0.093			
LR-HPV														
HPV-6	28 (1.2)	21 (0.9)	7 (0.3)	-	2 (0.09)	-	5 (0.2)	-	-	-	-			
HPV-11	21 (0.9)	13 (0.6)	8 (0.3)	-	2 (0.09)	-	5 (0.2)	1 (0.04)	-	1.0(0.13-7.65)	0.994			
HPV-42	6 (0.3)	2 (0.09)	4 (0.2)	-	0 (0)	-	4 (0.2)	-	-	-	-			
HPV-43	2 (0.09)	1 (0.04)	1 (0.04)	-	1 (0.04)	-	-	-	-	-	-			
HPV-44	9 (0.4)	4 (0.2)	5 (0.2)	-	2 (0.09)	1 (0.04)	1 (0.04)	1 (0.04)	-	3.2(0.36-29.1)	0.269			
HPV-81	64 (2.7)	38 (1.6)	26 (1.1)	1	8 (0.3)	-	13 (0.6)	3 (0.1)	1 (0.0)	1.0(0.31-3.34)	0.976			

The 95% confidence interval (CI) was obtained by using binomial distribution analysis model. P value: the HPV prevalence between NILM and abnormal cytological findings was compared using X2 test. Bold type indicates statistically significant values.

Abbreviations: %, prevalence percentage; AGC, atypical glandular cells; ASC-H, atypical squamous cells not possible exclude HSIL; ASC-US, atypical squamous cells of undetermined significance; HR-HPV, high-risk human papillomavirus; HSIL, high-grade squamous intraepithelial lesion; LR-HPV, low-risk human papillomavirus; LSIL, low-grade squamous intraepithelial lesion; n, number of cases; NILM, negative for intraepithelial lesion or malignancy; OR, odds ratio; SCC, squamous cell carcinoma.

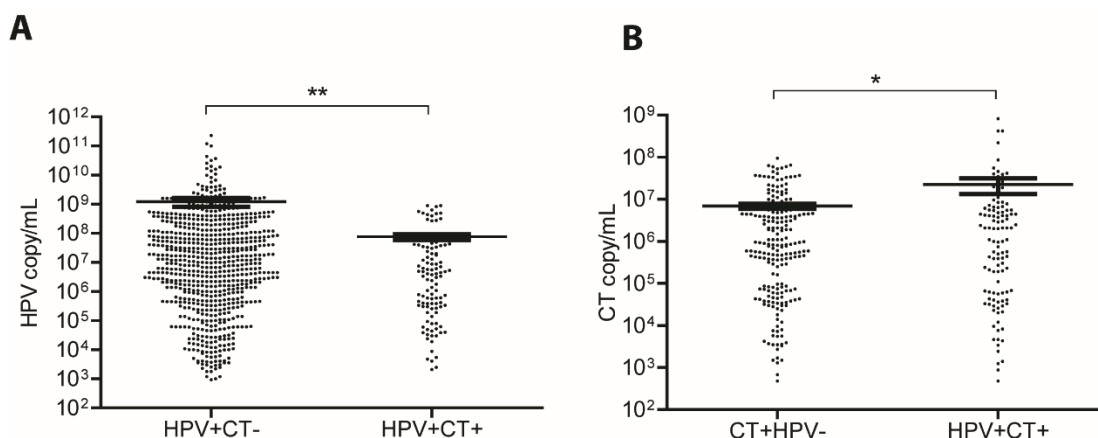
## HPV Genotype Distribution by Cytology Status

Among the 2345 participants, samples from 590 (26.7%) women exhibited the following abnormal cytology: 6 AGC (0.3%), 256 ASC-US (10.9%), 19 ASC-H (0.8%), 170 LSIL (7.2%), 136 HSIL (5.8%), and 3 SCC (0.1%) (Table 1).

Of the 844 HPV-DNA-positive women, samples from 431 had ASC-US or higher grade cytology (18.4%, 431/2345), of which 2 were AGC (0.09%, 2/2345), 140 ASC-US (6.0%), 15 ASC-H (0.6%), 143 LSIL (6.1%), 128 HSIL (5.5%), and 3 SCC (0.1%) (Table 1). The HR-HPV genotypes (except for HPV51, HPV59, and HPV68) were more prevalent in participants with HSIL cytological findings than those with NILM. Among the LR-HPV genotypes, HPV42, 44, and 81 were more frequently found among participants with HSIL than those with NILM. HPV16, 53, and 81 were detected in the 3 SCC cases.

## The Prevalence of *C. trachomatis* and HPV Coinfection

Among the overall participants, 335 were *C. trachomatis* DNA positive (14.3%). Among these, 113 were coinfecting with HPV. The DNA load of HPV was significantly higher in the mono-infection group (median =  $1.2 \times 10^7$  copy/mL) compared to the coinfection cases (median =  $4.1 \times 10^6$  copy/mL). In contrast, the DNA load of *C. trachomatis* was significantly higher in the coinfection group (median =  $1.1 \times 10^6$  copy/mL) than in the mono-infection group (median =  $9.2 \times 10^5$  copy/mL; Figure 1).



**Figure 1. Genomic DNA load of HPV and *C. trachomatis* within mono-infection or co-infection. A.** Genomic DNA load of HPV (copy/ml) among women within mono-infection (median= $1.2 \times 10^7$  copy/ml) and co-infection (median= $4.1 \times 10^6$  copy/ml) (\*\* $P < 0.01$ ) (with Mann Whitney test). **B.** Genomic DNA load of *C. trachomatis* (copy/ml) among women within mono-infection (median= $1.1 \times 10^6$  copy/ml) and co-infection (median= $9.2 \times 10^5$  copy/ml) (\* $P < 0.05$ ) (with T test). Median and interquartile range indicated by solid lines in the graph.

## Age Quintile and HPV Genotype Distribution in Coinfected Patients

In HPV-positive patients, the prevalence of HPV dropped from age quintile 1 (36.2%) to age quintile 3 (30.7%) and then increased progressively to age quintile 5 (42.4%) (Table 2 and Figure 2A). In the coinfection group, most participants were in age quintile 1 (10.4%) (Table 2 and Figure 2B).

Among the coinfecting participants, 14 HR-HPV genotypes (except for HPV35) and 4 LR-HPV genotypes (except for HPV42 and 43) were present. HR-HPV genotypes, including HPV66 (1.5%), HPV52 (1.2%), HPV16 (1.1%), HPV56 (1.0%), and HPV58 (0.9%) were identified in the 113 coinfecting cases (Table 3). HPV11 was the most common LR-HPV genotype (0.5%).

## Potential Risk Factors

Table 4 shows the association between HPV mono- or coinfection with residence areas, ethnicity, education background, profession, marital status, and lifetime number of sex partners. Most indicators, except ethnicity, were significantly associated with HPV-positive results. For coinfection with HPV and *C. trachomatis*, marital status and lifetime sex partners were potential risk factors. Single women (OR = 6.0; 95% CI, 4.044–8.782) or women with >2 sex partners (OR = 1.9; 95% CI, 1.324–2.824) were more likely to be infected with both HPV and *C. trachomatis*.

## Coinfection of *C. trachomatis* and HPV Increases the Risk of Abnormal Cytology

Among the 335 participants with *C. trachomatis* infection, 77 had ASC-US (3.3%, 77/2345), 1 ASC-H (0.04%), 52 LSIL (2.2%), 36 HSIL (1.5%), and 3 SCC (0.1%) (Table 1). Among the 113 coinfecting participants, 90 had abnormal cytological findings and 36 HSIL (Table 3). Thus, coinfection has dramatically increased the risk for abnormal cytology (OR = 11.6; 95% CI, 7.29–18.6).

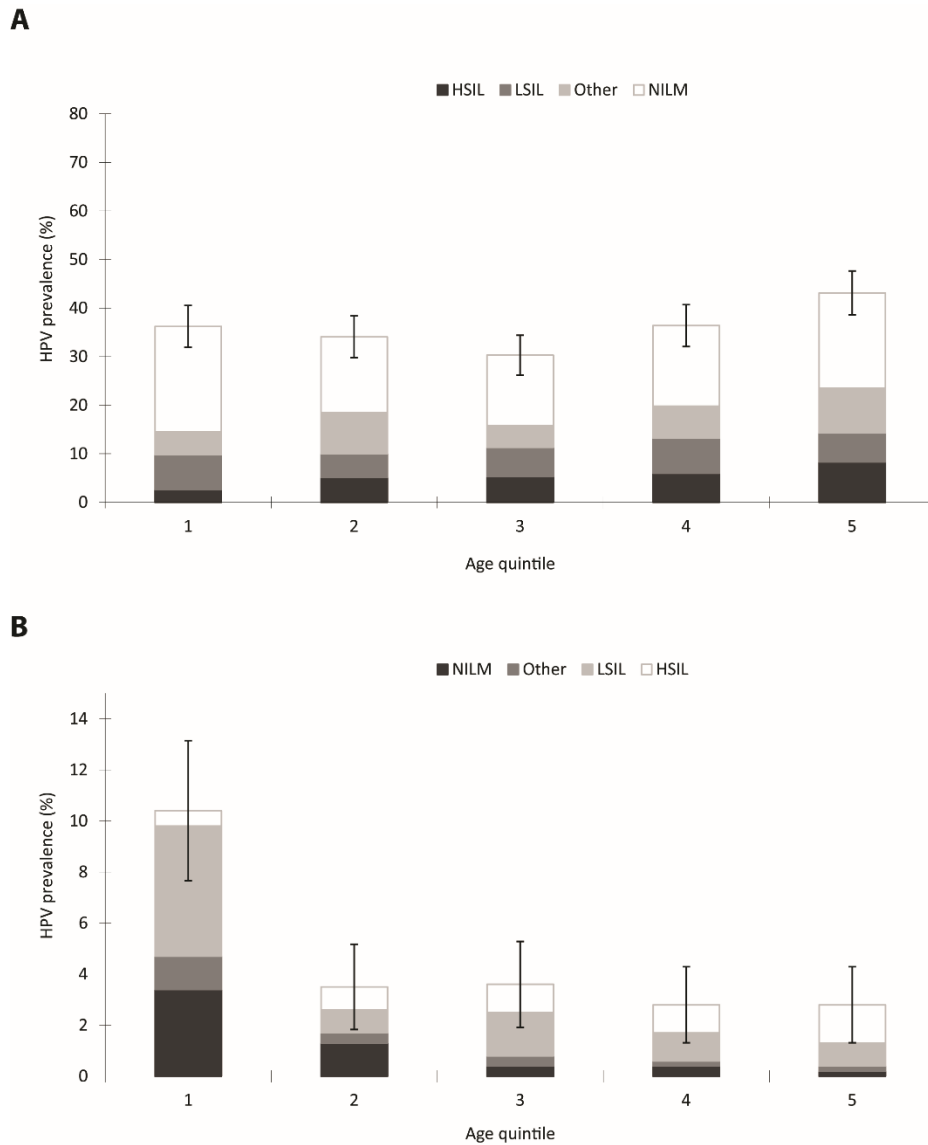
Specifically, coinfection with different HR-HPV genotypes, including HPV16, 18, 31, 33, 39, 52, 53, 56, 58, and 66, showed a higher risk for ASC-US or higher grade cytology. Coinfection with HR-HPV genotypes, including HPV16, 18, 31, 39, 52, 58, 66, and 68, showed increased risk for HSIL. Of note, all the 3 SCC cases detected in this study were in the coinfecting population. Among the coinfections with LR-HPV genotypes, HPV11 and HPV81 showed a higher risk for ASC-US or higher grade cytology and no LR-HPV genotypes exhibited increased risk for HSIL.

Furthermore, the rate of LSIL or HSIL in each age quintile with coinfection was clearly higher than those with HPV mono-infection. Of the 3 SCC cases, 2 were identified in age quintile 3 and 1 in age quintile 5 of the coinfection group.

**Table 2. Comparison of cytological-associated HPV genotypes of HPV-DNA and HPV/C. trachomatis co-infection by age quintiles.**

Groups	Total	NILM Positive (%)	Other cytological findings					LSIL Positive (%)	HSIL Positive (%)	SCC Positive (%)
			AGC Positive (%)	ASC-US Positive (%)	ASC-H Positive (%)	LSIL Positive (%)	HSIL Positive (%)			
Age 1 (19-32 years)	469									
HPV-DNA	170	102 (21.7)	0 (0)	20 (4.3)	2 (0.4)	34 (7.2)	12 (2.6)	0 (0)		
Co-infection	49	16 (3.4)	0 (0)	6 (1.3)	0 (0)	24 (5.1)	3 (0.6)	0 (0)		
Age 2 (32-37 years)	469									
HPV-DNA	160	73 (15.6)	0 (0)	36 (7.7)	4 (0.9)	23 (4.9)	24 (5.1)	0 (0)		
Co-infection	16	6 (1.3)	0 (0)	2 (0.4)	0 (0)	4 (0.9)	4 (0.9)	0 (0)		
Age 3 (37-42 years)	469									
HPV-DNA	144	68 (14.5)	1 (0.2)	17 (3.6)	3 (0.6)	28 (6.0)	25 (5.3)	2 (0.4)		
Co-infection	21	2 (0.4)	0 (0)	3 (0.6)	0 (0)	8 (1.7)	6 (1.3)	2 (0.4)		
Age 4 (42-49 years)	469									
HPV-DNA	171	78 (1.7)	1 (0.2)	27 (5.8)	3 (0.6)	34 (7.2)	28 (6.0)	0 (0)		
Co-infection	13	2 (0.4)	0 (0)	1 (0.4)	0 (0)	5 (1.1)	5 (1.1)	0 (0)		
Age 5 (49-83 years)	469									
HPV-DNA	199	92 (19.6)	0 (0)	40 (8.5)	3 (0.6)	24 (5.1)	39 (8.3)	1 (0.2)		
Co-infection	14	1 (0.2)	0 (0)	1 (0.2)	0 (0)	4 (0.9)	7 (1.5)	1 (0.2)		

Abbreviations: AGC, atypical glandular cells; ASC-H, atypical squamous cells not possible exclude HSIL; ASC-US, atypical squamous cells of undetermined significance; HPV, human papillomavirus; HSIL, high-grade squamous intraepithelial lesion; LSIL, low-grade squamous intraepithelial lesion; NILM, negative for intraepithelial lesion or malignancy; SCC, squamous cell carcinoma.



**Figure 2. Prevalence of HPV infection and cytological findings with or without *C. trachomatis* infection by age quintile and corresponding 95% CIs. A.** 2345 participants were divided into five groups according to their age quintile. 844 HPV DNA positive women were classified based on the age quintile. The prevalence of HPV drops from age quintile 1 (36.2%) to age quintile 3 (30.7%) and then increases progressively to age quintile 5 (42.4%). **B.** 113 women with HPV/*C. trachomatis* co-infection were classified based on their age quintile. In this group, most participants are located at age quintile 1 (10.4%). Furthermore, the rate of LSIL or HSIL in each age quintile with co-infection is clearly higher than those with HPV mono-infection. Of the three SCC cases, two were identified in age quintile 3 and one in age quintile 5 within the co-infection group.

Table 3. Genotypes distribution of HPV-DNA in HPV/C. trachomatis co-infection by cytology status

Co-infection	Cytological findings				≥ASC-us				Cytological findings with ≥ASC-US				HSIL	
	Overall (n=2345) n (%)	NILM (n=1755) n (%)	≥ASC-us (n=590) n (%)	P-value	OR(95%CI)	P-value	AGC (n=6) n (%)	ASC-US (n=256) n (%)	ASC-H (n=19) n (%)	LSIL (n=170) n (%)	HSIL (n=136) n (%)	SCC (n=3) n (%)	OR(95%CI)	P-value
HPV DNA	113 (4.8)	23 (1.0)	90 (3.8)	<0.001	11.6(7.29-18.6)	<0.001	0 (0)	8 (0.3)	0 (0)	43 (1.8)	36 (1.5)	3 (0.1)	20.2(11.6-35.1)	<0.001
Multiple infection	45 (1.9)	12 (0.5)	33 (1.4)	<0.001	8.18(4.20-15.9)	<0.001	-	6 (0.3)	-	17 (0.7)	9 (0.4)	1 (0.04)	9.68(4.01-23.4)	<0.001
Single infection	68 (2.0)	11 (0.5)	57 (2.4)	<0.001	15.4(8.03-29.6)	<0.001	-	2 (0.09)	-	26 (1.1)	27 (1.1)	2 (0.04)	31.7(15.4-65.2)	<0.001
HR-HPV														
HPV-16	26 (1.1)	4 (0.2)	24 (1.0)	<0.001	17.8(6.17-51.6)	<0.001	-	2 (0.09)	-	7 (0.3)	13 (0.6)	2 (0.09)	41.9(13.5-130.4)	<0.001
HPV-18	15 (0.6)	3 (0.1)	12 (0.5)	<0.001	11.9(3.35-42.3)	<0.001	-	1 (0.04)	-	4 (0.2)	7 (0.3)	-	30.1(7.70-117.8)	<0.001
HPV-31	10 (0.4)	3 (0.1)	7 (0.3)	0.001	6.94(1.79-26.9)	0.001	-	1 (0.04)	-	3 (0.1)	3 (0.1)	-	12.9(2.58-64.5)	<0.001
HPV-33	12 (0.5)	4 (0.2)	8 (0.3)	0.001	5.95(1.79-19.8)	0.001	-	-	-	3 (0.1)	5 (0.2)	-	16.1(4.28-60.8)	<0.001
HPV-35	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HPV-39	6 (0.3)	1 (0.04)	5 (0.2)	0.001	14.9(1.73-127.6)	0.001	-	-	-	3 (0.1)	2 (0.09)	-	25.8(2.53-286.4)	<0.001
HPV-45	7 (0.3)	1 (0.04)	3 (0.1)	0.022	8.92(0.93-86.0)	0.022	-	1 (0.04)	-	2 (0.09)	-	-	-	-
HPV-51	7 (0.3)	1 (0.04)	3 (0.1)	0.022	8.92(0.93-86.0)	0.022	-	-	-	2 (0.09)	-	-	-	-
HPV-52	27 (1.1)	5 (0.2)	11 (0.5)	<0.001	6.54(2.26-18.9)	<0.001	-	-	-	4 (0.2)	7 (0.3)	-	12.9(0.80-207.4)	0.020
HPV-53	15 (0.6)	3 (0.1)	6 (0.3)	0.004	5.95(1.48-23.9)	0.004	-	-	-	4 (0.2)	1 (0.04)	1 (0.04)	18.1(5.66-57.7)	<0.001
HPV-56	23 (1.0)	3 (0.1)	10 (0.4)	<0.001	9.92(2.72-36.1)	<0.001	-	-	-	7 (0.3)	1 (0.04)	-	4.3(0.44-41.6)	0.170
HPV-58	22 (0.9)	4 (0.09)	9 (0.4)	<0.001	6.69(2.05-21.8)	<0.001	-	1 (0.04)	-	5 (0.2)	3 (0.1)	-	9.7(2.14-43.7)	<0.001
HPV-59	3 (0.1)	1 (0.04)	1 (0.09)	0.419	2.98(0.19-47.6)	0.419	-	-	-	1 (0.04)	-	-	-	-
HPV-66	35 (1.4)	3 (0.1)	16 (0.7)	<0.001	15.9(4.61-54.6)	<0.001	-	3 (0.1)	-	10 (0.4)	3 (0.1)	-	12.9(2.58-64.5)	<0.001
HPV-68	11 (0.5)	3 (0.1)	4 (0.2)	0.052	3.97(0.89-17.8)	0.052	-	-	-	2 (0.09)	2 (0.09)	-	8.6(1.43-51.9)	0.005
LR-HPV														
HPV-6	4 (0.2)	-	2 (0.09)	-	-	-	-	1 (0.04)	-	1 (0.04)	-	-	-	-
HPV-11	11 (0.5)	1 (0.04)	5 (0.2)	0.001	14.9(1.73-127.6)	0.001	-	2 (0.09)	-	3 (0.1)	-	-	-	-
HPV-42	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HPV-43	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HPV-44	1 (0.04)	-	1 (0.04)	-	-	-	-	-	-	-	-	-	-	-
HPV-81	9 (0.4)	1 (0.04)	4 (0.2)	0.005	11.9(1.32-106.7)	0.005	-	-	-	1 (0.04)	1 (0.04)	1 (0.04)	12.9(0.80-207.4)	0.020

The 95% confidence interval (CI) was obtained by using binomial distribution analysis model.

P value: the HPV prevalence between NILM and abnormal cytological findings was compared using X2 test. Bold type indicated statistically significant values.

Abbreviations: %, prevalence percentage; AGC, atypical glandular cells; ASC-H, atypical squamous cells not possible exclude HSIL; ASC-US, atypical squamous cells of undetermined significance; HR-HPV, high-risk human papillomavirus; HSIL, high-grade squamous intraepithelial lesion; LR-HPV, low-risk human papillomavirus; LSIL, low-grade squamous intraepithelial lesion; n, number of cases; NILM, negative for intraepithelial lesion or malignancy; OR, odds ratio; SCC, squamous cell carcinoma...



**Table 4. Baseline Features of the Participants With HPV DNA Positive and Coinfection**

Variables	Total (N=2345)	HPV-DNA (n=844)		Co-infection (n=113)	
		n	OR (95% CI)	n	OR (95% CI)
Residence area					
Urban	2030	772	1	96	1
Rural	315	72	0.6 (0.459-0.787)	17	1.1 (0.672-1.034)
Ethnicity					
Han	2241	810	1	112	1
Mongolian	92	32	0.96 (0.639-1.45)	0	-
Others	12	2	0.5 (0.103-2.065)	1	1.7 (0.215-12.937)
Education					
Junior high school or less	281	146	1	8	1
Senior high school	1471	421	0.6(0.439-0.691)	73	1.7 (0.831-3.657)
College and higher	593	277	0.9 (0.703-1.149)	32	1.9 (0.862-4.167)
Professions					
Job	1976	759	1	93	1
No job	369	85	0.6 (0.467-0.770)	20	1.2 (0.701-1.891)
Marital status					
Married	1978	679	1	55	1
Single	344	153	2.3 (1.884-2.751)	57	6.0 (4.044-8.782)
Divorced or widowed	23	12	1.7 (0.856-3.501)	1	1.6 (0.207-11.787)
Lifetime sex partners					
1	1479	403	1	53	1
≥2	866	441	1.9 (1.594-2.191)	60	1.9 (1.324-2.824)

Bold type indicated statistically significant values.

Abbreviations: CI, confidence interval; HPV, human papillomavirus; OR, odds ratio.

## Discussion

In this study, 2345 women who visited the Department of Gynecology, Inner Mongolian Maternal and Child Care Hospital were recruited for an HPV and *C. trachomatis* screening program. The prevalence of HPV-DNA was 36.0% in this population, whereas HPV prevalence from other parts of China ranges from 9.03% to 16.8%<sup>[14, 15]</sup>. Furthermore, we identified 335 participants to be *C. trachomatis* DNA positive (14.3%) by real-time PCR. The sensitivity and specificity of this method is 95.7% and 100%, respectively<sup>[12]</sup>.

Among the *C. trachomatis* infected participants, 113 were coinfecting with HPV, and HPV16, HPV58, and HPV52 were the most common genotypes in this population. This finding is similar to that in most regions of China and other areas of East Asian, although there are variations in different region<sup>[15–19]</sup>. HPV16 and HPV52 were more common in HSIL cytology. HPV16 was the most prevalent genotype in normal and abnormal cytological findings. Interestingly, HPV66 represented the genotype most frequently detected in coinfecting participants and nearly half of the cases with this genotype developed abnormal cytology. HPV66 was also common in coinfection with LSIL or HSIL. The overall prevalence of HPV66 is usually low in HPV screening of Chinese women<sup>[9]</sup>. However, it appears more prone to coinfect with *C. trachomatis* than other genotypes. We found the rate of multiple HPV genotypes in coinfecting women was higher than that in those with HPV monoinfection. It has been previously reported that HPV multiple infection does not increase the risk of acquiring further infection but may impair the local immune system<sup>[20]</sup>. Interestingly, patients with coinfection had significantly higher load of *C. trachomatis* DNA compared with monoinfection of *C. trachomatis*. We speculate that this may be associated with host immune response to the infection of 2 pathogens, but further research is required to validate our hypothesis.

Risk factors that significantly increased the odds of HPV-*C. trachomatis* coinfection in Inner Mongolia included marital status and number of lifetime sex partners. We found that married women were less likely to be coinfecting as compared to single women. Whereas, women with 2 or more sexual lifetime partners had a higher risk of coinfection as compared to women with 1 sexual partner. We found that sexual contact with many partners increases the odds ratio for HPV monoinfection, which is consistent with previous studies<sup>[21–23]</sup>. As *C. trachomatis* is mainly sexually transmitted, we expect that coinfection is also likely associated with risky sex behaviors. Coinfection is possibly acquired by interactions with multiple hosts who may each be infected with a particular pathogen species. These interactions may change the transmission, clinical progression, and control of infectious diseases<sup>[24, 25]</sup>. We found that younger age was associated with a higher *C. trachomatis* infection prevalence, which may be related to more risky sexual behavior. Risk factors, including residence areas, education, and profession, were significantly associated with HPV infection, but were not significantly related

to coinfection. Also, unexpectedly, no significant association was observed between ethnicity and risk of coinfection. This is probably due to extensive ethnic miscegenation in the region, or the size of studied population was not large enough.

The association of cervical abnormalities and HPV-*C. trachomatis* coinfection was in line with previous studies with small cohorts<sup>[26, 27]</sup>. Coinfection is known to trigger a much stronger inflammatory reaction and *C. trachomatis* infection is often associated with HPV persistence<sup>[28-30]</sup>. Persistent HPV infection in basal keratinocytes of mucosal epithelium requires an altered epithelial environment. *C. trachomatis* infection may lead to epithelial disruption and therefore may facilitate entry of the virus<sup>[31]</sup>. Chlamydial infection might also disturb the immune response that is necessary to clear the virus<sup>[31]</sup>. Furthermore, we found that coinfection of HR-HPV and *C. trachomatis* exhibited more severe pathogenesis, in particular higher risk of HSIL. Based on the numbers of abnormal cytological findings, *C. trachomatis* mono-infection was associated with low-grade cervical abnormalities (ASC-US or LSIL), but was less likely to be associated with high-grade cervical lesions (HSIL or SCC). The number of cases with HSIL in the coinfection group was significantly higher than that in the HPV mono-infection group. Among 844 HPV-positive cases 128 (15.2%) had HSIL. Among the 113 participants with coinfection, 36 (31.9%) had HSIL. Importantly, all 3 SCC cases were found in patients with coinfection. Interestingly, the genomic DNA loads of both pathogens were significantly different in mono- versus coinfection, indicating the active interactions of these 2 pathogens. But the exact biological implications of these findings remain to be further explored.

Of note, our study has some limitations. Information on the history of women with HPV or *C. trachomatis* infection prior to enrollment or in the follow-up is lacking. These data could help to better interpret the risk factors for persistent HPV infection. Other STD pathogens, including *Neisseria gonorrhoeae* and HSV-1 and -2, which may also contribute to the progression of cervical cancer, were not tested in the study.

In summary, we report the rates of HPV and *C. trachomatis* mono-infection and coinfection were 36.0%, 14.3%, and 4.8%, respectively, in women from Inner Mongolia, China. Factors including marital status and number of lifetime sex partners were significantly associated with coinfection. Importantly, coinfection resulted in more severe cervical pathogenesis and progression to carcinogenesis. These findings bear important implications for future screening and management of patients with HPV and *C. trachomatis* coinfection.

## References

1. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA Cancer J Clin* **2015**; 65:87–108.
2. World Health Organization. Globocan 2012: estimated cancer incidence, mortality and prevalence worldwide in 2012. [http:// globocan.iarc.fr/Default.aspx](http://globocan.iarc.fr/Default.aspx). Accessed 31 August 2014.
3. Yang L, Huangpu XM, Zhang SW, et al. Changes of mortality rate for cervical cancer during 1970's and 1990's periods in China. *Zhongguo Yi Xue Ke Xue Yuan Xue Bao* **2003**; 25:386–90.
4. Li J, Kang LN, Qiao YL. Review of the cervical cancer disease burden in mainland China. *Asian Pac J Cancer Prev* **2011**; 12:1149–53.
5. zur Hausen H. Papillomaviruses in the causation of human cancers—a brief historical account. *Virology* **2009**; 384:260–5.
6. IARC Working Group on the Evaluation of Carcinogenic Risk to Humans. Biological Agents. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, No. 100B. Lyon, France: International Agency for Research on Cancer, **2012**.
7. Brunham RC, Rey-Ladino J. Immunology of *Chlamydia* infection: implications for a *Chlamydia trachomatis* vaccine. *Nat Rev Immunol* **2005**; 5:149–61.
8. Paavonen J. *Chlamydia trachomatis* infections of the female genital tract: state of the art. *Ann Med* **2012**; 44:18–28.
9. Wang X, Ji Y, Li J, et al. Prevalence of human papillomavirus infection in women in the Autonomous Region of Inner Mongolia: a population based study of a Chinese ethnic minority. *J Med Virol* **2018**; 90:148–56.
10. Bauer HM, Greer CE, Manos MM. Determination of genital human papillomavirus infection using consensus PCR. In: Herrington CS, McGee JOD, eds. *Diagnostic molecular pathology: a practical approach*. Oxford: Oxford University Press, **1992**:132–52.
11. Yang H, Li LJ, Xie LX, et al. Clinical validation of a novel real-time human papillomavirus assay for simultaneous detection of 14 high-risk HPV type and genotyping HPV type 16 and 18 in China. *Arch Virol* **2016**; 161:449–54.
12. Jatou K, Bille J, Greub G. A novel real-time PCR to detect *Chlamydia trachomatis* in first-void urine or genital swabs. *J Med Microbiol* **2006**; 55:1667–74.
13. Nayar R, Wilbur D. The Bethesda system for cervicovaginal cytology. Definitions, criteria, and explanatory notes. 3rd ed. Cham, Switzerland: Springer, **2015**.
14. Lin M, Yang LY, Li LJ, Wu JR, Peng YP, Luo ZY. Genital human papillomavirus screening by gene chip in Chinese women of Guangdong province. *Aust N Z J Obstet Gynaecol* **2008**; 48:189–94.
15. Li LK, Dai M, Clifford GM, et al. Human papillomavirus infection in Shenyang City, People's Republic of China: a population-based study. *Br J Cancer* **2006**; 95:1593–7.
16. Clifford GM, Smith JS, Plummer M, Muñoz N, Franceschi S. Human papillomavirus types in invasive cervical cancer worldwide: a meta-analysis. *Br J Cancer* **2003**; 88: 63–73.
17. Lu S, Cong X, Li M, Chang F, Ma L, Cao YT. Distribution of high-risk human papillomavirus genotypes in HPV-infected women in Beijing, China. *J Med Virol* **2015**; 87:504–7.
18. Chan PK, Li WH, Chan MY, Ma WL, Cheung JL, Cheng AF. High prevalence of human papillomavirus type 58 in Chinese women with cervical cancer and precancerous lesions. *J Med Virol* **1999**; 59:232–8.
19. Quek SC, Lim BK, Domingo E, et al. Human papillomavirus type distribution in invasive cervical cancer and high-grade cervical intraepithelial neoplasia across 5 countries in Asia. *Int J Gynecol Cancer* **2013**; 23:148–56.
20. Comar M, Monasta L, Zanotta N, Vecchi Brumatti L, Ricci G, Zauli G. Human papillomavirus infection is associated with decreased levels of GM-CSF in cervico-vaginal fluid of infected women. *J Clin Virol* **2013**; 58:479–81.

21. Francis SC, Ao TT, Vanobberghen FM, et al. Epidemiology of curable sexually transmitted infections among women at increased risk for HIV in northwestern Tanzania: inadequacy of syndromic management. *PLoS One* **2014**; 9: e101221.
22. Wilson C, Sathiyasuman A. Associated risk factors of STIs and multiple sexual relationships among youths in Malawi. *PLoS One* **2015**; 10: e0134286.
23. de Lima YA, Turchi MD, Fonseca ZC, et al. Sexually transmitted bacterial infections among young women in Central Western Brazil. *Int J Infect Dis* **2014**; 25:16–21.
24. Sternberg ED, Lefevre T, Rawstern AH, de Roode JC. A virulent parasite can provide protection against a lethal parasitoid. *Infect Genet Evol* **2011**; 11:399–406.
25. Pedersen AB, Fenton A. Emphasizing the ecology in parasite community ecology. *Trends Ecol Evol* **2007**; 22:133–9.
26. Discacciati MG, Gimenes F, Pennacchi PC, et al. MMP-9/ RECK imbalance: a mechanism associated with high-grade cervical lesions and genital infection by human Papillomavirus and *Chlamydia trachomatis*. *Cancer Epidemiol Biomarkers Prev* **2015**; 24:1539–47.
27. Souza RP, de Abreu AL, Ferreira ÉC, et al. Simultaneous detection of seven sexually transmitted agents in human immunodeficiency virus-infected Brazilian women by multiplex polymerase chain reaction. *Am J Trop Med Hyg* **2013**; 89:1199–202.
28. Castle PE, Escoffery C, Schachter J, et al. *Chlamydia trachomatis*, herpes simplex virus 2, and human T-cell lymphotropic virus type 1 are not associated with grade of cervical neoplasia in Jamaican colposcopy patients. *Sex Transm Dis* **2003**; 30:575–80.
29. Samoff E, Koumans EH, Markowitz LE, et al. Association of *Chlamydia trachomatis* with persistence of high-risk types of human papillomavirus in a cohort of female adolescents. *Am J Epidemiol* **2005**; 162:668–75.
30. Silins I, Ryd W, Strand A, et al. *Chlamydia trachomatis* infection and persistence of human papillomavirus. *Int J Cancer* **2005**; 116:110–5.
31. Magaña-Contreras M, Contreras-Paredes A, Chavez Blanco A, Lizano M, De la Cruz-Hernandez Y, De la Cruz Hernandez E. Prevalence of sexually transmitted pathogens associated with HPV infection in cervical samples in a Mexican population. *J Med Virol* **2015**; 87:2098



# Chapter 4

## **Optimizing Cervical Cancer Screening and Prevention in Rural Areas of China: A Modeling Analysis up to 2030**

Yunpeng Ji, Hua Liu, Yong Ye, Kai Zhang, Maikel P. Peppelenbosch, Zhongren Ma, Qiuwei Pan

Embargoed





## **PART II**

### ***HEV cross-species transmission and infection in pregnant women***



# Chapter 5

## Estimating the global prevalence of hepatitis E virus in swine and pork products

Pengfei Li<sup>#</sup>, Yunpeng Ji<sup>#</sup>, Yunlong Li, Zhongren Ma, Qiuwei Pan

(#: contributed equally)



## **Abstract**

Zoonotic transmission of hepatitis E virus (HEV), in particular the genotype (GT) 3 and GT4 strains, constitutes a major one health issue. Swine serves as an important reservoir and the processed pork products essentially contribute to foodborne transmission. This study comprehensively estimated HEV prevalence in domestic pigs, wild boars, and pork products at global scale. At global level, we found nearly 60% domestic pigs and 27% wild boars have ever encountered HEV infection based seroprevalence rate. Nearly 13% domestic and 9.5% wild swine are actively infected based HEV RNA positivity. Importantly, about 10% of commercial pork products are HEV RNA positive, although available data are limited in this respect. Our results indicate the high prevalence rate of HEV infection in pigs and widespread contamination in pork products, although there are substantial variations at regional and country levels. These findings are important for better understanding the global epidemiology and clinical burden of HEV infection in human population related to zoonotic transmission.

**Keywords:** Hepatitis E virus; Swine; Prevalence; Epidemiology; Zoonotic transmission

## **Introduction**

Hepatitis E virus (HEV) is a non-enveloped, single-stranded positive-sense RNA virus. It is recognized as the leading cause of acute viral hepatitis. Globally, it is estimated approximately 939 million corresponding to 1 in 8 individuals have ever been infected with HEV (Li, Liu, et al., 2020). Among the different HEV genotypes (GT) that affect human health, GT3 and GT4 are zoonotic, which have been found in various animal species (Zhou et al., 2019). Although HEV infection is usually self-limiting or asymptomatic in healthy individuals, GT3 and GT4 HEV infection in organ transplant patients is prone to develop chronic hepatitis (Kamar et al., 2008; Wang, Liu, Pan, & Zhao, 2020).

Pigs serve as the major reservoir for the zoonotic HEV strains. Anti-HEV antibodies have been widely detected in both domestic pigs and wild boars (Baechlein et al., 2010; Fredriksson-Ahomaa et al., 2020; Geng et al., 2010). There are different routes of HEV transmission from pigs to human, such as direct contact with the animal, indirectly through contaminated environment and the consumption of pork products. However, the contribution of these different transmission routes can vary tremendously among different settings attributing to multi-factors, such as socioeconomic status, farming systems, food chains and life styles. Nevertheless, the widespread consuming of pork products is inevitably posing a major risk of HEV foodborne transmission in public health. Hepatitis E cases linking to consuming undercooked pork or wild boar meat have been widely reported (Matsuda, Okada, Takahashi, & Mishiro, 2003; Yazaki et al., 2003).

Globally, the epidemiological feature and clinical burden of zoonotic HEV infection in human population are distinct among different countries/regions (Li, Liu, et al., 2020). It is intriguing to postulate whether this is associated with the specific prevalence rate of HEV in local swine population and available pork products. In this study, we aim to estimate the global prevalence of HEV in both domestic pigs and wild boars, as well as pork products in retailers.

## **Methods**

### **Data sources, search strategies and study selection**

A systematic search was conducted in Medline, Embase, Web of science, Cochrane CENTRAL and Google scholar. Databases were searched for articles in the English language from inception until 31 May 2021. Studies were included if they contained epidemiological data about HEV in domestic pig or wild boar. The full search strategies and study selection criteria are provided in the Supplementary file S1-S2.

### **Data extraction, quality assessment and statistical analysis**

Eligible studies were further divided into three study populations: domestic swine, wild boars and market/retailer pork products. Studies were scored according to Joanna Briggs Institute checklist for prevalence studies. A 95% confidence interval (95% CI) was estimated using Wilson score method, and pooled prevalence rate was calculated by the DerSimonian-Laird random-effects model with Freeman-Tukey double arcsine transformation. Funnel plots and Egger regression test were used to assess potential publication biases. 'Meta' package in the R-3.5.3 statistical software was used for meta-analysis as previously described (Li, Ikram, Peppelenbosch, Ma, & Pan, 2020; Liu et al., 2021). Sensitivity analysis was performed by using 'metainf' to investigate the effects of group source and potentially unrepresentative samples. The details of quality assessment and statistical analysis are provided in supplementary S3.

## Results and Discussion

By comprehensively searching 5 databases (supplementary S1-S3), we identified a total of 215 studies met the inclusion criteria, which were processed for analysis of HEV prevalence in domestic swine, wild boars and pork products (sFig. 1). First, we estimated anti-HEV seroprevalence (indication of ever exposure) and HEV RNA positivity (indication of active infection) in domestic pigs. A total of 84 studies were included to estimate the global anti-HEV seroprevalence, resulting in a pooled rate of 59.33% (37 countries, 95% CI 53.64-64.90,  $I^2=99\%$ ; sFig. 2). The highest seroprevalence was found in Oceania (75%, 95% CI 64.28-84.40), but this is only based on one study which likely causes bias in estimation. The second highest seroprevalence rate was found in Asia (67.45%, 95% CI 53.50-79.99,  $I^2=100\%$ ), followed by Europe (57.46%, 95% CI 49.82-64.93,  $I^2=99\%$ ), Africa (53.46%, 95% CI 43.26-63.52,  $I^2=92\%$ ), and South America (53.03%, 95% CI 33.79-71.81,  $I^2=99\%$ ) (sFig. 3). Based on 118 studies from 45 countries/territories, the global estimation of HEV RNA positive rate was 12.71% (95% CI 10.81-14.73,  $I^2=97\%$ ). The highest rate was found in North America (18.10%, 95% CI 8.71-29.84,  $I^2=97\%$ ), followed by Europe (17.19%, 95% CI 13.16-21.61,  $I^2=98\%$ ), South America (15.67%, 95% CI 6.75-27.33,  $I^2=98\%$ ), Africa (12.29%, 95% CI 0.01-38.70,  $I^2=99\%$ ), Asia (8.23%, 95% CI 6.21-10.49,  $I^2=97\%$ ), and Oceania (6.52%, 95% CI 2.23-12.63) (sFig. 4 and 5).

At country level, HEV prevalence in domestic pigs varies substantially, from 9.90% (Thailand, 95% CI 8.01-11.96) to 84.02% (India, 95% CI 44.05-100.00,  $I^2=99\%$ ) of anti-HEV seroprevalence, and from 0% (Croatia, 95% CI 0.00-0.37) to 76.67% (Nigeria, 95% CI 67.32-84.89) of HEV RNA positivity (Table 1 and 2, Figure 1). Importantly, we have collected genotyping information of swine HEV. GT3 is universally prevalent across the globe, whereas GT4 is mainly present in Western Pacific region. Interestingly, GT3 and GT4 are co-circulating in countries/territories, such as mainland China, Taiwan, Korea and Japan (Table 1 and 2, Figure 1). This is consistent with clinical observations that both GT3 and GT4 HEV patients have been reported from these regions (Kitaura et al., 2020; Owada et al., 2020; L. Wang et

al., 2020; Wang et al., 2018). Although Europe is dominated by GT3, GT4 HEV has been identified in some peculiar cases including chronically infected patients(Micas et al., 2021). Here, we found that GT3 and GT4 are also co-circulating in domestic pig populations in Belgium (Figure 1). Thus, the emergence of GT4, which is thought to be more pathogenic, requires more attention from both public health and patient care perspectives.

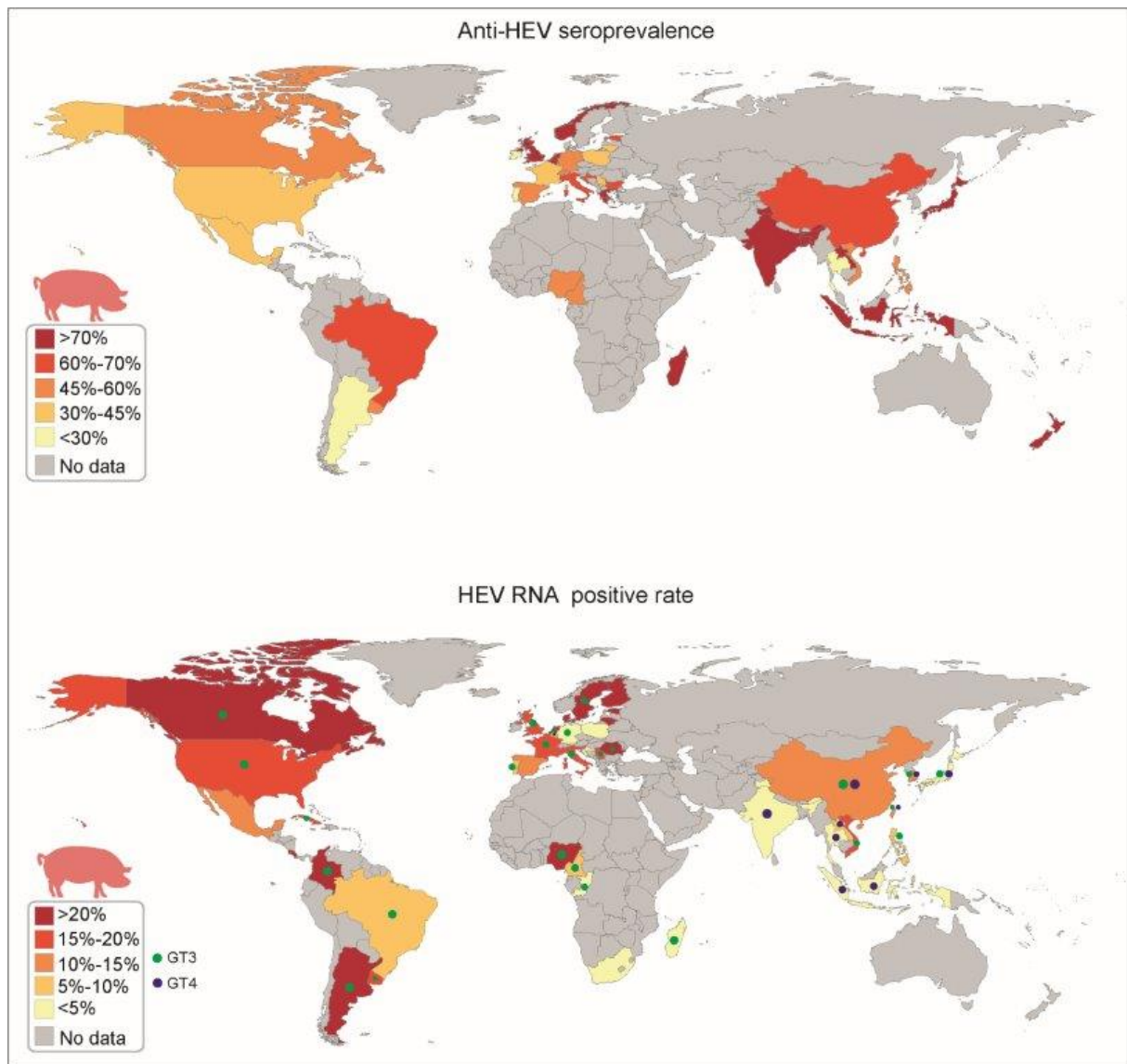


**Table 1. HEV seroprevalence in domestic swine.**

Continent	Country	No. of studies	Events	Anti-HEV antibodies			Genotype	
				Tested (n)	Prevalence (%)	95% CI		
Asia	Bangladesh	1	82	100	82.00	73.80-88.98	-	
	China	13	7036	15461	65.73	46.01-82.99	Major GT4; Minor GT3	
	India	2	297	360	84.02	44.05-100.00	GT4	
	Indonesia	2	224	307	72.99	67.85-77.84	GT4	
	Korea	-	-	-	-	-	Major GT3; Minor GT4	
	Japan	3	327	442	74.33	39.02-97.42	GT3+GT4	
	Laos	2	769	899	81.36	26.33-100.00	GT4	
	Philippines	1	155	299	51.84	46.16-57.49	GT3	
	Vietnam	1	300	586	51.19	47.14-55.24	GT3	
	Thailand	1	87	879	9.90	8.01-11.96	GT4	
	Taiwan	-	-	-	-	-	GT3+GT4	
	Total	26	9277	19333	67.45	53.50-79.99	-	
	Europe	Belgium	1	307	420	73.10	68.74-77.23	GT3+GT4
		Bulgaria	4	652	1049	61.69	48.15-74.38	-
Croatia		2	524	1484	64.94	8.05-100.00	-	
Czech		-	-	-	-	-	GT3	
Denmark		-	-	-	-	-	-	
Estonia		1	234	380	61.58	56.63-66.41	GT3	
Finland		-	-	-	-	-	GT3	
France		3	1785	7814	38.19	9.19-72.96	GT3	
Germany		3	1949	3861	54.79	54.92-77.30	GT3	
Greece		1	76	96	79.17	70.41-86.76	-	
Hungary		-	-	-	-	-	GT3	
Ireland		1	89	330	26.97	22.31-31.90	-	
Italy		6	2961	5737	66.58	54.92-77.30	GT3	
Lithuania		1	168	384	43.75	38.82-48.74	-	
Netherlands		2	775	976	75.23	60.91-87.18	GT3	
Norway		1	484	663	73.00	69.55-76.32	-	
Poland		1	63	143	44.06	35.99-52.28	-	
Portugal		1	4	29	13.79	3.21-29.13	GT3	
Romania		-	-	-	-	-	GT3	
Serbia		2	271	654	41.14	28.63-54.25	GT3	
Slovenia		-	-	-	-	-	GT3	
Spain		5	651	1925	49.66	27.37-72.02	-	
Sweden		-	-	-	-	-	GT3	
Switzerland	2	1281	2199	58.27	56.20-60.33	-		
UK	2	692	805	79.56	42.36-99.54	GT3		
Total	39	12966	28949	57.46	49.82-64.93	-		
Oceania	New Zealand	1	54	72	75.00	64.28-84.40	-	
	New Caledonia	-	-	-	-	-	GT3	
	Total	1	54	72	75.00	64.28-84.40	-	
North America	Canada	1	594	998	59.52	56.45-62.55	GT3	
	Costa Rica	-	-	-	-	-	GT3	
	Mexico	3	964	2055	44.98	27.55-63.07	-	
	USA	2	2036	5117	39.73	38.39-41.08	GT3	
	Cuba	-	-	-	-	-	GT3	
	Total	6	3594	8170	45.06	35.45-54.86	-	
South America	Argentina	1	22	97	22.68	14.84-31.59	GT3	
	Brazil	5	1021	1542	60.38	37.62-81.01	GT3	
	Uruguay	1	103	220	46.82	40.25-53.44	GT3	
	Colombia	-	-	-	-	-	GT3	
	Total	7	1146	1859	53.03	33.79-71.81	-	
Africa	Cameroon	2	286	615	46.50	42.56-50.46	GT3	
	Madagascar	1	178	250	71.20	65.42-76.66	GT3	
	Nigeria	2	204	406	51.76	41.76-61.70	GT3	
	Congo	-	-	-	-	-	GT3	
	South Africa	-	-	-	-	-	-	
	Total	5	668	1271	53.46	43.26-63.52	-	
Overall	Global	84	27705	59654	59.33	53.64-64.90	GT3+GT4	

**Table 2. HEV virology prevalence in domestic swine.**

Continent	Country	No. of studies	Events	HEV RNA		95% CI	Genotype
				Tested (n)	Prevalence (%)		
	Bangladesh	-	-	-	-	-	-
Asia	China	26	1233	19493	10.40	7.48-13.74	Major GT4; Minor GT3
	India	4	30	746	3.35	0.71-7.57	GT4
	Indonesia	2	3	307	0.93	0.06-2.48	GT4
	Korea	5	162	1294	11.13	4.63-19.85	Major GT3; Minor GT4
	Japan	4	82	869	3.01	0.00-13.24	GT3+GT4
	Laos	2	26	455	5.69	0.01-18.78	GT4
	Philippines	1	22	299	7.36	4.65-10.62	GT3
	Vietnam	1	148	774	19.12	16.42-21.97	GT3
	Thailand	1	25	875	2.86	1.85-4.07	GT4
	Taiwan	2	82	816	10.17	0.00-43.06	GT3+GT4
	Total	48	1813	25928	8.23	6.21-10.49	-
Europe	Belgium	1	8	115	6.96	2.91-12.43	GT3+GT4
	Bulgaria	-	-	-	-	-	-
	Croatia	1	0	469	0.00	0.00-0.37	-
	Czech	-	-	-	-	-	GT3
	Denmark	1	48	97	49.48	39.53-59.45	-
	Estonia	1	103	449	22.94	19.16-26.95	GT3
	Finland	1	15	67	22.39	13.11-33.23	GT3
	France	6	343	5949	15.60	7.61-25.66	GT3
	Germany	1	3	120	2.50	0.31-6.24	GT3
	Greece	-	-	-	-	-	-
	Hungary	1	52	248	20.97	16.11-26.27	GT3
	Ireland	-	-	-	-	-	-
	Italy	8	311	2031	19.20	9.68-30.95	GT3
	Lithuania	1	106	470	22.55	18.88-26.45	-
	Netherlands	2	55	161	38.90	1.20-88.42	GT3
	Norway	-	-	-	-	-	-
	Poland	1	5	146	3.42	0.97-7.11	-
	Portugal	2	44	229	7.94	0.00-40.29	GT3
	Romania	1	6	19	31.58	12.27-54.50	GT3
	Serbia	1	51	330	15.45	11.74-19.57	GT3
	Slovenia	2	142	896	15.70	13.36-18.18	GT3
Spain	5	64	427	11.77	4.72-21.23	-	
Sweden	2	150	603	25.40	18.14-33.42	GT3	
Switzerland	-	-	-	-	-	-	
UK	5	232	1483	19.31	6.05-37.45	GT3	
Total	44	1768	14422	17.19	13.16-21.61	-	
Oceania	New Zealand	-	-	-	-	-	-
	New Caledonia	1	6	92	6.52	2.23-12.63	GT3
Total	1	6	92	6.52	2.23-12.63	-	
North America	Canada	2	32	200	22.70	0.00-80.63	GT3
	Costa Rica	1	19	52	36.54	23.90-50.16	GT3
	Mexico	2	28	130	10.40	0.00-56.02	-
	USA	3	366	5256	15.52	3.57-33.50	GT3
	Cuba	1	10	53	18.87	9.32-30.65	GT3
	Total	9	455	5691	18.10	8.71-29.84	-
South America	Argentina	2	59	189	47.03	0.00-100.0	GT3
	Brazil	7	123	1601	7.19	1.86-15.41	GT3
	Uruguay	1	25	150	16.67	11.09-23.09	GT3
	Colombia	1	87	250	34.80	29.01-40.83	GT3
	Total	11	294	2341	15.67	6.75-27.33	-
Africa	Cameroon	1	8	136	5.88	2.45-10.56	GT3
	Madagascar	1	3	345	0.87	0.11-2.19	GT3
	Nigeria	1	69	90	76.67	67.32-84.89	GT3
	Congo	1	1	40	2.50	0.00-10.42	GT3
	South Africa	1	7	160	4.38	1.66-8.18	-
	Total	5	88	771	12.29	0.01-38.70	-
Overall	Global	118	4424	49245	12.71	10.81-14.73	GT3+GT4



**Figure 1. Global prevalence of anti-HEV seroprevalence and HEV RNA positivity among domestic swine.**

Next, we performed subgroup analysis of pigs at different developmental stages of their life. As expected, the anti-HEV seroprevalence rate increases over time, from 42.19% (95% CI 26.79-58.40,  $I^2 = 97\%$ ) in 0-4 month old pigs, 49.27% (95% CI 30.37-68.29%,  $I^2 = 98\%$ ) in 5-8 month pigs, 66.20% (95% CI 55.78-75.89,  $I^2 = 97\%$ ) in over 9 month age pigs. In contrast, the positive rate of HEV RNA showed a reverse pattern, with positive rate of 17.62% (95% CI 12.83-22.96,  $I^2 = 91\%$ ) in pigs of 0-4 month age, 10.75% (95% CI 4.26-19.51,  $I^2 = 89\%$ ) of 5-8 month age, and 6.59% (95% CI 0.86-16.27,  $I^2 = 95\%$ ) over 9 month age (Figure 2, sFig 6 and 7).

Considering the clear differences in husbandry and natural habitat between wild and domesticated pigs, we separately estimated HEV prevalence in wild boars. Based on data extracted from 33 studies from 19 countries/territories, we estimated that the overall anti-HEV seroprevalence was 26.82% (95% CI 21.69-32.28,  $I^2 = 98\%$ ) (Table 3 and 4, sFig. 8). Based

on 37 studies from 18 countries/territories, the pooled rate of HEV RNA positivity was 9.45% (95% CI 6.42-12.96,  $I^2 = 96\%$ ) (Table 3 and 4, sFig. 9).

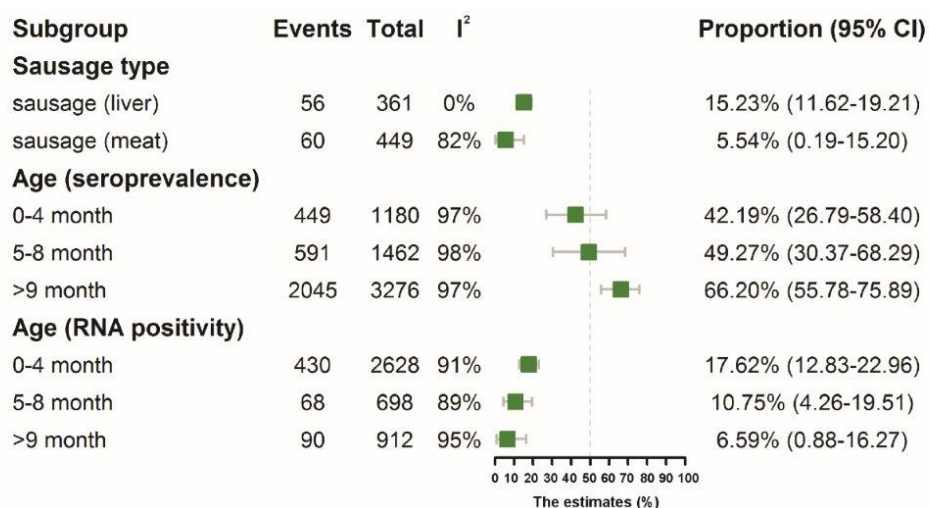


Figure 2. Subgroup analysis of anti-HEV seroprevalence or HEV RNA positivity.

Table 3. HEV seroprevalence in wild boars.

Country	No. of studies	Events	Anti-HEV antibodies			Genotype
			Tested (n)	Prevalence (%)	95% CI	
China	1	186	758	24.52	21.51-27.67	-
Korea	1	1041	2736	38.05	36.24-39.88	Major GT4, minor GT3
Japan	6	262	1139	19.26	10.13-30.37	Major GT3, minor GT4
Thailand	-	-	-	-	-	GT3
Bulgaria	1	98	240	40.83	34.68-47.13	-
Croatia	1	311	1000	31.10	28.27-34.01	-
Czech	1	31	366	8.47	5.82-11.56	-
Estonia	1	81	471	17.20	13.92-20.75	GT3
France	2	160	767	21.07	8.35-37.62	GT3
Germany	1	81	180	45.00	37.78-52.33	GT3
Hungary	-	-	-	-	-	-
Italy	6	560	3416	33.59	15.76-54.21	GT3
Lithuania	1	178	312	57.05	51.51-62.50	-
Netherlands	1	293	1029	28.47	25.76-31.27	-
Poland	1	90	290	31.03	25.83-36.49	-
Portugal	-	-	-	-	-	GT3
Romania	-	-	-	-	-	GT3
Slovenia	1	87	288	30.21	25.03-35.65	-
Spain	4	409	1299	41.51	25.30-58.72	-
Sweden	-	-	-	-	-	GT3
Switzerland	1	38	303	12.54	9.03-16.52	-
Turkey	1	0	93	0	0.00-1.84	-
EU/multiples	1	12	104	11.54	6.02-18.47	-
Uruguay	1	31	140	22.14	15.62-29.43	-
Total	33	3949	14931	26.82	21.69-32.28	-

**Table 4. HEV virology prevalence in wild boars.**

Country	No. of studies	Events	HEV RNA		95% CI	Genotype
			Tested (n)	Prevalence (%)		
China	-	-	-	-	-	-
Korea	1	24	1859	1.29	0.82-1.86	Major GT4, minor GT3
Japan	5	93	2609	3.20	2.51-3.95	Major GT3, minor GT4
Thailand	1	1	31	3.23	0.00-13.33	GT3
Bulgaria	-	-	-	-	-	-
Croatia	1	17	150	11.33	6.70-16.95	-
Czech	-	-	-	-	-	-
Estonia	1	13	81	16.05	8.77-24.93	GT3
France	2	15	637	2.35	1.28-3.71	GT3
Germany	4	157	701	22.33	3.34-51.18	GT3
Hungary	1	8	75	10.67	4.53-18.80	-
Italy	10	232	1697	13.56	6.57-22.46	GT3
Lithuania	1	86	505	17.03	13.87-20.44	-
Netherlands	2	8	158	2.67	0.00-14.30	-
Poland	-	-	-	-	-	-
Portugal	1	24	120	20.00	13.28-27.67	GT3
Romania	1	9	50	18.00	8.41-30.02	GT3
Slovenia	1	1	288	0.35	0.00-1.49	-
Spain	2	43	296	14.47	6.48-24.86	-
Sweden	1	13	159	8.18	4.36-13.00	GT3
Switzerland	-	-	-	-	-	-
Turkey	-	-	-	-	-	-
EU/multiples	1	4	104	3.85	0.83-8.58	-
Uruguay	1	13	140	9.29	4.97-14.72	-
Total	37	761	9660	9.45	6.42-12.96	-

Given the important role of foodborne transmission, we collected data on HEV RNA detection rates of pork meat, liver and sausage in retailers. This generated positive rate of 9.5% (95% CI 5.14-14.90,  $I^2=94\%$ ), with 13.27% (95% CI 0.99-35.12,  $I^2=98\%$ ) in meat, 6.59% (95% CI 1.83-13.49,  $I^2=92\%$ ) in liver and 11.70% (95% CI 7.62-16.47,  $I^2=71\%$ ) in sausage (sFig. 10). Sausage production represents a very large industry across the globe, particularly in Europe. Sausages are popular in groceries and sold in a variety of species. We thus furtherly compared the HEV positivity between liver sausage and pork sausage. Notably, we estimated a nearly 3-fold HEV positivity rate of 15.23% (95% CI 11.62-19.21,  $I^2=0\%$ ) in liver sausage, compared with 5.54% (95% CI 0.19-15.20,  $I^2=82\%$ ) in pork sausage (Figure 2, sFig.11-12).

Finally, we performed sensitivity analysis for the prevalence analysis in domestic pigs and wild boars. In this meta-analysis, no significant change was observed by arbitrarily excluding any study from these groups. This low sensitivity supports the reliability of our estimation. However, funnel plot and Egger's test indicate the presence of publication bias ( $p>0.05$ ) in three analyses, including seroprevalence and HEV RNA prevalence among domestic swine, and RNA prevalence among wild boars, which may potentially compromise the accuracy of prevalence estimation (sFig. 13-26). Another limitation of our study is that we were unable to estimate HEV prevalence in pork products at regional/country levels and clarify the original place of the products, due to limited data available. Because the current food production and

supply chains are diverse and complicated; it has become increasingly important to trace the origin of the contaminated products.

In summary, we found nearly 60% domestic pigs and 27% wild boars have ever encountered HEV infection at global level. Nearly 13% domestic and 9.5% wild swine are experiencing active infection. The risk of potential foodborne transmission is highlighted by our estimation that around 10% commercial pork products are HEV RNA positive. However, there remains gaps of translating these knowledge for better understanding the global epidemiology and clinical burden of HEV infection in human population related to zoonotic transmission. Because HEV zoonosis also involves many other factors, including socioeconomic status, farming style, food production and supply, as well as life styles. Nevertheless, our findings have set a stage for future research to further study the role of swine related HEV zoonosis and to facilitate the development of prevention and mitigation strategies.

(the supplementary information of this chapter can be found with this link: <https://www.sciencedirect.com/science/article/pii/S235277142100152X?via%3Dihub>)

## References

1. Baechlein, C., Schielke, A., Johne, R., Ulrich, R. G., Baumgaertner, W., & Grummer, B. (2010). Prevalence of Hepatitis E virus-specific antibodies in sera of German domestic pigs estimated by using different assays. *Vet Microbiol*, 144(1-2), 187-191. doi: S0378-1135(09)00602-6 [pii]10.1016/j.vetmic.2009.12.011
2. Fredriksson-Ahomaa, M., London, L., Skrzypczak, T., Kantala, T., Laamanen, I., Bistrom, M., Gadd, T. (2020). Foodborne Zoonoses Common in Hunted Wild Boars. *Ecohealth*, 17(4), 512-522. doi:10.1007/s10393-020-01509-510.1007/s10393-020-01509-5 [pii]
3. Geng, Y., Wang, C., Zhao, C., Yu, X., Harrison, T. J., Tian, K., & Wang, Y. (2010). Serological prevalence of hepatitis E virus in domestic animals and diversity of genotype 4 hepatitis E virus in China. *Vector Borne Zoonotic Dis*, 10(8), 765-770. doi:10.1089/vbz.2009.0168
4. Kamar, N., Selves, J., Mansuy, J. M., Ouezzani, L., Peron, J. M., Guitard, J., Rostaing, L. (2008). Hepatitis E virus and chronic hepatitis in organ-transplant recipients. *N Engl J Med*, 358(8), 811-817. doi:358/8/811 [pii]10.1056/NEJMoa0706992
5. Kitaura, S., Wakabayashi, Y., Okazaki, A., Okada, Y., Okamoto, K., Ikeda, M., Moriya, K. (2020). The First Case Report of Acute Symptomatic HEV Genotype 4 Infection in an HIV-positive Patient in Japan. *Intern Med*, 59(13), 1655-1658. doi:10.2169/internalmedicine.4505-20
6. Li, P., Ikram, A., Peppelenbosch, M. P., Ma, Z., & Pan, Q. (2020). Systematically mapping clinical features of infections with classical endemic human coronaviruses. *Clin Infect Dis*. doi:5905561 [pii]10.1093/cid/ciaa1386
7. Li, P., Liu, J., Li, Y., Su, J., Ma, Z., Bramer, W. M., Pan, Q. (2020). The global epidemiology of hepatitis E virus infection: A systematic review and meta-analysis. *Liver Int*, 40(7), 1516-1528. doi:10.1111/liv.14468
8. Liu, J., Ayada, I., Zhang, X., Wang, L., Li, Y., Wen, T., Pan, Q. (2021). Estimating global prevalence of metabolic dysfunction-associated fatty liver disease in overweight or obese adults. *Clin Gastroenterol Hepatol*. doi: S1542-3565(21)00208-1 [pii] 10.1016/j.cgh.2021.02.030

9. Matsuda, H., Okada, K., Takahashi, K., & Mishiro, S. (2003). Severe hepatitis E virus infection after ingestion of uncooked liver from a wild boar. *J Infect Dis*, 188(6), 944. doi: JID31294 [pii]1086/378074
10. Micas, F., Suin, V., Peron, J. M., Scholtes, C., Tuailon, E., Vanwollegem, T., Abravanel, F. (2021). Analyses of Clinical and Biological Data for French and Belgian Immunocompetent Patients Infected with Hepatitis E Virus Genotypes 4 and 3. *Front Microbiol*, 12, 645020. doi:10.3389/fmicb.2021.645020
11. Owada, Y., Oshiro, Y., Inagaki, Y., Harada, H., Fujiyama, N., Kawagishi, N., Ohkohchi, N. (2020). A Nationwide Survey of Hepatitis E Virus Infection and Chronic Hepatitis in Heart and Kidney Transplant Recipients in Japan. *Transplantation*, 104(2), 437-444. doi:10.1097/TP.000000000000280100007890-202002000-00035 [pii]
12. Wang, L., Yan, L., Jiang, J., Zhang, Y., He, Q., Zhuang, H., & Wang, L. (2020). Presence and persistence of hepatitis E virus RNA and proteins in human bone marrow. *Emerg Microbes Infect*, 9(1), 994-997. doi:10.1080/22221751.2020.1761762
13. Wang, Y., Liu, S., Pan, Q., & Zhao, J. (2020). Chronic hepatitis E in an immunocompetent patient. *Clin Res Hepatol Gastroenterol*, 44(3), e66-e68. doi: S2210-7401(19)30184-6 [pii]10.1016/j.clinre.2019.08.001
14. Wang, Y., Wang, S., Wu, J., Jiang, Y., Zhang, H., Li, S., Zhao, J. (2018). Hepatitis E virus infection in acute non-traumatic neuropathy: A large prospective case-control study in China. *EBioMedicine*, 36, 122-130. doi: S2352-3964(18)30345-1 [pii]10.1016/j.ebiom.2018.08.053
15. Yazaki, Y., Mizuo, H., Takahashi, M., Nishizawa, T., Sasaki, N., Gotanda, Y., & Okamoto, H. (2003). Sporadic acute or fulminant hepatitis E in Hokkaido, Japan, may be food-borne, as suggested by the presence of hepatitis E virus in pig liver as food. *J Gen Virol*, 84(Pt 9), 2351-2357. doi:10.1099/vir.0.19242-0
16. Zhou, J. H., Li, X. R., Lan, X., Han, S. Y., Wang, Y. N., Hu, Y., & Pan, Q. (2019). The genetic divergences of codon usage shed new lights on transmission of hepatitis E virus from swine to human. *Infect Genet Evol*, 68, 23-29. doi: S1567-1348(18)30511-2 [pii]10.1016/j.meegid.2018.11.024





# Chapter 6

## **Estimating the burden and modeling mitigation strategies of pork-related hepatitis E virus foodborne transmission in representative European countries**

Yunpeng Ji<sup>#</sup>, Pengfei Li<sup>#</sup>, Yueqi Jia, Xiaohua Wang, Qinyue Zheng, Maikel P. Peppelenbosch,  
Zhongren Ma, Qiuwei Pan  
(#: contributed equally)

One Health. 2021, 13:100350



## **Abstract**

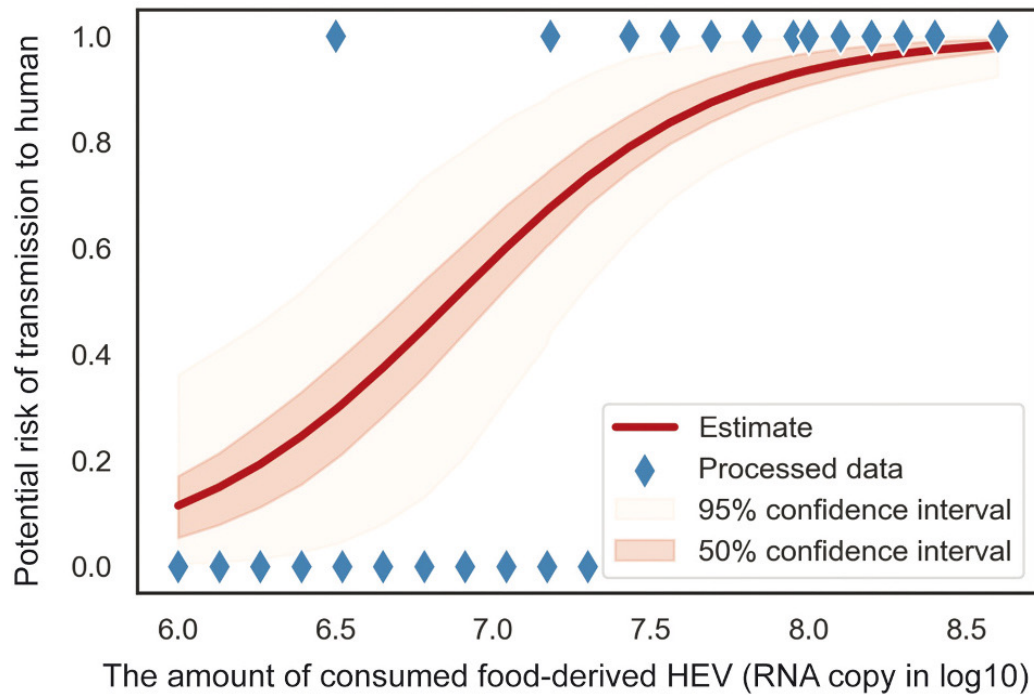
Hepatitis E virus (HEV) is an emerging zoonotic pathogen posing global health burden, and the concerns in Europe are tremendously growing. Pigs serve as a main reservoir, contributing to pork-related foodborne transmission. In this study, we aim to simulate this transmission route and to assess potential interventions. We firstly established a dose-response relationship between the risk of transmission to human and the amount of ingested viruses. We further estimated the incidence of HEV infection specifically attributed to pork-related foodborne transmission in four representative European countries. Finally, we demonstrated a proof-of-concept of mitigating HEV transmission by implementing vaccination in human and pig populations. Our modeling approach bears essential implications for better understanding the transmission of pork-related foodborne HEV and for developing mitigation strategies.

**Keywords:** Hepatitis E virus, cross-species transmission, pork product consumption, European countries, Mathematical modeling

Hepatitis E virus (HEV), a positive-sense single-stranded RNA virus, is a leading cause of acute liver inflammation. It has been estimated that approximately 939 million corresponding to 1 in 8 individuals have ever experienced HEV infection worldwide<sup>[1]</sup>. Among the eight defined genotypes, HEV genotypes 3 and 4 are zoonotic and primarily circulating in developed countries<sup>[2]</sup>. Genotype 3 HEV has been isolated from various mammals including human, swine, wild boar, cattle, goat, deer and rabbit, but pigs are recognized as the main reservoir contributing to transmission to humans<sup>[3]</sup>. HEV has been detected in the liver, gastrointestinal tract, blood, meat and different other organs of pigs. Association of HEV infection with consumption of pork-derived food products has been well-established<sup>[4,5]</sup>, and consuming HEV contaminated food acts as an important route of transmission.

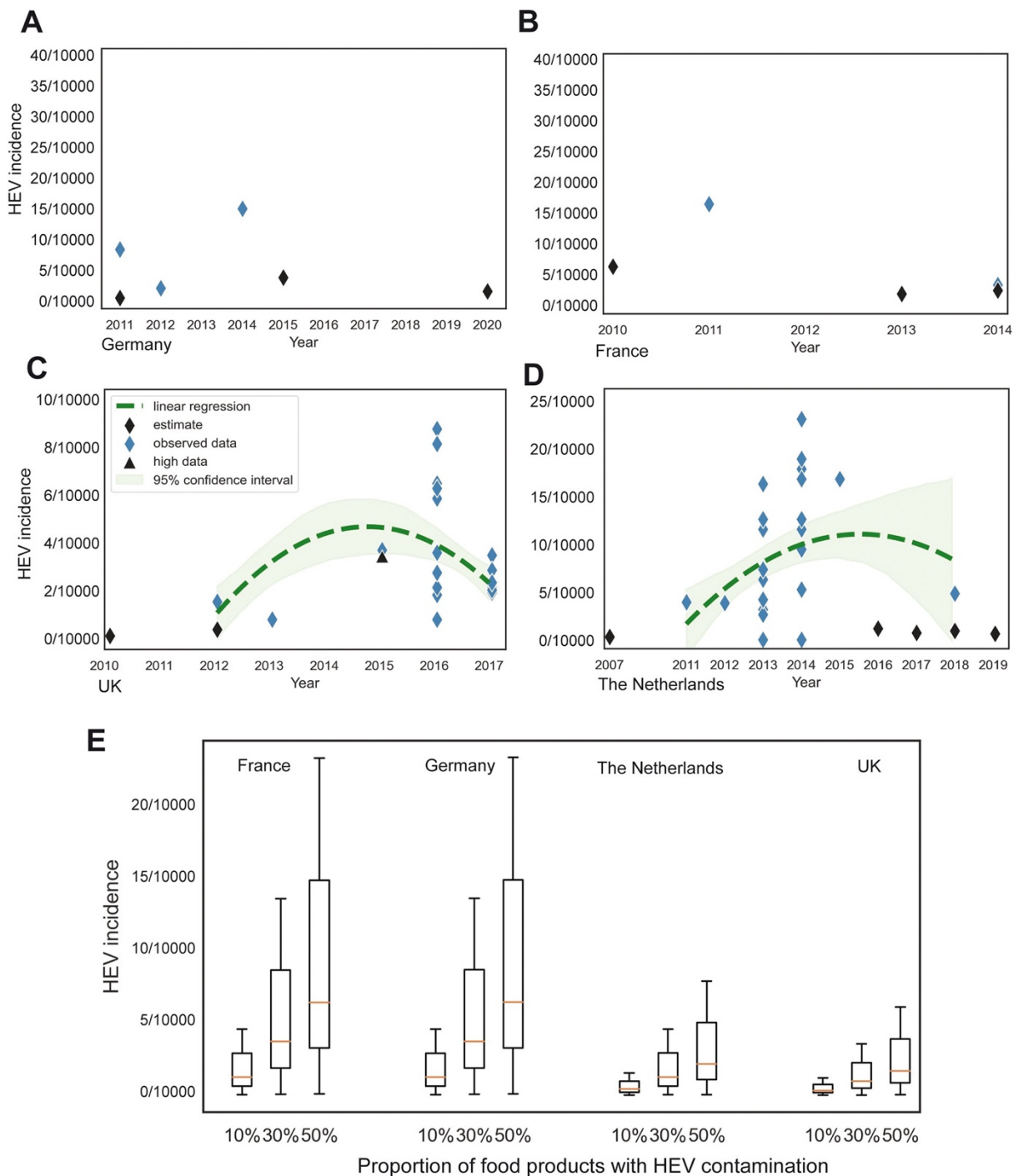
The concerns of health burdens caused HEV infection in Europe are tremendously growing<sup>[6]</sup>. In particular, chronic hepatitis E is frequently reported in Europe, especially in immunocompromised organ transplantation patients, which can be caused by both genotype 3 and 4 HEV<sup>[7, 8]</sup>. Pork-related foodborne transmission is expected to largely contribute to the HEV burden, since consumption of pork-derived food products is common in Europe<sup>[9]</sup>. Given the lack of sufficient real-world data to define the exact risk and contribution of HEV foodborne transmission, this study aimed to estimate the burden of pork-related HEV foodborne transmission in four representative European countries and the effect of potential mitigation strategies by mathematical modeling.

We first attempted to establish the relationship between the risk of HEV infection in human and the amount of acquired HEV through food consumption. By searching published studies, we collected human cases likely to have acquired HEV infection from a food source. HEV genomic RNA copy numbers of the tested food samples in shops or markets where patients habitually visit were also collected. In total, four studies describing 28 HEV RNA-positive human cases<sup>[10-13]</sup>, reported from 2003 to 2014, matched the inclusion criteria. Food products are derived from the meat or organs of animal reservoirs including pigs<sup>[10-12]</sup> and deer<sup>[13]</sup>. By building a logistic dose-response regression model (see details in Supplementary Methods), we estimated the dose-response relation between the risk of transmission to human and the total accrued HEV levels from food (Fig. 1). The probability of infection by oral ingestion of one HEV particle is  $2.5 \times 10^{-9}$  (95% CI  $6.8 \times 10^{-10}$ - $1.5 \times 10^{-8}$ ). The estimated orally ingested amount of HEV at which the probability of infection equals 50% is  $8.1 \times 10^6$  (95% CI  $2.4 \times 10^6$ - $2.0 \times 10^7$ ) viral genomes.



**Figure 1. Logistic dose-dependent relationship between risk of transmission to human and the amount ingested HEV.** HEV genomic RNA copy number is indicated by blue diamond. Model fitting is indicated by the curve, and confidence interval by the shade.

Next, to estimate the contribution of pork-derived foodborne infection, we collected data on HEV incidence from four European countries, Germany, UK, France and the Netherlands. Information on the proportion of pork-derived food contamination with HEV in the food chains were also collected. Technically, we combined the logistic dose-response relationship described above (Fig. 1) and a model of foodborne transmission (Supplementary Methods) for the simulation. The foodborne transmission model describes the process from intake of contaminated pork food to final infection without human-to-human spread. Based on the simulation, the estimated incidence of pork-derived foodborne HEV transmission in the four countries ranges from 1/120784 to 1/2724 (Fig. 2 A-D), from 2001 to 2020. The mean incidence of pork-derived foodborne infection nationwide, based on the non-continuous estimations, is 1/4792 (95%CI 1/559749-1/1679) in Germany, 1/2117 (95% CI 1/366851-1103) in France, 1/29644 (95% CI 1/2693012-1/8028) in UK, and 1/8627 (95% CI 1/1478521-1/4407) in the Netherlands. Correspondingly, the mean number of these HEV cases per year is 17362 in Germany, 31648 in France, 2226 in UK, and 1982 in the Netherlands, respectively.

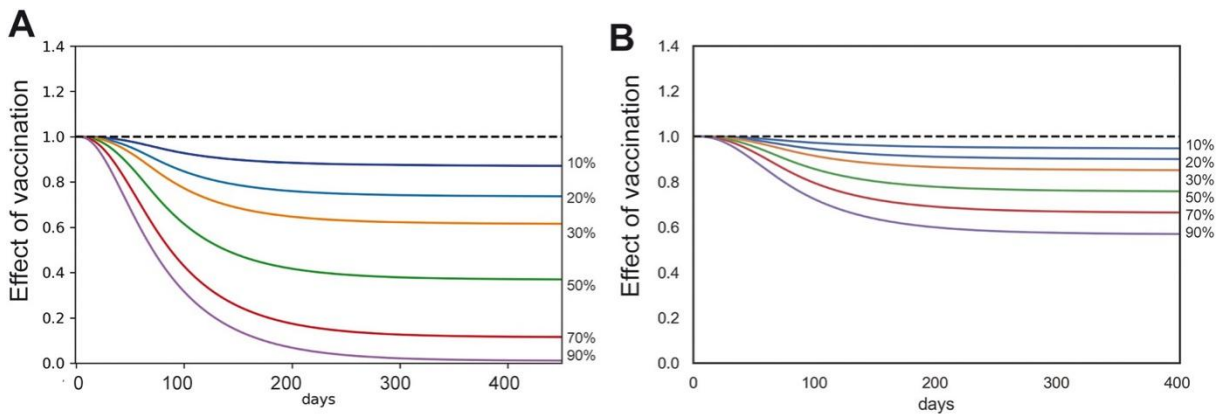


**Figure 2. Estimated HEV incidence of pork-related foodborne infection in four European countries.** Estimated incidence (black triangle) in Germany (A), France (B), UK (C), and the Netherlands (D) based on the yearly available data of food contamination with HEV. The reported overall incidence of in each country is indicated by blue diamond. Trends of incidence in UK and the Netherlands were predicted (green curve) by linear regression incorporating monthly HEV incidence data of specific years. (E). Estimated incidence assuming fixed rates (10%, 30% and 50%) of pork-derived food contamination with HEV.

Because HEV incidence has been reported monthly for UK and the Netherlands<sup>[14,15]</sup>, and we thus extracted these data to visualize the incidence trend in these two countries by a polynomial linear regression method. The modeled HEV incidence in UK and the Netherlands showed a similar shape, reaching the summit around 2015 and then gradually decreasing (Fig. 2C and 2D). We further comparatively simulated the incidence in the four countries with the same levels of pork-derived food contamination with HEV (Fig. 2E). The estimated incidence in France is 1/8069 (95% CI 1/2188-1/732551) (10% contamination), 1/2687 (95% CI 1/243355-1/734) (30%), and 1/1553 (95% CI 1/140265-1/428) (50%), similar to that in Germany but higher than that in the Netherlands and UK.

Our simulation results collectively suggest a substantial burden of pork-related foodborne HEV transmission in Europe. A subsequent question is whether such risk can be prevented through interventions. Effective prevention of HEV transmission likely requires joint efforts from multi-stakeholders. Here, we investigated the potential impact of applying vaccination. A recombinant vaccine, HEV 239, has been licensed in China, which is well-tolerated and effective in the prevention of hepatitis E in the general population<sup>[16]</sup>. Taking Germany as an example (Supplementary Methods), assuming different vaccination coverage rates (from 10% to 90%) implemented in the general human population, the burden of pork-derived HEV foodborne transmission would be reduced coverage-dependently (Fig. 3A). If targeting at a subpopulation with high frequency of pork-related food consumption, the burden would be reduced by 5.3% (with 10% coverage), 10.0% (20%), 14.9% (30%), 24.3% (50%), 33.6% (70%), and 43.1% (90%), respectively (Fig. 3B).

Surveillance and interventions throughout the pork production chain are essential for preventing HEV foodborne transmission. A previous study of Switzerland shows that active interventions in food chains is likely to prevent most human cases<sup>[17]</sup>. However, the current food production and supply chains are diverse, and it has become increasingly difficult to trace the origin of the contaminated products<sup>[18]</sup>. We believe vaccinating pigs is an attractive option to mitigate the HEV burden in human population, although no approved vaccine is available for preventing HEV infection in pigs. The most important effects of vaccinating pigs are expected to reduce the susceptibility of uninfected animals and the contagiousness of animals once get infected<sup>[19]</sup>. Here, we estimated the impact of applying vaccine in pigs on the risk of HEV transmission to humans with the model of Germany (Supplementary Methods). We assumed that the vaccination targets piglets of 10 weeks and is completed in two weeks. We chose this delayed approach of vaccination considering that early vaccination (e.g. for pig of 4 weeks) is likely to be interfered by maternal antibody that produces a strong immunity in newborn piglets.

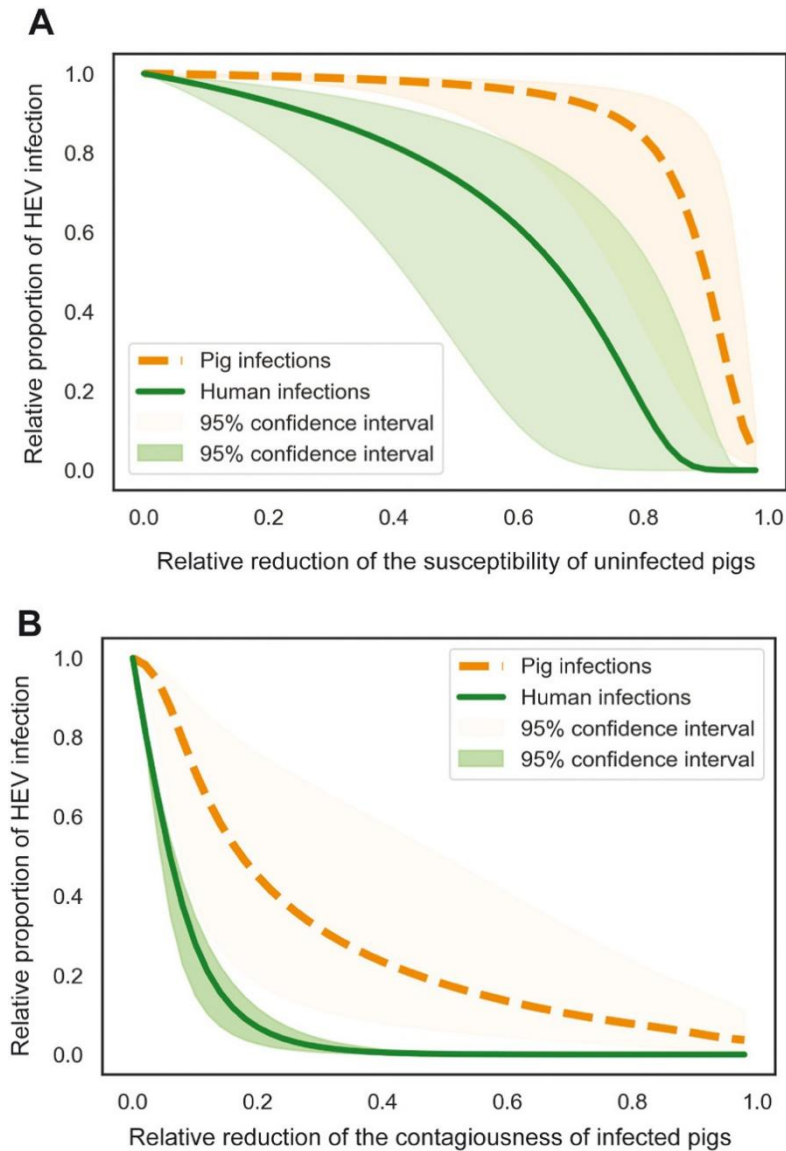


**Figure 3. The impact of vaccinating human population on pork-related foodborne HEV transmission.** The licensed HEV vaccine 293 was assumed to be applied in Germany. A three-dose vaccination would be completed in six months, and the corresponding efficacy after one dose (0 month), two doses (one month) and three doses (six months) is 95.5% (95% CI 66.3-99.4%), 100% (95% CI 9.1-100.0%), and 100% (95% CI 72.1-100.0%), respectively. The coverage rate of vaccination assumingly ranges from 10% to 90%. The effects estimated in general population (A) and a subpopulation with high frequency of pork food consumption (B).

Because HEV vaccine for pigs is not available, we thus theoretically simulated the effects assuming reduction of the susceptibility of uninfected animals (Fig. 4A) and the contagiousness of infected animals (Fig. 4B), respectively. Based on relationship between human risk and levels of virus ingested (Fig.1), we were able to estimate the infections in both pig and human populations. For example, in the first scenario (Fig. 4A), a 78% decrease of susceptibility of uninfected animals would result in a 13.7% (95%CI 3.7%-45.9%) reduction of HEV incidence in pigs at slaughterer age, and a 80% (95%CI 40.9%-99.9%) reduction of human cases. In the second scenario (Fig. 4B), a 14% reduction of the contagiousness of infected animals would lead to a 36% (95%CI 14.5%-68.6%) reduction of HEV incidence in animals at slaughter age, and a 80% (95%CI 72.2%-89.8%) decrease of human cases.

Mapping transmission across the human-animal interface is one of the most important challenges in the control of cross-species infectious diseases in one health. By retrieving published data and examining the drivers of HEV foodborne transmission, we have successfully built a relationship between human risk and the amount of ingested viruses. We observed a dose-dependent effect, meaning that the probability of acquiring infection associates with the quantity of active viral particles ingested orally in a meal. However, in this study, we did not consider the differential susceptibility of different human populations, due to the lack of sufficient data in this respect. The collected data on human cases suggest that susceptibility to HEV infection by consuming pork-derived food appears to be independent of age, but is higher in immunocompromised patients even when the viral titer in contaminated pork product is very lower<sup>[12]</sup>.





**Figure 4. The impact of vaccinating pigs on pork-related foodborne HEV transmission.** (A). The effects of reducing the susceptibility of uninfected animals by vaccination on HEV infections in both pig and human populations were simulated. (B). The effects of reducing the contagiousness of infected animals by vaccination on HEV infections in both pig and human populations.

Although HEV foodborne transmission does not develop a human-to-human spread, consistent dietary exposure of a population still enables constant transmission. The transmission rate is conditional corresponding to the levels of contaminated food and consumption style. In this study, we only selectively estimated the burden of pork-derived foodborne HEV transmission in four European countries, because of the scarcity of available real-world data required by our mathematical models. But our approach would be applicable for any other countries, if the relevant data become available. Importantly, the estimations by our model appear robust. For example, our estimated mean number of cases of pork-derived foodborne infection in Germany based on estimated incidences in the year of 2011, 2015, and

2020 is 17945. This is close to the previous estimation (1500 cases) in Switzerland<sup>[20]</sup>, considering the size of its population is around one tenth of Germany.

Given the substantial burden of pork-derived foodborne HEV transmission as estimated, it is essential to develop effective prevention strategies. We hypothetically simulated the applications of vaccine in both human and pig populations, and showed the effectiveness of both approaches. Nevertheless, a major challenge of applying HEV vaccine in general population is the acceptance, although one vaccine has already been licensed in China. Interestingly, we have demonstrated a proof-of-concept of applying vaccine in high risk population with high frequency of consuming pork products (Fig. 3B). This concept is in line with a large Phase IV clinical trial evaluating the effectiveness in protection of pregnant women by HEV vaccine in Bangladesh (ClinicalTrials.gov Identifier: NCT02759991).

Our results simulating vaccination in pigs quantitatively illustrated the reduction of HEV prevalence in both pig and human populations. However, HEV infection does not affect pig health or the economic performance of swine herds. Intuitively, it would be a challenge to motivate the development of such a vaccine and the subsequent applications by farmers. Therefore, future studies are required to in-depth evaluate the cost-benefit of vaccinating pigs by incorporating the public health consequences on human population.

In summary, this study has established the relation between the risk of transmission to human and the amount of ingested HEV. We estimated the burden of foodborne HEV infection in four European countries, and demonstrated proof-of-concept of mitigating transmission by implementing vaccination in human and pig populations. Our mathematical modeling approach bears essential implications for better understanding the public health burden and developing mitigation strategies for foodborne HEV as well as many other pathogens.

## References

1. P. Li, J. Liu, Y. Li, et al., The global epidemiology of hepatitis E virus infection: A systematic review and meta-analysis, *Liver Int.* 40 (7) (2020) 1516-1528.
2. J.H. Zhou, X.R. Li, X. Lan, et al., The genetic divergences of codon usage shed new lights on transmission of hepatitis E virus from swine to human, *Infect Genet Evol.* 68 (2019) 23-29.
3. N. Kamar, R. Bendall, F. Legrand-Abravanel, et al., Hepatitis E, *Lancet.* 379 (9835) (2012) 2477-2488.
4. O. Wichmann, S. Schimanski, J. Koch, et al., Phylogenetic and case-control study on hepatitis E virus infection in Germany, *J Infect Dis.* 198 (12) (2008) 1732-1741.
5. F. Legrand-Abravanel, N. Kamar, K. Sandres-Saune, et al., Characteristics of autochthonous hepatitis E virus infection in solid-organ transplant recipients in France, *J Infect Dis.* 202 (6) (2010) 835-844.
6. L. The, Growing concerns of hepatitis E in Europe, *Lancet.* 390 (10092) (2017) 334.
7. N. Kamar, J. Selves, J.M. Mansuy, et al., Hepatitis E virus and chronic hepatitis in organ-transplant recipients, *N Engl J Med.* 358 (8) (2008) 811-817.

8. Y. Wang, G. Chen, Q. Pan, J. Zhao, Chronic Hepatitis E in a Renal Transplant Recipient: The First Report of Genotype 4 Hepatitis E Virus Caused Chronic Infection in Organ Recipient, *Gastroenterology*. 154 (4) (2018) 1199-1201.
9. S.R. Pallerla, S. Schembecker, C.G. Meyer, et al., Hepatitis E virus genome detection in commercial pork livers and pork meat products in Germany, *J Viral Hepat*. 28 (1) (2021) 196-204.
10. C. Renou, A.M. Roque-Afonso, N. Pavio, Foodborne transmission of hepatitis E virus from raw pork liver sausage, France, *Emerg Infect Dis*. 20 (11) (2014) 1945-1947.
11. Y. Guillois, F. Abravanel, T. Miura, et al., High Proportion of Asymptomatic Infections in an Outbreak of Hepatitis E Associated with a Spit-Roasted Piglet, France, 2013, *Clin Infect Dis*. 62 (3) (2016) 351-357.
12. M. Riveiro-Barciela, B. Minguez, R. Girones, F. Rodriguez-Frias, J. Quer, M. Buti, Phylogenetic demonstration of hepatitis E infection transmitted by pork meat ingestion, *J Clin Gastroenterol*. 49 (2) (2015) 165-168.
13. S. Tei, N. Kitajima, K. Takahashi, S. Mishiro, Zoonotic transmission of hepatitis E virus from deer to human beings, *Lancet*. 362 (9381) (2003) 371-373.
14. H. Harvala, P.E. Hewitt, C. Reynolds, et al., Hepatitis E virus in blood donors in England, 2016 to 2017: from selective to universal screening, *Euro Surveill*. 24 (10) (2019)
15. B.M. Hogema, M. Molier, M. Sjerps, et al., Incidence and duration of hepatitis E virus infection in Dutch blood donors, *Transfusion*. 56 (3) (2016) 722-728.
16. F.C. Zhu, J. Zhang, X.F. Zhang, et al., Efficacy and safety of a recombinant hepatitis E vaccine in healthy adults: a large-scale, randomised, double-blind placebo-controlled, phase 3 trial, *Lancet*. 376 (9744) (2010) 895-902.
17. P. Ripellino, E. Pianezzi, G. Martinetti, et al., Control of Raw Pork Liver Sausage Production Can Reduce the Prevalence of HEV Infection, *Pathogens*. 10 (2) (2021)
18. I.L.A. Boxman, C.C.C. Jansen, G. Hagele, et al., Porcine blood used as ingredient in meat productions may serve as a vehicle for hepatitis E virus transmission, *Int J Food Microbiol*. 257 (2017) 225-231.
19. J.A. Backer, A. Berto, C. McCreary, F. Martelli, W.H. van der Poel, Transmission dynamics of hepatitis E virus in pigs: estimation from field data and effect of vaccination, *Epidemics*. 4 (2) (2012) 86-92.
20. A. Muller, L. Collineau, R. Stephan, A. Muller, K.D.C. Stark, Assessment of the risk of foodborne transmission and burden of hepatitis E in Switzerland, *Int J Food Microbiol*. 242 (2017) 107-115.

## Supplementary information

### Supplementary Methods

#### 1. Relationship between human risk and HEV levels in food exposure.

We collected human cases likely to have acquired infection from a food source of HEV by searching related literatures or case reports. Data on levels of genomic RNA copy were extracted and prepared. Indirective data attained by testing samples stored in shops or markets where patients habitually visit were also collected in consideration of shortage of information. Over the study period, four studies describing 28 RNA-positive human cases<sup>[1-4]</sup>, reported from 2003 to 2014, matched the inclusion criteria. Food source is products made from the meat or organs of animal reservoirs including pigs (three literatures) and deer (one literature). It is assumed that data of viral genomic levels follow a uniform distribution if merely a range was given.

A logistic dose-response regression model was used to describe the relationship between human risk and HEV genomic levels. The expected number of human cases caused by counts of HEV is  $F(\log(count))$ , where  $F(x)$  is given by

$$F(x) = \frac{1}{1 + \exp(-(x-m)/s)},$$

where  $m$  and  $s$  are the midpoint and scale parameters of the logistic function, respectively. Parameter estimation was performed using maximum likelihood method. To estimate model fitness, we introduced a log-linear model and compared their performances with likelihood ratio test or Akaike information criterion. The robustness of the better model was assessed by bootstrapping. As a result, the logistic regression model offers a better fit than the log-linear model over 97% of the time.

#### 2. Simulating the incidence of pork-derived foodborne infection

We also collected data on HEV incidence of four European countries, namely, Germany, UK, France and the Netherlands. Data of non HEV genotype 3 strain were excluded. Information on proportion of pork-derived food contamination in the food chains were also collected for modeling contribution of foodborne infection to HEV cases (Supplementary table 1).

Specifically, we simulated the burden of foodborne infection by combining the logistic regression model described above and a model of foodborne transmission. The foodborne transmission model describes a process from intake of contaminated pork food to final infection without a human-to-human spread. We introduced a SEIR framework for this foodborne transmission part,

$$\frac{dE}{dt} = \beta_f f_i S - (\mu + \sigma)E,$$

$$\frac{dI}{dt} = \sigma E - (\mu + \gamma)I,$$

where  $E$ ,  $I$ , and  $S$  denote the numbers of exposed, infected and susceptible persons, respectively, while it is assumed that all people in a population is susceptible;  $\mu$  is the per capita death rate,  $\gamma$  is called the removal or recovery rate, and  $\sigma$  is the rate at which individuals move from the exposed to the infectious classes.  $f_i$  represents frequency of meat consumption of class  $i$ , and  $\beta_f$  is the foodborne transmission rate.  $\beta_f$  is calculated from a simplified dose-response equation simulating an oral-to-fecal infection<sup>[5]</sup>,

$$P_{inf} = 1 - e^{-rD} = 1 - e^{-\frac{1}{2}\beta_f},$$

of which  $P_{inf}$  is the probability of infection,  $r$  is the probability of infection per HEV particle ingested and estimated in the logistic regression model above, and  $D$  is amount of HEV dose ingested.  $\beta$  can be further expanded as,

$$\beta_f = 2rD = 2rdmpf_r,$$

where  $d$  denotes mean dose of HEV per gram contaminated food contains,  $m$  is mean account of pork product ingested in a meal,  $p$  is proportion of food contamination, and  $f_r$  is frequency of raw pork-derived food consumption. Required parameters and data are summarized in Supplementary table 1-4.

Supplementary table 1. Parameters involved in the pork-derived foodborne transmission model

Index	Meaning	Value	Source
$\gamma$	recovery rate	1/40	[6]
$\sigma$	rate at which individuals move from the exposed to the infectious classes	1/20	[6]
$r$	probability of infection per HEV particle ingested	$2.5 \times 10^{-9}$	estimated
$m$	account of pork product ingested in a meal (gram)	50-150	[7]
$d$	mean dose of HEV per gram contaminated food contains (genomic copy)	$1 \times 10^4$ ( $110-7 \times 10^4$ )	[8]

Supplementary table 2. Frequency of pork food consumption in four European countries from high to low.

Area	Frequency of meal with pork food	Number or proportion of the consumers	Reference
Germany	everyday	20.60%	[9]
	$\geq$ once a week	53.30%	
	$\geq$ once a month	11.30%	
	Never	14.60%	
France	everyday	11	[10]
	$\geq$ once a week	385	
	$\geq$ once a month	142	
	$\leq$ once a month	31	
UK	$\geq$ 4.5 times a week	35768	[7]
	2.25-4 times a week	102360	
	1.25-2 times a week	95911	
	0.25-1 times a week	170708	
	never	29613	
the Netherlands	everyday	5.40%	[11]
	$\geq$ once a week	35.2%	
	$\geq$ once a month	31.9%	
	$\leq$ once a month	27.5%	

Supplementary table 3. Data on food contamination of specific years in four European countries.

Area	Year	Food sources	Pork-derived food type	Incidence (%)	Reference
Germany	2011	grocery store	Liver	8/200 (4)	[12]
	2015	retail	sausages	13/50 (26)	[13]
				11/50 (22)	
				1/10 (10)	
	2020	grocery & butcher	liver sausage	5/40 (13)	[14]
				liver pate	
pork sausage				0/10 (0)	
Netherlands	2007	butcher	Liver	4/62 (6.5)	[15]
	2016	market	Liver	10/79 (12.7)	[8]
			Liverwurst	70/99 (70.7)	
			liver pate	62/90 (68.9)	
			fresh sausage	0/103 (0)	
			pork chop	0/98 (0)	
			wild boar	0/52 (0)	
			Total	142/521 (27.3)	
	2017	retail	sausage	14/90 (15.6)	[16]
	2018			23/113 (20.4)	
2019	7/52 (13.5)				
France	2010	market	raw liver sausage	7/12 (58.3)	[17]
	2014	market	figatelli & fitone	42/140 (30)	[18]
			liver sausage	49/169 (29)	
			Quenelles	13/55 (25)	
			dried salted livers	7/230 (3)	
2013	market	liver sausage	1/4 (25)	[19]	
UK	2010	retail	Liver	1/76 (1.3)	[20]
	2012	butcher	Sausages	6/63 (9.3)	[21]

Supplementary table 4. Demographic parameters in four European countries

Parameters	Germany	France	UK	the Netherlands
Size of population	83.19×10 <sup>6</sup>	65.40×10 <sup>6</sup>	68.18×10 <sup>6</sup>	17.13×10 <sup>6</sup>
Birth rate (per 1000 people yearly)	9.397	11.042	11.488	10.101
Mortality rate (per 1000 people yearly)	11.3	9	9.1	9.2

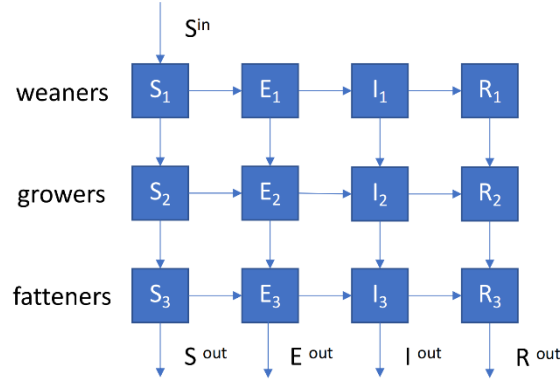
### 3. Simulating the impact of human vaccination

We used the model above to estimate the impact of human vaccination on the risk of pork-derived food exposure. We referred the features and efficacy of the recombinant vaccine HEV 239, a HEV vaccine licensed in China [22]. The three-dose vaccination completes in 6 months, and corresponding efficacy after one dose (0 month), two doses (one month) and three doses (six months) is 95.5% (95% CI 66.3–99.4%), 100% (95% CI 9.1–100.0%), and 100% (95% CI 72.1–100.0%), respectively. Effect of vaccine was estimated by offering different coverage rates for the whole population or a specific group with high frequency of pig meat consumption.

### 4. Simulating the human benefits of pig vaccination

We performed this estimation based on a delayed vaccination, a program applied to piglets of 10 weeks and completing in two weeks. We select delayed vaccination as that early vaccination is likely to be intervened by maternal antibody that produces a strong immunity in newborn piglets but gradually decays. It is assumed that the pig vaccine mainly reduces HEV infectiousness and susceptibility in pigs, respectively.

We quantified the impacts of vaccinations by introducing a HEV transmission in a pig herd. The simulation adopts a SEIR framework and describes how HEV transmits in animals in the pig farm. Specifically, in a wean-to-finish farm, pigs enter the herd at a weaning age of 4 weeks, after which they pass through three different stages: weaners (4-9 weeks of age), growers (10-12 weeks of age) and fatteners (13-26 weeks of age), until slaughter at 26 weeks of age. The pig herd is further subdivided to susceptible, exposed, infectious and recovered animals (Supplementary figure 1). It is assumed that the farm adopts a 5-week batch management system, in which the herd divides into four pig groups, and each move in with a five-week interval. The total cycle is completed after 20 weeks. the size of the herd is assumed to be around 3000.



Supplementary figure 1. In a wean-to-finish farm, a pig herd is divided to weaners (0-4 weeks), growers (5-9 weeks) and fatteners (10-24 weeks), and further subdivided to susceptible, exposed, infectious and recovered animals for modeling.

Reduction of infectiousness is described by factor  $\rho$ , varying between 0 (no effect) and 1 (full effect). The values of transmission factor  $\beta_p$  in pig herd changes according to age periods<sup>[23]</sup>,

$$\beta_p = \begin{cases} \beta_m, & 0 \sim 9 \text{ weeks} \\ \beta_0(1 - \frac{\rho t}{14}), & 10 \sim 12 \text{ weeks}, \\ \beta_0(1 - \rho), & 13 \sim 26 \text{ weeks} \end{cases}$$

where the subscripts 0 denote the parameter values without vaccination,  $t$  is time elapsed since pigs enter the period, and  $\beta_m$  is transmission rate for pigs with maternal immunity from immune mothers. Specifically,

$$\beta_m = \beta_0(1 - e^{-0.06t}) \text{ [5]}$$

The SEIR framework of HEV transmission in pigs can be expanded as,

$$\frac{dE_i}{dt} = \beta_d I_{d_i} S_i + \beta_{ind} I_{ind_i} S_i - (\mu_p + \sigma_p) E_i,$$

$$\frac{dI_i}{dt} = \sigma_p E_i - (\mu_p + \gamma_p) I_i,$$

where  $S_i$ ,  $E_i$  and  $I_i$  denote the number of susceptible, exposed and infectious pigs in patch class  $i$ .  $\beta_d$  is direct transmission rate for piglets in a pen, and  $\beta_{ind}$  is indirect transmission rate, or oral-to-fecal transmission rate, between different pens.  $\mu_p$ ,  $\sigma_p$ , and  $\gamma_p$  are the mortality rate, the rate moving from the exposed to the infectious classes, and the removal or recovery rate of pigs, respectively. The pig herd was initially susceptible and an infectious piglet was populated before model running.  $\beta_{ind}$  can be further expanded as,

$$\beta_{ind} = 2f_p r_p D(1 - c),$$

of which  $f_p$  is the fraction of pens with infectious pigs,  $r_p$  is the probability of pig infection per HEV particle ingested,  $D$  is amount of HEV dose shed by pigs in the nearby pens, and  $c$  is the



clean rate of the room where pigs are kept.

Thus, the adjusted transmission rate of foodborne transmission in human is,

$$\beta_f = 2rd\tilde{m}\tilde{p}f_r,$$

where  $\tilde{p}$  is the new proportion of food contamination, estimated based on the decreased prevalence in vaccinated pigs at slaughter age, and  $\tilde{m}$  is the new mean account of pork product sourced from vaccinated pigs and equals to  $m(1 - p)$ . Generally, vaccine reduces HEV infectiousness in animals by decreasing the viral dose in the animal body<sup>[24]</sup>.

For vaccination mainly reducing viral susceptibility, the SEIR framework of HEV transmission in pigs can be adjusted as,

$$\frac{dE_i}{dt} = \beta_d I_{d_i} S_i + \beta_{ind} I_{ind_i} S_i - (\mu_p + \sigma_p) E_i, 0 \sim 9 \text{ weeks}$$

$$\frac{dE_i}{dt} = \beta_d I_{d_i} S_i (1 - \frac{\rho t}{14}) + \beta_{ind} I_{ind_i} S_i (1 - \frac{\rho t}{14}) - (\mu_p + \sigma_p) E_i, 10 \sim 12 \text{ weeks}$$

$$\frac{dE_i}{dt} = \beta_d I_{d_i} S_i (1 - \rho) + \beta_{ind} I_{ind_i} S_i (1 - \rho) - (\mu_p + \sigma_p) E_i, 13 \sim 26 \text{ weeks}$$

where the transmission factor is independent of factor  $\rho$ . The vaccination would expectedly reduce the prevalence of HEV in animals at slaughter age, and finally influence the fraction of food contamination. Required parameters and data are summarized in Supplementary table 5.

Supplementary table 5. Parameters involved in the pig transmission model

Index	Meaning	Mean value	Source
$\beta_d$	direct transmission rate for piglets in a pen	0.16 (0.082-0.29)	[25]
$\mu_p$	natural death rate of pigs	0.00028 day <sup>-1</sup>	[26]
$\sigma_p$	rate at which pigs move from the exposed to the infectious classes, its reciprocal determines the average latent period	1/6	[27]
$\gamma_p$	the HEV removal or recovery rate of pigs, its reciprocal determines the average infectious period	1/24 (1/39-1/13)	[25,27]
$r_p$	probability of pig infection per HEV particle ingested	$1.3 \times 10^{-9}$	[5]
$D$	amount of HEV contained in daily ingested feces	$1 \times 10^{-7}$	[5]
$c$	the clean rate of the room where pigs are kept	0.5	assumed

## References

1. Christophe Renou, Anne-Marie Roque Afonso, and Nicole Pavio. Foodborne Transmission of Hepatitis E Virus from Raw Pork Liver Sausage, France. *Emerging Infectious Diseases*, 2014, 20(11): 1945-1947.
2. Yvonnick Guillois, Florence Abravanel, Takayuki Miura, Nicole Pavio, Véronique Vaillant, Sébastien Lhomme, Françoise S. Le Guyader, Nicolas Rose, Jean-Claude Le Saux, Lisa A. King, Jacques Izopet, Elisabeth Couturier. High Proportion of Asymptomatic Infections in an Outbreak of Hepatitis E Associated with a Spit-Roasted Piglet, France. *Clinical Infectious Diseases*, 2013, 2016(62): 351-357.
3. Shuchin Tei, Naoto Kitajima, Kazuaki Takahashi, Shunji Mishiro. Zoonotic transmission of hepatitis E virus from deer to human beings. *Lancet*, 2003, 362(2): 371-373.
4. Mar Riveiro-Barciela, MD, Beatriz Miñguez, Rosa Girone's, Francisco Rodriguez-Fri'as, Josep Quer, Mari'a Buti. Phylogenetic Demonstration of Hepatitis E Infection Transmitted by Pork Meat Ingestion. *J Clin Gastroenterol*, 2015, 49(2):165-168.
5. Bouwknecht M, Teunis PF, Frankena K, de Jong MC, de Roda Husman AM. Estimation of the likelihood of fecal-oral HEV transmission among pigs. *Risk Anal*. 2011 Jun;31(6):940-950.
6. Dalton HR, Bendall R, Ijaz S, Banks M. Hepatitis E: an emerging infection in developed countries. *Lancet Infect Dis*. 2008 Nov;8(11):698-709.
7. Thomas J. Littlejohns, Naomi L. Neal, Kathryn E. Bradbury, Hendrik Heers, Naomi E. Allen, Ben W. Turney. Fluid Intake and Dietary Factors and the Risk of Incident Kidney Stones in UK Biobank: A Population-based Prospective Cohort Study, *European Urology Focus*, 2020, 6(4): 752-761.
8. Ingeborg L.A. Boxman, Claudia C.C. Jansen, Geke Hägele, Ans Zwartkruis-Nahuis, Aloys S.L. Tijmsma, Harry Vennema. Monitoring of pork liver and meat products on the Dutch market for the presence of HEV RNA. *International Journal of Food Microbiology*, 2019, 296: 58-64.
9. Wim Verbeke, Federico J.A. Pérez-Cueto, Klaus G. Grunert. To eat or not to eat pork, how frequently and how varied? Insights from the quantitative Q-PorkChains consumer survey in four European countries. *Meat Science*. 2011, 88(4): 619-626.
10. T.M Ngapo, J.-F Martin, E Dransfield. Consumer choices of pork chops: results from three panels in France. *Food Quality and Preference*, 2004, 15(4):349-359.
11. Yung Hung, Theo M. de Kok, Wim Verbeke. Consumer attitude and purchase intention towards processed meat products with natural compounds and a reduced level of nitrite. *Meat Science*, 2016, 121: 119-126.
12. Jürgen J. Wenzel, Julia Preiß, Mathias Schemmerer, Barbara Huber, Annelie Plentz, Wolfgang Jilg. Detection of hepatitis E virus (HEV) from porcine livers in Southeastern Germany and high sequence homology to human HEV isolates, *Journal of Clinical Virology*, 2011, 52(1): 50-54.
13. Kathrin Szabo, Eva Trojnar, Helena Anheyer-Behmenburg, Alfred Binder, Ulrich Schotte, Lüppo Ellerbroek, Günter Klein, Reimar John. Detection of hepatitis E virus RNA in raw sausages and liver sausages from retail in Germany using an optimized method. *International Journal of Food Microbiology*, 2015, 215: 149-156.
14. Pallerla SR, Schembecker S, Meyer CG, Linh LTK, John R, Wedemeyer H, Bock CT, Kreamsner PG, Velavan TP. Hepatitis E virus genome detection in commercial pork livers and pork meat products in Germany. *J Viral Hepat*. 2021 Jan;28(1):196-204.
15. Martijn Bouwknecht, Froukje Lodder- Verschoor, Wim H. M. Van der Poel, Saskia A. Rutjes, Ana Maria de Roda Husman. Hepatitis E Virus RNA in Commercial Porcine Livers in The Netherlands. *J Food Prot* 1 December 2007; 70 (12): 2889–2895.
16. Ingeborg L.A. Boxman, Claudia C.C. Jansen, Ans J.T. Zwartkruis-Nahuis, Geke Hägele, Nils P. Sosef, René A.M. Dirks. Detection and quantification of hepatitis E virus RNA in ready to eat raw pork sausages in the Netherlands. *International Journal of Food Microbiology*, 2020, 333:108791.

17. Philippe Colson, Patrick Borentain, Benjamin Queyriaux, Mamadou Kaba, Valérie Moal, Pierre Gallian, Laurent Heyries, Didier Raoult, René Gerolami, Pig Liver Sausage as a Source of Hepatitis E Virus Transmission to Humans, *The Journal of Infectious Diseases*, 2010, 202(6), 825–834.
18. Pavio N, Merbah T, Thébault A. Frequent hepatitis E virus contamination in food containing raw pork liver, France. *Emerg Infect Dis*. 2014;20(11):1925-1927.
19. Berto A, Grierson S, Hakze-van der Honing R, Martelli F, John R, Reetz J, Ulrich RG, Pavio N, Van der Poel WHM, Banks M. Hepatitis E virus in pork liver sausage, France. *Emerg Infect Dis*. 2013;19(2):264-266.
20. Banks M, Martelli F, Grierson S, Fellows HJ, Stableforth W, Bendall R, Dalton HR, 2010. Hepatitis E virus in retail pig livers. *Veterinary Record*, 166, 29.
21. Berto A, Martelli F, Grierson S, Banks M. Hepatitis E virus in pork food chain, United Kingdom, 2009-2010. *Emerging Infectious Diseases*, 2012, 18, 1358-1360.
22. Zhu FC, Zhang J, Zhang XF, Zhou C, Wang ZZ, Huang SJ, Wang H, Yang CL, Jiang HM, Cai JP, Wang YJ, Ai X, Hu YM, Tang Q, Yao X, Yan Q, Xian YL, Wu T, Li YM, Miao J, Ng MH, Shih JW, Xia NS. 2010. Efficacy and safety of a recombinant hepatitis E vaccine in healthy adults: a large-scale, randomized, double-blind placebo-controlled, phase 3 trial. *Lancet* 376:895–902.
23. J.A. Backer, A. Bertoa, C. McCreary, F. Martelli, W.H.M. van der Poel. Transmission dynamics of hepatitis E virus in pigs: Estimation from field data and effect of vaccination. *Epidemics* 4 (2012) 86–92.
24. Chase-Topping M, Xie J, Pooley C, Trus I, Bonckaert C, Rediger K, Bailey RI, Brown H, Bitsouni V, Barrio MB, Gueguen S, Nauwynck H, Doeschl-Wilson A. New insights about vaccine effectiveness: Impact of attenuated PRRS-strain vaccination on heterologous strain transmission. *Vaccine*. 2020, 23;38(14):3050-3061.
25. Backer JA, Berto A, McCreary C, Martelli F, van der Poel WH. Transmission dynamics of hepatitis E virus in pigs: estimation from field data and effect of vaccination. *Epidemics*. 2012, 4(2):86-92.
26. Bouwknecht M, Frankena K, Rutjes SA, Wellenberg GJ, de Roda Husman AM, van der Poel WH, de Jong MC. Estimation of hepatitis E virus transmission among pigs due to contact-exposure. *Vet Res*. 2008, 39(5):40.
27. Reynolds JJ, Torremorell M, Craft ME. Mathematical modeling of influenza A virus dynamics within swine farms and the effects of vaccination. *PLoS One*. 2014, 27;9(8): e106177.



# Chapter 7

## **Prevalence and clinical features of hepatitis E virus infection in pregnant women: A large cohort study in Inner Mongolia, China**

Xiaoxia Ma<sup>#</sup>, Yunpeng Ji<sup>#</sup>, Li Jin, Zulqarnain Baloch, Derong Zhang, Yijin Wang, Qiuwei Pan,  
Zhongren Ma  
(#: contributed equally)

Clinics and Research in Hepatology and Gastroenterology. 2021, 45(4):101536



**Abstract**

*Background and aim.* Hepatitis E virus (HEV) infection causes severe maternal and fetal outcomes in pregnant women. These patients are exclusively from resource-limited regions with genotype 1 HEV infection, but not from western countries with genotype 3 prevalence. Since the circulating strains in China have evolved from the waterborne genotype 1 to the zoonotic genotype 4 HEV in the past decades, this study aims to evaluate the prevalence and clinical features of HEV infection in a large cohort of pregnant women in Inner Mongolia, China. *Methods.* A total of 3278 pregnant women who visited the Inner Mongolia Maternal and Child Care hospital during 2018 were enrolled. Serum samples were examined for anti-HEV IgG and anti-HEV IgM antibodies using ELISA. Demographic information, results of clinical biochemical tests, maternal and neonatal outcomes were collected. *Results.* Among the recruited 3278 pregnant women, 6.0% were anti-HEV IgG antibody positive, 0.3% were anti-HEV IgM antibody positive and 0.3% were positive for both anti-HEV IgG and anti-HEV IgM antibodies. HEV viral RNA was not detected. Pregnant women with recent/ongoing HEV infection indicated by anti-HEV IgM positivity have slightly higher ALT level, and potential risk of developing hyperlipidemia, preterm delivery and neonatal jaundice. *Conclusions.* These findings indicated that HEV infection is associated with a possible increase in adverse maternal, fetal and neonatal outcomes in our cohort. Thus, the burden of HEV infection in pregnant women in China appears distinct from resource-limited regions and western countries. Nevertheless, future studies are required to confirm and extend our findings.

**Keywords:** Hepatitis E virus, Pregnant women, Sero-Prevalence, Outcomes, Risk factors, China

## Introduction

Hepatitis E virus (HEV) is a positive-sense, single-stranded RNA virus that represents as a leading cause of acute viral hepatitis globally. There is only one single serotype of HEV, but classified into eight genotypes. Anti-HEV IgM antibodies can be detected shortly after the infection followed by the appearance of anti-HEV IgG antibodies. The presence of anti-HEV IgM antibodies indicates recent/ongoing infection, whereas the presence of IgG antibodies alone indicates previous infection<sup>[1]</sup>. Genotypes 1 and 2 HEV exclusively infecting humans are found mainly in developing countries, and responsible for many water-borne outbreaks. In contrast, HEV genotypes 3 and 4 are zoonotic and responsible for sporadic infections mainly in developed countries<sup>[2-5]</sup>.

Although HEV infection is usually acute and self-limiting in the general population, infection in pregnant women can cause severe outcomes including fulminant hepatic failure (FHF) with fatality rate up to 30%<sup>[6]</sup>. FHF resulted from fulminant hepatitis is the main cause of high death rate among pregnant women, especially in the third trimester<sup>[7-9]</sup>. These patients are almost exclusively from recourse-limited regions with genotype 1 HEV infection, but not from western countries with prevalence of genotype 3 HEV. Recently, studies from Namibia and China have indicated that genotype 2 and 4 HEV can also negatively affect the clinical outcomes of pregnant women<sup>[4,10]</sup>.

China is an endemic area for HEV. In 1991, there was a large outbreak occurred in Xinjiang province, caused by genotype 1 HEV. The attack rate in pregnant women was significantly higher than that of the non-pregnant population. This resulted in severe clinical outcomes in pregnant patients, with fatality rate of 5.88% and abortion rate of 17.64%<sup>[6,11]</sup>. Of note, the epidemiology of HEV is dramatically changing in China. Currently, genotype 4 instead of 1 is circulating in the population. It is zoonotic and supposed to be more pathogenic<sup>[12]</sup>. This study aims to investigate the prevalence and clinical features of HEV in a large hospital-based population cohort of pregnant women in Inner Mongolian. Inner Mongolian is one of the five autonomous regions for ethnic minorities in China, including Han, Mongolian, Hui, Ewenki Manchu and Daur. It is a less developed area with animal husbandry as the mainstay of economy. Therefore, this region is ideal for studying diseases associated with genetics, environmental factors and life styles.



## Materials and methods

### Study design

Pregnant women who visited the Inner Mongolia Maternal and Child Care hospital from January to December, 2018 were enrolled. Demographic information, results of clinical biochemical tests, maternal and neonatal outcomes were collected from the hospital's computerized pregnancy information database. A control non-pregnant cohort of 290 participants who attended physical examination in Inner Mongolia was included. The records of all participants were anonymously analyzed. All participants provided informed consent that allow future testing of archived bio-samples including leftover serum. This study was approved by ethical committee of Inner Mongolia Maternal and Child Care Hospital, and Northwest Minzu University, China.

### Laboratory tests of HEV infection

Serum samples of pregnant women were screened for the presence of anti-HEV IgG and IgM antibodies using commercially-available enzyme immunoassay kits (Wantai Biological Pharmacy Enterprise, Beijing, China). Every 30 samples were pooled for detecting HEV viral RNA by conventional qRT-PCR. All anti-HEV IgM positive samples were again tested for HEV RNA using commercially available Fluorescence Quantitative PCR kit (Beijing Kinghawk Pharmaceutical CO., LTD, Beijing, China) according to the manufacturer's instructions.

The anti-HEV IgM antibody positive and HEV RNA negative samples were further tested using nested RT-PCR. Total RNA was extracted from the serum using the QIAamp Viral RNA mini-kit (Qiagen, Germany) according to the manufacturer's instructions. cDNA was synthesized from 8 µl purified RNA using 2 µl reverse transcriptase (promega, USA). A nest-PCR was carried out to produce a 348-nucleotide amplicon from HEV open reading frame 2 (ORF2). Briefly, the first round PCR was in a 20 µl reaction, including 5 µl cDNA, 10 µl Green Taq Mix (Takara, Japan), 1 µl primers (Forward, 5-AATTATGCYAGTAYCGRGTTG-3. Reverse, 5-CCCTTA(G) TCC(T)TGCTGA (C)GCATTCTC-3) and 4 µl ddH<sub>2</sub>O. The PCR parameters including a denaturation step at 94 °C for 5 min, followed by 35 cycles of denaturation for 30 s at 94 °C, annealing for 30 s at 42 °C, extension for 50 s at 72 °C, and a final incubation at 72 °C for 5 min. The second-round PCR was performed using 5 µl of first-round PCR product and internal primers (forward, 5-GTT(A)ATGCTT(C)TGCATA(T)CATGGCT-3. reverse, 5-AGCCGACGAAATCAATTCTGTC-3), and parameters were the same with that in the first-round, except shortening the extension time to 30 s.

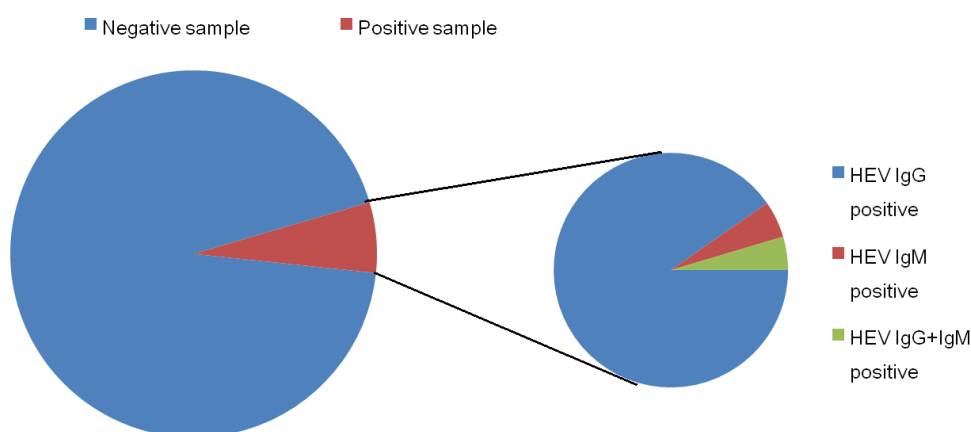
### Statistical analysis

SPSS version 13.0 (Chicago, IL, USA) was used to perform all the statistical analysis. Statistical analysis was performed by univariate analysis. Significance was set at  $P < 0.05$ .

## Results

### HEV sero-prevalence

In total, 3278 pregnant women were enrolled for HEV screening. The overall sero-prevalence of anti-HEV IgG and IgM antibodies was 6.62% (217/3278). 6.0% (196) were anti-HEV IgG antibody positive IgM antibody negative, indicating past HEV infection. 0.62% (21) were anti-HEV IgM antibody positive, including 10 anti-HEV IgM and IgG positive and 11 anti-HEV IgM positive IgG negative, indicating recent/ongoing infection (Fig. 1). In the non-pregnant cohort, 13 out of 290 samples (4.5%) were positive for anti-HEV IgG, but none was positive for anti-IgM antibody. Detection of HEV RNA in patient blood and stool is essential to conform the active infection<sup>[1]</sup>. However, we were not able to detect viral RNA by pool of every 30 serum samples and in any of the individual anti-HEV IgM antibody positive serum samples, whereas stool samples were not available for these participants.



**Fig 1. Study population** .3278 women were recruited and screened. Among these samples, 6.0% were anti-HEV IgG positive, 0.3% were anti-HEV IgM positive and 0.3% were positive for both anti-HEV IgG and anti-HEV IgM respectively

### Demographic characteristics and potential association with recent/ongoing HEV infection

We next performed analysis on the characteristics of HEV sero-positive pregnant women (Table 1). We focused on the nested-cohort of the 217 sero-positive participants, and this enabled us to compare the differences between recent/ongoing with past infection as the

control. All 217 patients were classified into four groups according to the status of anti-HEV IgG or/and IgM antibody positivity. There is no significant difference regarding pregnant trimester, mean age, childbearing history, residence area and profession. Interestingly, ethnicity in particular Tujia people appears to have higher rate of recent/ongoing infection (IgM antibody positivity). But this sub-population is too small to draw firm conclusion.

**Table 1 Demographic characteristics according to anti-HEV IgG or/and IgM antibody positivity and factors potentially associated with recent/ongoing HEV infection.**

Variables	IgM <sup>+</sup> IgG <sup>-</sup>		IgM <sup>+</sup> IgG <sup>+</sup>		(IgM <sup>+</sup> IgG <sup>+</sup> ) & (IgM <sup>-</sup> IgG <sup>+</sup> )		P value
	(n=11)	(n=10)	(n=21)	(n=196)	(IgM <sup>+</sup> IgG <sup>-</sup> ) vs (IgG <sup>+</sup> IgM <sup>-</sup> )	(IgM <sup>+</sup> IgG <sup>+</sup> ) vs (IgG <sup>+</sup> IgM <sup>-</sup> )	
Trimester, n (%)							
First trimester	8 (72.7)	9 (90)	17 (81)	103 (52.6)	0.497	0.338	0.212
Second trimester	3 (27.3)	1 (10)	4 (19)	93 (47.4)	0.398	0.081	0.093
Mean age (SD), y	28 (2.4)	27 (3.8)	28 (3.2)	28 (3.4)	0.875	0.294	0.423
Childbearing history (SD)							
Mean gravida	0.9 (0.3)	0.67 (0.7)	0.82 (0.5)	0.74 (0.8)	0.542	0.835	0.719
Mean para	0.2 (0.4)	0.17 (0.3)	0.18 (0.4)	0.19 (0.4)	0.965	0.879	0.935
Ethnicity, n (%)							
Han	7 (63.6)	7 (70)	14 (66.7)	160 (81.6)	0.614	0.76	0.574
Mongolian	3 (27.2)	1 (10)	4 (19)	26 (13.3)	0.283	0.791	0.534
Hui	1 (9.1)	0 (0)	1 (4.8)	5 (2.6)	0.234		0.571
Man	0 (0)	1 (10)	1 (4.8)	4 (2)		0.132	0.445
Tujia	0 (0)	1 (10)	1 (4.8)	1 (0.5)		0.005	0.059
Residence area, n (%)							
Urban	11 (100)	9 (90)	20 (95.2)	175 (89.3)	0.796	0.987	0.845
Rural	0 (0)	1 (10)	1 (4.8)	21 (10.7)		0.949	0.428
Profession, n (%)							
Job	6 (54.5)	7 (70)	13 (61.9)	156 (79.6)	0.464	0.799	0.495
No job	5 (45.6)	3 (30)	8 (38.1)	40 (20.4)	0.148	0.569	0.16

### **Blood biochemical profiling**

Regarding blood biochemical parameters, there is no significant difference for mean hemoglobin, mean platelet count, mean serum albumin level, median aspartate aminotransferase (AST) level and median alkaline phosphatase level. Interestingly, median alanine aminotransferase (ALT) level among anti-HEV IgM positive group is significantly higher compared with that in the IgG antibody positive IgM antibody negative population (Table 2). Because the level of ALT increase is very mild, it is difficult to conclude whether this is related to possible induction of liver injury by HEV infection.

### **Pregnancy complications**

To assess the potential impact of HEV infection on maternal outcome, we analyzed eight variables including postpartum hemorrhage, premature rupture of membranes, maternal mild anemia, intrauterine asphyxia, pregnant hypertension, pregnant hyperlipidemia, gestational diabetes mellitus and severe preeclampsia. No significant differences were observed between anti-HEV IgM antibody positive and the control group, except for pregnant hyperlipidemia at the second trimester ( $P = 0.03$ ) (Table 3). None of the sero-positive pregnant women died in our study.

Table 2 Blood biochemical profiles according to anti-HEV IgG or/and IgM antibody positivity

Variables	IgM <sup>+</sup> IgG <sup>-</sup> (n=11)	IgM <sup>+</sup> IgG <sup>+</sup> (n=10)	(IgM <sup>+</sup> IgG <sup>+</sup> & IgM <sup>-</sup> IgG <sup>+</sup> )	IgG <sup>-</sup> IgM <sup>-</sup> (n=196)	P value		
					(IgM <sup>+</sup> IgG <sup>-</sup> ) vs (IgG <sup>+</sup> IgM <sup>-</sup> )	(IgM <sup>+</sup> IgG <sup>+</sup> ) vs (IgG <sup>+</sup> IgM <sup>+</sup> )	
Mean hemoglobin level (SD), g/l	129 (8.5)	125 (6.3)	127 (7.9)	127 (10.8)	0.668	0.565	0.985
Median leukocyte count (range), cells ×10 <sup>9</sup> /l	8.9 (5.6-15.3)	8.1 (5.5-12.7)	10.9 (5.5-15.3)	9.8 (4.4-15.6)	0.364	0.253	0.135
Mean platelet count (SD), cells ×10 <sup>9</sup> /l	220.9 (55)	238.1 (49)	227.6 (53)	259.9 (48)	0.097	0.651	0.126
Mean serum bilirubin level (SD)	6.6 (3.0)	8.73 (4.1)	7.6 (3.7)	6.93 (3.8)	0.809	0.226	0.521
Mean serum albumin level (SD), g/l	41.6 (2.1)	41 (4.7)	41.3 (3.6)	42.14 (2.7)	0.561	0.304	0.276
Median alanine aminotransferase level (range), U/l	18.5 (12.2-22.9)	19.9 (12.6-23.5)	18.7 (12.2-23.5)	14.1 (4.2-47)	<b>0.028</b>	<b>0.049</b>	<b>0.02</b>
Median aspartate aminotransferase level (range), U/l	16.2 (10.5-25)	17.5 (12.2 - 30.2)	17.5 (10.5-30.2)	16.1 (9.2-51.9)	0.479	0.649	0.712
Median alkaline phosphatase level (range), U/l	47 (28.3-67.5)	49.1 (39.6- 164.6)	49.1 (28.3- 164.6)	51.6 (19.3- 100.4)	0.479	0.699	0.417

**Table 3 Comparison of the risk of obstetric complications**

Variables	IgM <sup>+</sup> IgG <sup>-</sup> (n=11)	IgM <sup>+</sup> IgG <sup>+</sup> (n=10)	(IgM <sup>+</sup> IgG <sup>-</sup> & IgM <sup>+</sup> IgG <sup>+</sup> ) (n=21)	IgG <sup>+</sup> IgM <sup>-</sup> (n=196)	P value		
					(IgM <sup>+</sup> IgG <sup>-</sup> ) vs (IgG <sup>+</sup> IgM <sup>-</sup> )	(IgM <sup>+</sup> IgG <sup>-</sup> ) vs (IgG <sup>+</sup> IgM <sup>+</sup> )	(IgM <sup>+</sup> IgG <sup>-</sup> & IgM <sup>+</sup> IgG <sup>+</sup> ) vs (IgG <sup>+</sup> IgM <sup>+</sup> )
Postpartum hemorrhage, n/n (%)	0 (0)	0 (0)	0 (0)	2/196 (1)	0.391	0.323	0.876
First trimester	0 (0)	0 (0)	0 (0)	2/103 (1.9)	0.56	0.61	0.987
Second trimester	0 (0)	0 (0)	0 (0)	0/93 (0)		0.115	0.86
Premature rupture of membranes, n/n (%)	1/11 (9.1)	4/10 (40)	5/21 (23.8)	43/196 (22)	0.391	0.323	0.876
First trimester	1/8 (12.5)	3/9 (33.3)	4/17 (23.5)	24/103 (23.3)	0.56	0.61	0.987
Second trimester	0/3 (0)	1/1 (100)	1/4 (25)	19/93 (20.4)		0.115	0.86
Maternal mild anemia (Hb 90-110 g/dl), n/n (%)	5/11 (45.5)	1/10 (10)	6/21 (28.6)	51/196 (26)	0.315	0.35	0.848
First trimester	5/8 (62.5)	1/9 (11.1)	6/17 (35.3)	35/103 (34)	0.306	0.274	0.941
Second trimester	0 (0)	0 (0)	0 (0)	16/93 (17.2)			
Intrauterine asphyxia, n/n (%)	1/11 (9.1)	0 (0)	1/21 (4.8)	11/196 (5.6)	0.655		0.878
First trimester	1/8 (12.5)	0 (0)	1/17 (5.9)	6/103 (5.8)	0.494		0.993
Second trimester	0 (0)	0 (0)	0 (0)	5/93 (5.4)			
Pregnant hypertension, n/n (%)	0 (0)	0 (0)	0 (0)	4/196 (2.0)			
First trimester	0 (0)	0 (0)	0 (0)	3/103 (2.9)			
Second trimester	0 (0)	0 (0)	0 (0)	1/93 (1.1)			
Pregnant hypertyphidemia, n/n (%)	0 (0)	1/10 (10)	1/21 (4.8)	3/196 (1.5)		0.073	0.31
First trimester	0 (0)	0 (0)	0 (0)	2/103 (1.9)			
Second trimester	0 (0)	1/1 (100)	1/4 (25)	1/93 (1.1)		<0.001	0.03
Gestational diabetes mellitus, n/n (%)	0 (0)	0 (0)	0 (0)	12/196 (6.1)			
First trimester	0 (0)	0 (0)	0 (0)	5/103 (4.9)			
Second trimester	0 (0)	0 (0)	0 (0)	7/93 (7.5)			
Severe preeclampsia, n/n (%)	0 (0)	0 (0)	0 (0)	2/196 (1.0)			
First trimester	0 (0)	0 (0)	0 (0)	2/103 (1.9)			
Second trimester	0 (0)	0 (0)	0 (0)	0 (0)			

### **Fetal and neonatal outcomes**

To finally assess the impact on fetal and neonatal outcomes, eight variables were analyzed, including preterm delivery, mean Apgar values, precipitate labour, caesarean section, neonatal weight, neonatal death, neonatal retinal hemorrhage and neonatal jaundice. The rates of preterm delivery and neonatal jaundice appear to be higher in anti-HEV IgM antibody positive compared to the anti-HEV IgG antibody positive IgM antibody negative control group (Table 4).



**Table 4 Comparison of fetal/neonatal outcomes**

Variables	IgM <sup>+</sup> IgG <sup>-</sup>	IgM <sup>+</sup> IgG <sup>+</sup>	(IgM <sup>+</sup> IgG <sup>-</sup> ) & (IgM <sup>+</sup> IgG <sup>+</sup> )	IgG <sup>+</sup> IgM <sup>-</sup>	P value		
	(n=11)	(n=10)	(n=21)	(n=196)	(IgM <sup>+</sup> IgG <sup>-</sup> ) vs (IgG <sup>+</sup> IgM <sup>-</sup> )	(IgM <sup>+</sup> IgG <sup>+</sup> ) vs (IgG <sup>+</sup> IgM <sup>-</sup> )	(IgM <sup>+</sup> IgG <sup>-</sup> & IgM <sup>+</sup> IgG <sup>+</sup> ) vs (IgG <sup>+</sup> IgM <sup>-</sup> )
Preterm delivery (<37w)	1/11 (9.1)	2/10 (20)	3/21 (14.3)	7/196 (3.6)	0.385	0.026	0.041
Mean Apgar values (SD)	9.4 (0.5)	9.2 (0.4)	9.3 (0.4)	9.1 (0.4)	0.331	0.758	0.287
Precipitate labour; n (%)	0 (0)	0 (0)	0 (0)	3/196 (1.5)	-	-	-
Neonatal deaths; n (%)	0 (0)	0 (0)	0 (0)	0 (0)	-	-	-
Caesarean section; n/n (%)	1/11 (9.1)	2/10 (20)	3/21 (14.3)	47/196 (24)	0.341	0.818	0.412
First trimester	1/8 (12.5)	1/9 (11.1)	2/17 (11.8)	28/103 (27.2)	0.462	0.332	0.269
Second trimester	0 (0)	1/1 (100)	1/4 (25)	19/93 (20.4)	0.223	0.223	0.86
Neonatal weight (SD), g	3490 (504)	3333 (302)	3404 (414)	3356 (422)	0.496	0.896	0.719
Neonatal retinal hemorrhage; n/n (%)	1/11 (9.1)	1/10 (10)	2/21 (9.5)	6/196 (3.1)	0.31	0.267	0.361
First trimester	1/8 (12.5)	1/9 (11.1)	2/17 (11.8)	5/103 (4.9)	0.397	0.46	0.299
Second trimester	0 (0)	0 (0)	0 (0)	1/93 (1.1)	-	-	-
Neonatal jaundice; n/n (%)	3/11 (18.2)	1/10 (10)	4/21 (19)	9/196 (4.6)	0.07	0.47	0.018
First trimester	3/8 (25)	1/9 (11.1)	4/17 (23.5)	7/103 (6.8)	0.016	0.659	0.055
Second trimester	0 (0)	0 (0)	0 (0)	2/93 (2.2)	-	-	-

## Discussion

The overall sero-prevalence of HEV was 6.62% in our large cohort of over 3,000 pregnant women in Inner Mongolia China. 6.0% of the samples were anti-HEV IgG antibody positive, 0.3% were anti-HEV IgM antibody positive, and 0.3% were positive for both anti-HEV IgG and anti-HEV IgM antibodies. In our non-pregnant control cohort, we found a prevalence rate of anti-HEV IgG antibody as 4.5%. This is comparable but slightly lower than that in pregnant women. Of note, the average age ( $18.84 \pm 0.78$  years old) of the control group is slightly younger than the average pregnant age (about 26) in China (Supplementary Table 1). It is well-recognized that age is associated with anti-HEV IgG prevalence<sup>[1]</sup>, but we were not able to find a perfectly age-matched control cohort.

The sero-prevalence in our pregnant cohort is relatively lower than the average level (11.66-41.71% for IgG and 0.43—2.8 % for IgM) in the general population in China<sup>[13]</sup>, suggesting that pregnancy itself is likely not a risk factor of HEV infection. It is also lower than the average level of Chinese pregnant women reported in other studies, which is 10.24-16.2% for IgG and 2.56-3.2% for IgM<sup>[14,15]</sup>. Our overall sero-prevalence rate is much lower than the results of similar studies done in Egypt (84.3%), Sudan (41%), India (33.6%) and other developing countries<sup>[16-18]</sup>. In these endemic regions, HEV is predominantly transmitted via the faecal-oral route due to water contamination. This is especially common in rural areas having poor sanitation<sup>[19]</sup>. In China, local sanitation standards have been greatly improved during the past decades. With changes of the circulating genotype, HEV is mainly transmitted through contaminated food and causes sporadic infection by genotype 4 HEV<sup>[20,21]</sup>. We failed to detect HEV RNA, and therefore were not able to confirm the genotype. However, genotype 4 HEV has been widely detected in neighboring regions of Inner Mongolia<sup>[22-24]</sup>.

The low prevalence of HEV in Inner Mongolia may be associated with their dietary habit and life style. This region has large sheep and cow, but not swine farming. The local people mainly feed on mutton and milk products, whereas pigs are the main reservoir for food-borne transmission of HEV, although the host for HEV is not only restricted to swine<sup>[2]</sup>. We evaluated several factors including trimester, mean age, childbearing history, residence area and profession, but these were not associated with HEV infection. There may be potential relation with ethnicity that Tujia group appears to have higher rate of recent/ongoing infection. However, Tujia is not the main ethnicity in Inner Mongolia China and there is very limited patient number in our study to draw firm conclusion.

Severe clinical outcomes in pregnant women infected with HEV have been exclusively recorded from Asian and African countries, in particular resource-limited regions. In outbreak settings, HEV infection results in worse maternal and fetal outcomes, including jaundice,

malaise, anorexia, hepatalgia, nausea, vomiting and lethargy<sup>[6,25-27]</sup>. In hospital-based settings, pregnant women showing jaundice due to acute viral hepatitis have higher rates of FHF and mortality compared to women without HEV infection<sup>[10,28]</sup>. HEV infected pregnant women also have a significantly higher risk of developing obstetric complications, and poor fetal and neonatal outcomes<sup>[29-31]</sup>.

The largest HEV outbreak lasting from September 1986 to April 1988 was reported from Xinjiang, China. 120,000 suspected and 707 death cases were recorded with an overall attack rate of 3.0%. The attack rate in pregnant women was significantly higher resulting in severe clinical outcomes with fatality rate of 5.88 % and abortion rate of 17.64%<sup>[6,11]</sup>. In recent decade, to our knowledge, such severe clinical complications have not been reported in China. However, elevation of the liver injury marker ALT has been reported in pregnant women with anti-HEV IgM positivity in Yunnan<sup>[15]</sup>, but not in Jiangsu province<sup>[14]</sup>. Genotype 4 HEV infection associated adverse maternal outcomes among pregnant women were reported in Qinhuangdao, but no death occurred<sup>[10]</sup>. In our study, we observed a slight elevation of ALT, and potential relation to pregnant hyperlipidemia in the participants with recent/ongoing HEV infection. However, the patient number of anti-HEV IgM positive was too small to draw firm conclusions.

Vertical transmission from HEV infected mothers can cause poor fetal and neonatal outcomes<sup>[32,33]</sup>. The risk of vertical transmission was 100% in an antecedent study among the pregnant women. The mothers with active diseases gave birth to babies who were either preterm or had anicteric hepatitis<sup>[34]</sup>. Similarly, in another study, among the 186 deliveries, 84% were preterm. But a significantly increased risk of preterm deliveries occurred in HEV-infected women<sup>[28]</sup>. In our study, the babies born to mothers with anti-HEV IgM antibody positive were inclined to preterm (14.3%) or jaundice (23.5%) compared to those only with anti-HEV IgG antibody positive. Again, because of limited number of anti-HEV IgM positive patents, the effects on fetal and neonatal outcomes require to be further validated.

There are some limitations in this study. We did not collect information regarding the history of women with HEV prior to enrollment or in the follow-up. These data may be helpful for better understanding the risk factors for HEV infection. Pregnant women may be co-infected with other hepatotropic pathogens, which may also contribute to the clinical outcomes, but these were not tested in this study. We could not detect HEV RNA, therefore fail to confirm the exact genotype. Finally, the patient number with anti-HEV IgM positivity was too small to draw firm conclusions.

In summary, we report the rate of HEV sero-prevalence of 6.62% among pregnant women in Inner Mongolia, China, with IgG antibody positivity of 6% and IgM antibody positivity of 0.6%.

Importantly, recent/ongoing HEV infection might be associated with a slight increase in adverse pregnancy outcomes, obstetric complications, and poor fetal/neonatal outcomes. However, the patient number recent/ongoing HEV infection was too small and caution should be taken in interpreting the clinical outcome associations. Thus, future research is warranted to further confirm our findings in large populations in China, in order to better understand and control this health issue.

## References

1. Li P, Liu J, Li Y, et al. The global epidemiology of hepatitis E virus infection: a systematic review and meta-analysis[J]. *Liver Int* 2020.
2. Zhou JH, Li XR, Lan X, et al. The genetic divergences of codon usage shed new lights on transmission of hepatitis E virus from swine to human[J]. *Infect Genet Evol* 2019; 68: 23-9.
3. Zhou X, De Man RA, De Knecht RJ, et al. Epidemiology and management of chronic hepatitis E infection in solid organ transplantation: a comprehensive literature review[J]. *Rev Med Virol* 2013;23(5):295-304.
4. Bustamante ND, Matyenyika SR, Miller LA, et al. Notes from the field: Nationwide Hepatitis E Outbreak Concentrated in Informal Settlements - Namibia, 2017-2020[J]. *MMWR Morb Mortal Wkly Rep* 2020;69(12):355-7.
5. Wang B, Akanbi OA, Harms D, et al. A new hepatitis E virus genotype 2 strain identified from an outbreak in Nigeria, 2017[J]. *Virol J* 2018;15(1):163.
6. Hakim MS, Wang W, Bramer WM, et al. The global burden of hepatitis E outbreaks: a systematic review[J]. *Liver Int* 2017;37(1):19-31.
7. Patra S, Kumar A, Trivedi SS, et al. Maternal and fetal outcomes in pregnant women with acute hepatitis E virus infection[J]. *Ann Intern Med* 2007;147(1):28-33.
8. Navaneethan U, Al Mohajer M, Shata MT. Hepatitis E and pregnancy: understanding the pathogenesis[J]. *Liver Int* 2008;28(9):1190-9.
9. Tosone G, Simeone D, Spera AM, et al. Epidemiology and pathogenesis of fulminant viral hepatitis in pregnant women[J]. *Minerva Ginecol* 2018;70(4):480-6.
10. Li M, Bu Q, Gong W, et al. Hepatitis E virus infection and its associated adverse fetomaternal outcomes among pregnant women in Qinhuangdao, China[J]. *J Maternal Fetal Neonatal Med* 2019:1-5.
11. Aye TT, Uchida T, Ma XZ, et al. Complete nucleotide sequence of a hepatitis E virus isolated from the Xinjiang epidemic (1986-1988) of China[J]. *Nucleic Acids Res* 1992;20(13):3512.
12. Wang Y, Chen G, Pan Q, et al. Chronic hepatitis E in a renal transplant recipient: the first report of genotype 4 Hepatitis E virus caused chronic infection in organ recipient[J]. *Gastroenterology* 2018;154(4):1199–201.
13. Wang M, Fu P, Yin Y, et al. Acute, recent and past HEV infection among voluntary blood donors in China: a systematic review and meta-analysis[J]. *PLoS One* 2016;11(9): e0161089.
14. Gu G, Huang H, Zhang L, et al. Hepatitis E virus seroprevalence in pregnant women in Jiangsu, China, and postpartum evolution during six years[J]. *BMC Infect Dis* 2015; 15: 560.
15. Huang F, Wang J, Yang C, et al. Chinese pregnant women in their third trimester are more susceptible to HEV infection[J]. *Braz J Infect Dis* 2015;19(6):672-4.
16. Begum N, Devi SG, Husain SA, et al. Seroprevalence of subclinical HEV infection in pregnant women from north India: a hospital based study[J]. *Indian J Med Res* 2009;130(6):709-13.
17. Stoszek SK, Abdel-Hamid M, Saleh DA, et al. High prevalence of hepatitis E antibodies in pregnant Egyptian women[J]. *Trans R Soc Trop Med Hyg* 2006;100(2):95-101.

18. Abebe M, Ali I, Ayele S, et al. Seroprevalence and risk factors of Hepatitis E Virus infection among pregnant women in Addis Ababa, Ethiopia[J]. PLoS One 2017;12(6): e0180078.
19. Naik SR, Aggarwal R, Salunke PN, et al. A large waterborne viral hepatitis E epidemic in Kanpur, India[J]. Bull World Health Organ 1992;70(5):597–604.
20. Wang Y, Liu S, Pan Q, et al. Chronic hepatitis E in an immunocompetent patient[J]. Clin Res Hepatol Gastroenterol 2019.
21. Wang Y, Liu H, Liu S, et al. Incidence, predictors and prognosis of genotype 4 hepatitis E related liver failure: a tertiary nested case-control study[J]. Liver Int 2019;39(12):2291-300.
22. Zhu G, Qu Y, Jin N, et al. Seroepidemiology and molecular characterization of hepatitis E virus in Jilin, China[J]. Infection 2008;36(2):140-6.
23. Yu Y, Sun J, Liu M, et al. Seroepidemiology and genetic characterization of hepatitis E virus in the northeast of China[J]. Infect Genet Evol 2009;9(4):554-61.
24. Wang L, Liu L, Wei Y, et al. Clinical and virological profiling of sporadic hepatitis E virus infection in China[J]. J Infect 2016;73(3):271-9.
25. Gurley ES, Hossain MJ, Paul RC, et al. Outbreak of hepatitis E in urban Bangladesh resulting in maternal and perinatal mortality[J]. Clin Infect Dis 2014;59(5):658-65.
26. Rayis DA, Jumaa AM, Gasim GI, et al. An outbreak of hepatitis E and high maternal mortality at Port Sudan, Eastern Sudan[J]. Pathog Glob Health 2013;107(2):66-8.
27. Goumba CM, Yandoko-Nakoune ER, Kommas NPA. Fatal case of acute hepatitis E among pregnant women, Central African Republic[J]. BMC Res Notes 2010; 3: 103.
28. Javed N, Ullah SH, Hussain N, et al. Hepatitis E virus seroprevalence in pregnant women in Pakistan: maternal and fetal outcomes[J]. East Mediterr Health J 2017;23(8):559-63.
29. Kumar A, Beniwal M, Kar P, et al. Hepatitis e in pregnancy[J]. Int J Gynaecol Obstet 2004;85(3):240-4.
30. Singh S, Mohanty A, Joshi YK, et al. Mother-to-child transmission of hepatitis E virus infection[J]. Indian J Pediatr 2003;70(1):37–9.
31. Aziz AB, Hamid S, Iqbal S, et al. Prevalence and severity of viral hepatitis in Pakistani pregnant women: a five year hospital based study[J]. J Pak Med Assoc 1997;47(8):198-201.
32. Hakim MS, Ikram A, Zhou J, et al. Immunity against hepatitis E virus infection: implications for therapy and vaccine development[J]. Rev Med Virol 2018;28(2).
33. El Sayed Zaki M, El Razek MM, El Razek HM. Maternal-fetal hepatitis E. transmission: is it underestimated? [J]. J Clin Transl Hepatol 2014;2(2):117-23.
34. Kumar Acharya S, Kumar Sharma P, Singh R, et al. Hepatitis E virus (HEV) infection in patients with cirrhosis is associated with rapid decompensation and death[J]. J Hepatol 2007;46(3):387-94.



## **PART III**

### ***COVID-19 spread and control measures***





# Chapter 8

## Potential association between COVID-19 mortality and healthcare resource availability

Yunpeng Ji, Zhongren Ma, Maikel P Peppelenbosch, Qiuwei Pan

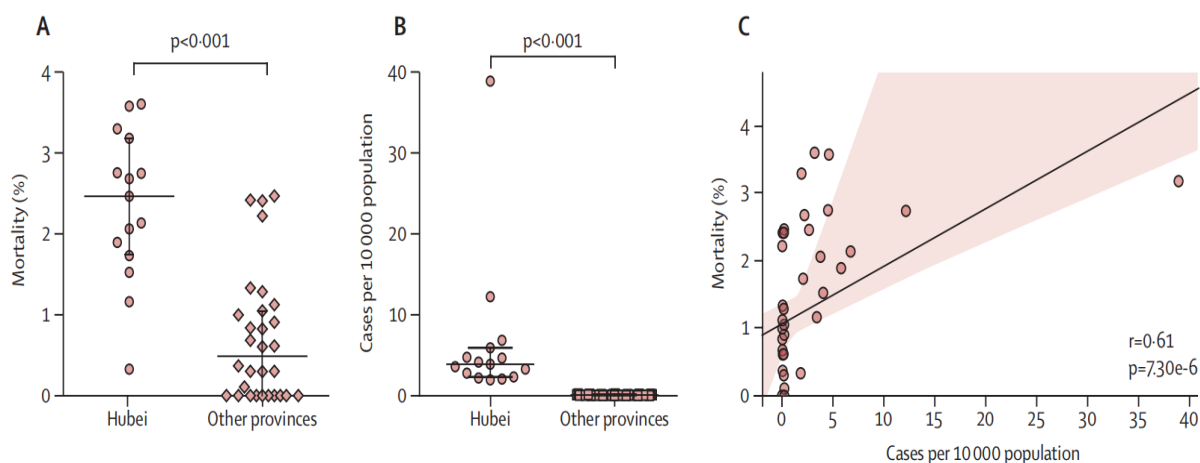
Lancet Global Health. 2020, 8, e480



The ongoing epidemic of coronavirus disease 2019 (COVID-19) is devastating, despite extensive implementation of control measures. The outbreak was sparked in Wuhan, the capital city of Hubei province in China, and quickly spread to different regions of Hubei and across all other Chinese provinces.

As recorded by the Chinese Center for Disease Control and Prevention (China CDC), by Feb 16, 2020, there had been 70641 confirmed cases and 1772 deaths due to COVID-19, with an average mortality of about 2.5%.<sup>1</sup> However, in-depth analysis of these data show clear disparities in mortality rates between Wuhan (>3%), different regions of Hubei (about 2.9% on average), and across the other provinces of China (about 0.7% on average). We postulate that this is likely to be related to the rapid escalation in the number of infections around the epicentre of the outbreak, which has resulted in an insufficiency of health-care resources, thereby negatively affecting patient outcomes in Hubei, while this has not yet been the situation for the other parts of China (figure A, B). If we assume that average levels of health care are similar throughout China, higher numbers of infections in a given population can be considered an indirect indicator of a heavier health-care burden. Plotting mortality against the incidence of COVID-19 (cumulative number of confirmed cases since the start of the outbreak, per 10 000 population) showed a significant positive correlation (figure C), suggesting that mortality is correlated with health-care burden.

In reality, there are substantial regional disparities in health-care resource availability and accessibility in China.<sup>2</sup> Such disparities might partly explain the low mortality rates-despite high numbers of cases-in the most developed southeastern coastal provinces, such as Zhejiang (0 deaths among 1171 confirmed cases) and Guangdong (four deaths among 1322 cases [0.3%]). The Chinese government has realized the logistical hurdles associated with medical supplies in the epicentre of the outbreak, and has strived to accelerate deliveries, mobilize the country's large and strong medical forces, and rapidly build new local medical facilities. These measures are essential for controlling the epidemic, protecting health workers on the front line, and mitigating the severity of patient outcomes. Acknowledging the potential association of mortality with health-care resource availability might help other regions of China, which are now beginning to struggle with this outbreak, to be better prepared. More importantly, as COVID-19 is already affecting at least 29 countries and territories worldwide, including one north African country, the situation in China could help to inform other resource-limited regions on how to prepare for possible local outbreaks.<sup>3</sup>



**Figure 1. Mortality and incidence of COVID-19 in Hubei and other provinces of China** Mortality (A) and cumulative number of confirmed cases of COVID-19 since the start of the outbreak per 10000 population (B) in Hubei and other provinces of China. Horizontal lines represent median and IQR. p values were from Mann-Whitney U test. (C) Correlation between mortality and number of cases per 10000 population (Spearman method). Data were obtained from the Chinese Center for Disease Control and Prevention to Feb 16, 2020. COVID-19=coronavirus disease 2019.

## References

1. Chinese Center for Disease Control and Prevention. Distribution of new coronavirus pneumonia. <http://2019ncov.chinacdc.cn/2019-nCoV/> (accessed Feb 16, 2020).
2. Yu M, He S, Wu D, et al. Examining the multi-scalar unevenness of high-quality healthcare resources distribution in China. *Int J Environ Res Public Health* 2019; 16: e2813.
3. Makoni M. Africa prepares for coronavirus. *Lancet* 2020; 395: 483.



# Chapter 9

## **A multi-regional, hierarchical-tier mathematical model of the spread and control of COVID-19 epidemics from epicenter to adjacent regions**

Qinyue Zheng<sup>#</sup>, Xinwei Wang<sup>#</sup>, Chunbing Bao<sup>#</sup>, Yunpeng Ji, Hua Liu, Qingchun Meng, Qiuwei Pan

(#: contributed equally)





## Abstract

Epicenters are the focus of COVID-19 research, whereas emerging regions with mainly imported cases due to population movement are often neglected. Classical compartmental models are useful, however, likely oversimplify the complexity when studying epidemics. This study aimed to develop a multi-regional, hierarchical-tier mathematical model for better understanding the complexity and heterogeneity of COVID-19 spread and control. By incorporating the epidemiological and population flow data, we have successfully constructed a multi-regional, hierarchical-tier SLIHR model. With this model, we revealed insight into how COVID-19 was spread from the epicenter Wuhan to other regions in Mainland China based on the large population flow network data. By comprehensive analysis of the effects of different control measures, we identified that Level 1 emergency response, community prevention and application of big data tools significantly correlate with the effectiveness of local epidemic containment across different provinces of China outside the epicenter. In conclusion, our multi-regional, hierarchical-tier SLIHR model revealed insight into how COVID-19 spread from the epicenter Wuhan to other regions of China, and the subsequent control of local epidemics. These findings bear important implications for many other countries and regions to better understand and respond to their local epidemics associated with the ongoing COVID-19 pandemic.

**Keywords:** COVID-19, Mathematical modeling, Epicenter, Local epidemic control, Population movement

## Introduction

The coronavirus disease 2019 (COVID-19) outbreak caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) was first sparked in Wuhan, the capital of Hubei province in China. From the epicenter, it subsequently spread to the entire country and now the globe. Since the declaration as a pandemic by WHO on March 11, 2020 (Green, 2020), it is infiltrating into every corner of the world, but local epidemics associated with this pandemic are highly dynamic. Some regions have developed as epicenters, whereas others may struggle with imported cases. There is currently intense debate and great confusion among political leaders, healthcare authorities and the general public on how to respond to the COVID-19 pandemic (The, 2020).

Since the outbreak in December, 2019, in Wuhan, the city rapidly grew into an epicenter. Because Wuhan is a major transportation hub located in the center of China and the outbreak coincided with a massive population movement due to the Chinese lunar new year holiday, it quickly spread to the entire country (Chen, Yang, Yang, Wang, & Barnighausen, 2020). Since January 23, 2020, the central government ordered heavy control measures, including city lockdown, travel restriction, and within-population quarantine. The pandemic was eventually under control by early March nationwide.

Extensive clinical, epidemiological and modeling studies have well-characterized the epidemic features of the epicenter Wuhan (Guan et al., 2020; Tian et al., 2020; Wu, Leung, & Leung, 2020). In contrast, little attention has been paid to the initial case importation and subsequent epidemic control of the different parts of Mainland China outside Wuhan. In this study, we aim to gain insight into this respect by mathematical modeling. Classical epidemic compartmental models, such as the susceptible-infected-recovered (SIR) or susceptible-exposed-infected-recovered (SEIR) model, have been widely and proven to be useful for modeling COVID-19 epidemics (Tolles & Luong, 2020). These basic models are easy to compute, but also oversimplify the complexity of disease processes, the heterogeneity of target population/society and the diversity of control measures.

In this study, we aim to develop a mathematical model to recapitulate the SARS-CoV-2 transmission patterns from epicenter to other regions during the early stages of the outbreak. To better recapitulate the real-world complexity and heterogeneity, we constructed a modified multi-regional, hierarchical-tier susceptible-latent-infected/hospitalized-recovered (SLIHR) model by incorporating the population flow network data. With this model, we estimated the overall and individual epidemics in different provinces of Mainland China outside Wuhan. We further performed comprehensive data/information mining to

understand the effectiveness of specific control measures in contributing to local epidemic control at provincial level.

## **Methods**

### **Model assumption**

A two-stage, hierarchical-tier, multi-regional SEIR model for COVID-19 epidemic was developed based on the population flow network. The first tier of our model was an open epidemic transmission system that described the risk of case importation. Before January 24, COVID-19 spread freely in Wuhan, and the isolation for infected individuals was incomplete under insufficient medical resources. The actively infected and latently infected individuals who were neither hospitalized nor isolated had been imported to other provinces in Mainland China as the population flow and then spread locally. In addition, our model specified infected individuals who could not be hospitalized for isolation due to medical resource constraints in Wuhan at the early stage. This population largely contributed to case importation from Wuhan to other provinces of Mainland China. Simultaneously, the model considered the heterogeneity of infection risk between close contacts and general contacts. The second tier of our model was multiple independent closed epidemiological transmission systems describing the local spread of COVID-19. From January 23 to January 25, the Level 1 emergency response was activated in all provinces of Mainland China except Xizang Province. The COVID-19 epidemic was confined to multiple separate closed systems and localized transmission occurred within each province under strict interprovincial traffic restrictions.

We simulated the possible epidemic scenarios assuming that there were no control measures with localized transmission in provinces under the same basic reproduction number. The gaps between the number of probable infections without control measures and real-world confirmed cases in each province represent the effectiveness of provincial responses on controlling the COVID-19 epidemic.

### **Data sources**

The Chinese government implemented strict control measures. Large-scale screening was conducted to identify infected cases and contacts with confirmed cases were closely tracked, resulting in rapid containment of the epidemic in China (Xing, Wong, Ni, Hu, & Xing, 2020). Thus, data of confirmed cases reported by China National Health Commission highly represents the real-world epidemic (Supplementary Table S1). Population flow network was based on "Baidu Migration" Big Data Platform, the largest database that reflects the size of the regional population inflow or outflow according to the geographic change of the users'

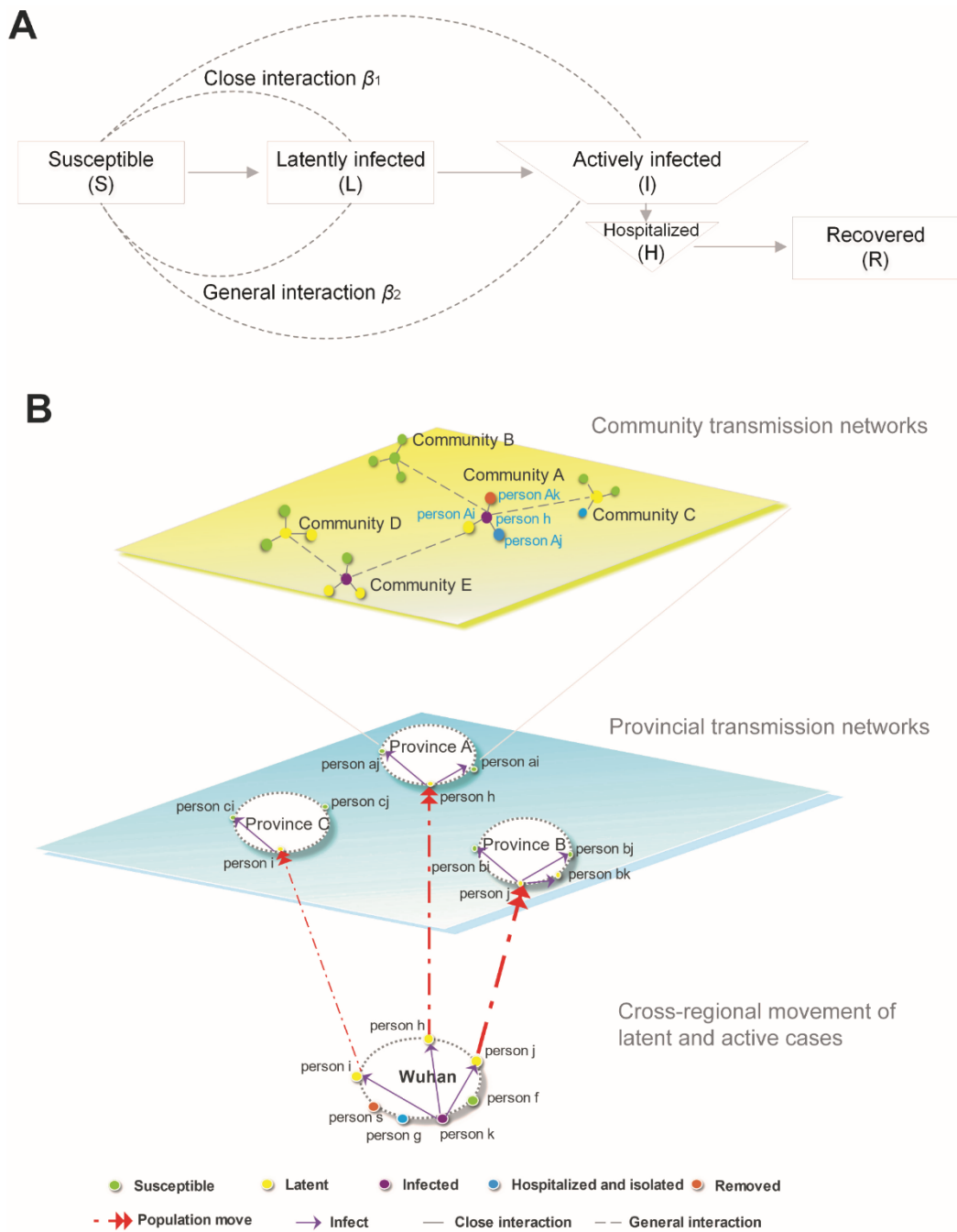
mobile devices in China (Baidu Online Network Technology (Beijing) Co. L. Baidu Map Smart). Data/Information on control measures of different provinces in Mainland China were comprehensively collected from local government documents, announcements and press conferences (Supplementary Table S2).

## Mathematical model

A modified hierarchical-tier, multi-regional SLIHR (susceptible - latently infected - un-isolated actively infected - hospitalized and isolated - dead) model with interregional and interpersonal network was built (Figure 1). For interregional transmission network, we considered that China has developed advanced transportation networks, and the early outbreak was coincided with a massive population migration, because of the Chinese lunar new year holiday. In the free spreading stage, Wuhan and its neighbors form a star network structure with traffic connections, and the latently (without symptom) or actively (with symptoms) infected individuals were exported from Wuhan in one direction. The initial COVID-19 cases of other regions in China were mainly imported from the epicenter. Thus, at the controlling stage, the initial values of multi-regional model were based on Wuhan population outflow network.

Considering the real-world situation of COVID-19 epidemics in China, we first developed an open SLIHR model (Figure 1A) for the epicenter Wuhan considering incomplete isolation and heterogeneous interactions. Transmission of SARS-CoV-2 occurs in the open system, and the total population of Wuhan varies with the cross-regional inflow and outflow of the population. The susceptible, the latently infected and infected but un-isolated individuals had been imported to other provinces of Mainland China. The probability of each subpopulation flowing out or staying in Wuhan was assumed equal. The infected individuals who were isolated, admitted to hospital or recovered could not flow out of Wuhan. In the controlling stage, SARS-CoV-2 was confined to several independent and closed systems under traffic restrictions after Level 1 emergency response was implemented. First generation cases in other regions outside Wuhan depended on population outflow from Wuhan and the number of latent or active patients without isolation in Wuhan before travel ban.

For better understanding the epidemic spread, interpersonal transmission network was also established (Figure 1B). The real-world interpersonal contact network is not homogeneous but close to community structure where individuals are grouped by different interaction frequency, which directly result in heterogeneous transmission in our model. The nodes (individuals) in the interaction network are grouped. The nodes (individuals) in the group (members of the same family) have high frequency of contact, while the nodes (individuals) between the groups (daily contact acquaintances) have low frequency of contact.



**Fig. 1 A multi-regional, hierarchical-tier SLIHR model for studying COVID-19 epidemics in China.** (A). An open SLIHR model considering incomplete isolation and heterogeneous interactions. (B). The structure of COVID-19 epidemic model with hierarchical transmission tiers.

In order to precisely capture the epidemic spread from epicenter to adjacent regions in Mainland China, we adopted parameters from widely cited studies on early epidemics in Wuhan and other regions in Mainland China (Chinazzi et al., 2020; Li et al., 2020). The details of our model were described in Supplementary methods, and the parameters in our model were referred to widely-cited literatures (Supplementary Table S3). We first estimated the epidemics in Wuhan and the exportation to other adjacent regions before travel ban. We then simulated the possible epidemic scenarios in other provinces except Wuhan under the assumption of no control measures. The heterogeneous effects of control measures in each province could be analyzed by comparing the number of probably infected individuals and the real-world confirmed cases.

### **Classification of control measures**

Based on collected data/Information on control measures of different provinces (Supplementary Table S2), we characterized their epidemic management systems in six dimensions and classified control measures into three levels based on the timeliness of implementation. The dimension of Level 1 emergency response reflects the speed in responding to the outbreak. Traffic restrictions represent the government's ability to control interprovincial traffic and urban public transport in a timely manner. The dimension of mask wearing claim portrays the intensity and timeliness of personal protection in public places requested by local governments. Community prevention means that governments conduct population screening in the communities. Big data tools are used by governments for contact tracing, accurate forecasting and control. The dimension of work resumption is expressed as the rate of return-to-work. The timeline for each control measure was determined by the median time of all provinces, and each province is assigned a rating based on the chronological order in which that measure was implemented.

### **Scope and timeframe**

The effectiveness and heterogeneity of epidemic containment across different parts of China mainly depend on the levels of case importation from epicenter and the subsequent control measures. Our network based mathematical model is capable of recapitulating and integrating these multiple factors. Because control measures were ordered by the central government on January 23 and a temporary diagnostic method that is different from the regular RT-PCR case diagnosis was adopted on February 12 in Hubei province, therefore we used the epidemiological data from January 24 to February 11 to assess the epidemic control.

### **Data visualization and Statistic analysis**

The mathematical models were solved by simulation methodology with the help of MATLAB 2016b (The MathWorks, Inc., Massachusetts). Epidemic maps were drawn using ArcGIS 10.1 (Environmental Systems Research Institute, Inc., California). Kendall's tau-b correlation coefficient was chosen to test the non-parametric correlation between ordered categorical variables and a continuous variable, where multiple mutual linear problems of independent variables were excluded by collinearity diagnosis. We used the Kendall's tau-b correlation coefficient method to test the non-parametric correlation between the varied control measures and the number of reduced probable infections. Statistical analysis was performed using IBM SPSS Statistics version 24 (International Business Machines Corporation, New York).  $P < 0.05$  was considered as statistically significant, and all tests were two-tailed.

## Results

### Simulating the epidemics in Wuhan and other regions of Mainland China

We collected and categorized the number of daily confirmed new cases in Wuhan and other regions of Mainland China, respectively (Figure 2A). As expected, the first case in Wuhan appeared much earlier, but the epidemic curve was delayed as compared to that of Mainland China outside Wuhan. This indicates delayed case diagnosis during the early epidemic in Wuhan. The overall epidemic period was much shorter in other regions of Mainland China compared to that in Wuhan. Of note, the unusual escalation of case number on February 12 reported by China National Healthcare Commission was attributed to the temporary inclusion of the “clinically diagnosed” cases without RT-PCR confirmation (Sun et al., 2020). This was intended to swiftly isolate and treat the large number of suspected cases in the epicenter, while the capacity of RT-PCR testing was limited at that time.

We first simulated the epidemic in Wuhan by an open SEIHR model considering incomplete isolation and heterogeneous interaction. The infected but un-isolated individuals were then estimated based on population flow network (Figure 1). The probable infections in each province except Wuhan were simulated assuming no control measures implemented. Our simulation revealed that the number of latently infected individuals without symptom was 1,600, and the total number of infected individuals with symptom was 1,581 in Wuhan as of January 24. We estimated that 369 latent or active patients already left Wuhan before implementation of the travel ban, thus spreading to other regions outside Wuhan.

Our model simulation fits with the reported cases during the early stages of the epidemics, for example by February 13 in Wuhan (Figure 2B) and by February 1 outside of Wuhan (Figure 2C). More specifically, the simulated number of isolated (hospitalized) active cases in Wuhan was in agreement with the reported confirmed cases by January 24. However, by that early

stage, screening capability and hospital beds were insufficient, there were infected but unidentified cases difficult to be estimated. From January 25 to February 14, the total number of simulated actively infected individuals was consistent with the reported data (Figure 2B). Similarly, in other regions outside Wuhan, the simulated infections assuming no control measures were in line with the reported data in January. But since February, there was an increasing gap between estimated and reported numbers, which exactly indicates the effects of implemented control measures in limiting epidemic spread (Figure 2C).

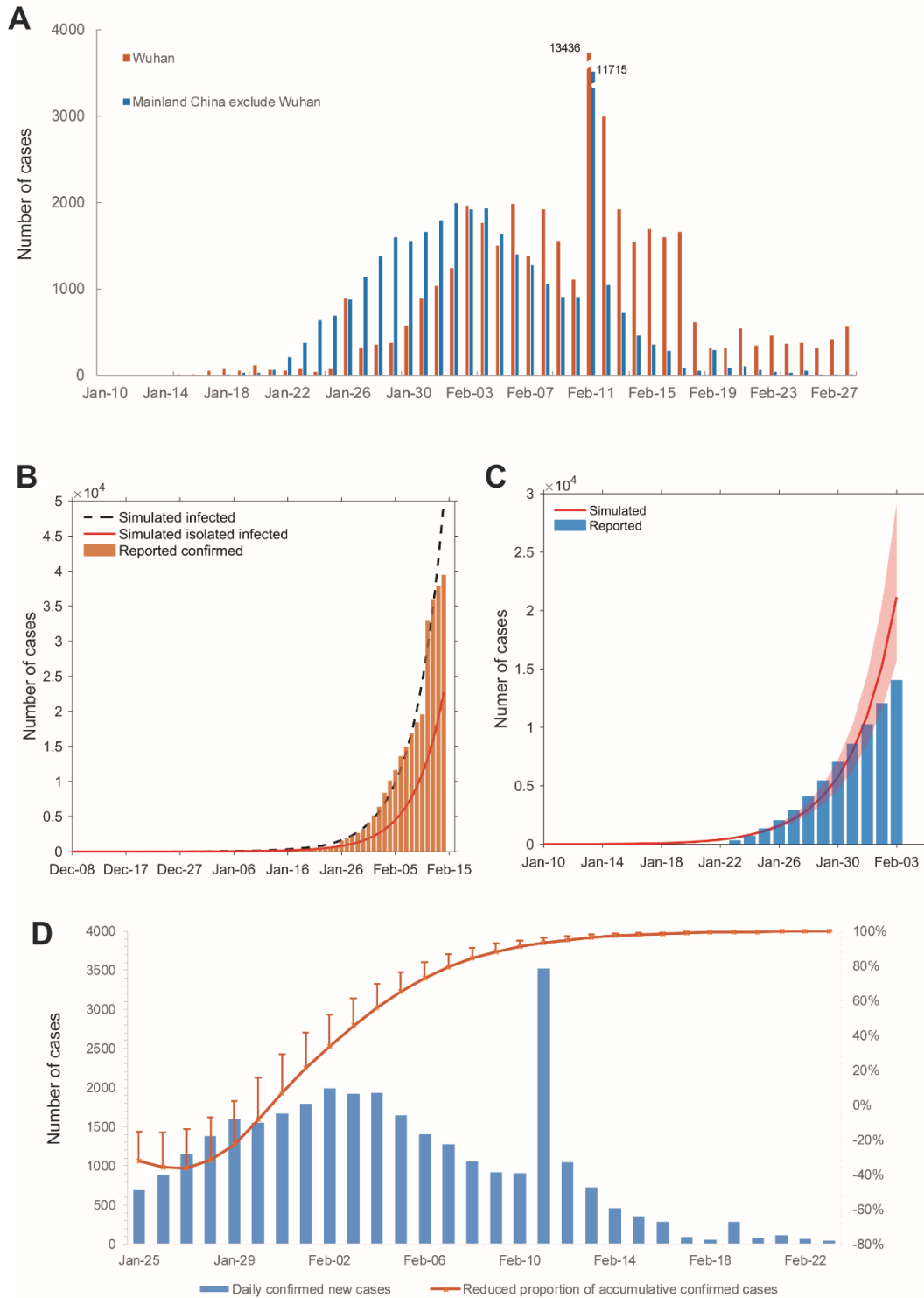
### **Simulating the overall effects on epidemic control in Mainland China outside of Wuhan**

From January 23, the Chinese central government began to implement heavy control measures, including city lockdown, travel ban, and within-population quarantine. Here, the effects of control measures were based on the reduced number of probable infections which was calculated by subtracting confirmed cases from the simulated probable infections assuming no control measures in each province.

We estimated that the probable number of actively infected individuals would exceed 17.4 (90% CI 8.1-39.1) million and the number of latent individuals would be 29.8 (90% CI 11.7-78.9) million in Mainland China outside Wuhan one month after January 24, if no control measures were implemented. In real-world, control measures were universally implemented albeit at various levels across different provinces of China. Thus, the number of reported daily new cases peaked around 10 days after implementing control measures but declined steadily thereafter. The rebound on February 12 was attributed to the temporary adoption of the new clinical diagnose method (Figure 2D).

In general, the simulated results before the epidemic peak were highly in agreement with the reported confirmed cases, both for Wuhan and outside Wuhan in Mainland China. In contrast, the reported confirmed cases after the peak were far less than the simulated numbers, which reflected the effects of control measures in mitigating SARS-CoV-2 spread. Two weeks of implementing control measures was estimated to reduce the probable number of infections by 56,535 (90% CI 31681-96743) in Mainland China outside Wuhan. This constitutes a 73% (90% CI 60.2-82.2%) reduction. One month after implementation, the number of infections was reduced by 99.8% (90% CI 99.6-99.9%), thus preventing 17.4 (90% CI 8.1-39.1) million people from infection.





**Fig. 2 Model simulation to fit the epidemics in China and estimating effectiveness of epidemic control.** (A). Daily confirmed new cases of the epicenter Wuhan and Mainland China excluding Wuhan. (B). Simulated probable numbers of currently infected individuals and currently isolated infected individuals in Wuhan. (C). Simulated probable infections assuming no control measures of Mainland China outside Wuhan. (D). Effectiveness of epidemic control in Mainland China outside Wuhan.

## **Heterogeneity in epidemic control across different provinces in Mainland China**

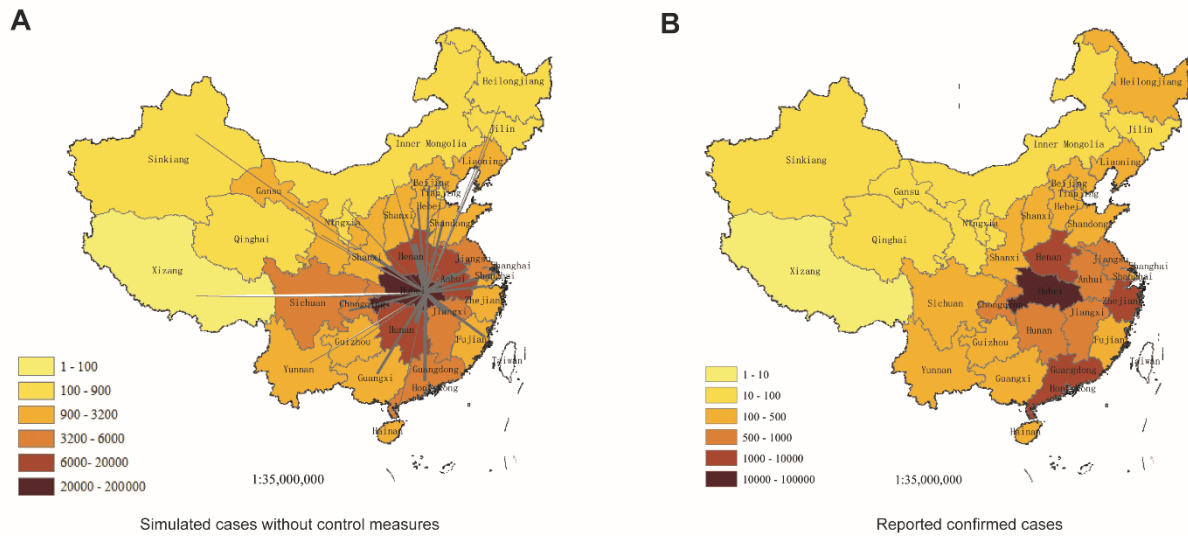
After January 24, the movement of people across provinces was restricted due to travel ban. Thus, the local transmission of COVID-19 in each province is considered as an individual unit, as described by the second tier of our model framework (Figure 1). COVID-19 spread in each province was independent of each other, and the control measures presented heterogeneity. In this case, different effectiveness of provincial responses could be estimated by comparing the real-world confirmed cases data and the number of probable infections assuming no control measure.

We generated a country map with cross-sectional comparison that visualizes the distinct effectiveness in containing COVID-19 epidemics across different provinces in Mainland China (Figure 3). The different gaps between reported cases and simulated numbers assuming without control measures were also shown for each province. The larger the gap indicates the more effective in epidemic control in that province (Figure 4). To further compare the differences at different stages, we mapped the real-world COVID-19 spread among different provinces during one month after implementing control measures (Supplementary Figure S1).

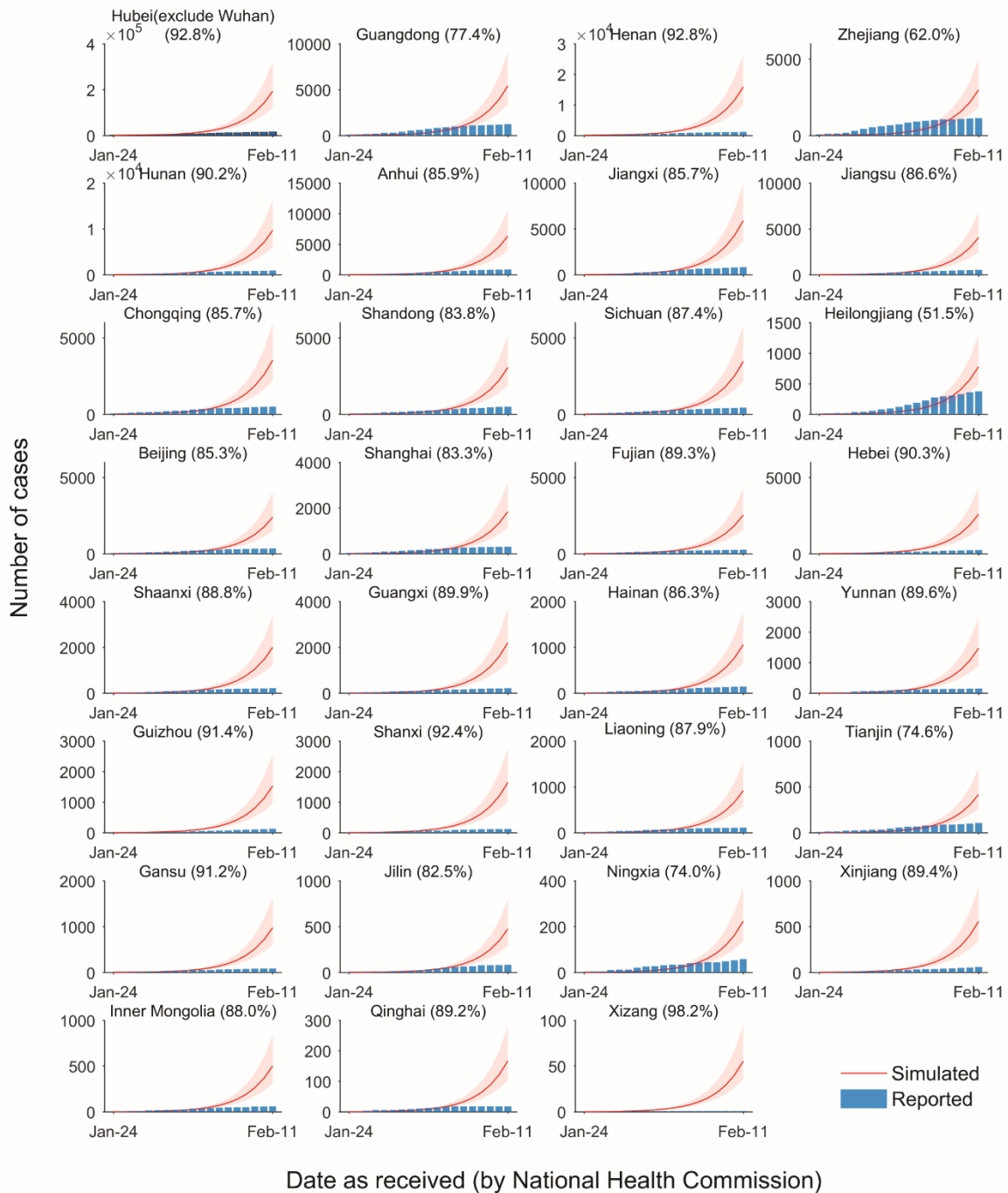
### **The effectiveness of specific control measures on local epidemic containment at provincial level**

To understand the effectiveness underlying the heterogeneity in epidemic control, we characterized their epidemic management systems in six dimensions and classified control measures into three levels based on the timeliness of implementation (Figure 5).

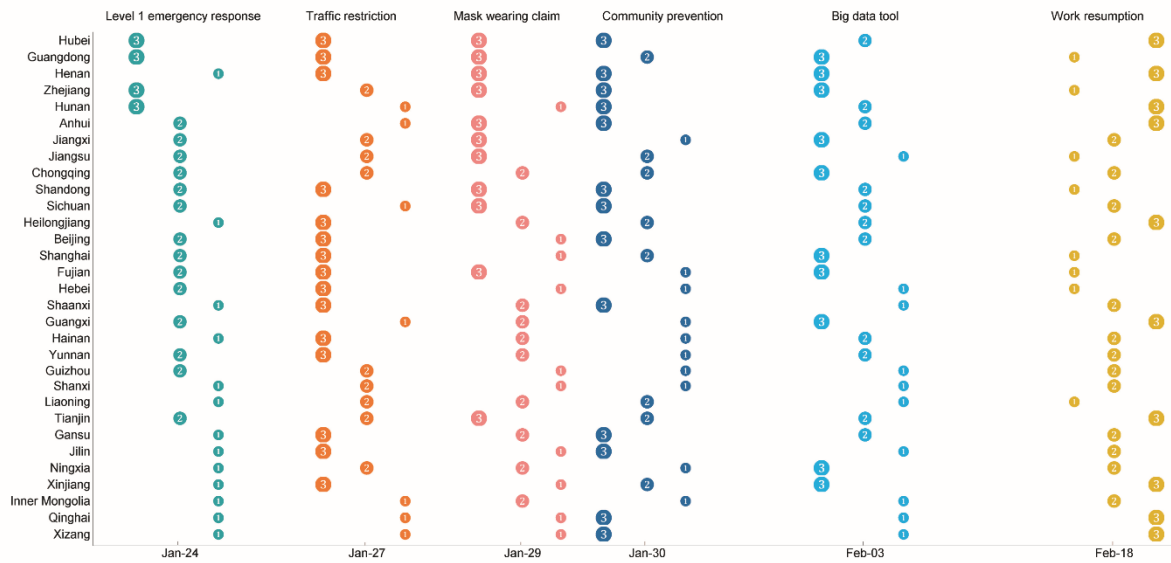
By correlation analysis, we found that Level 1 emergency response, community prevention and application of big data tools significantly correlated with the effectiveness of local epidemic containment (Table 1). Traffic restriction and mask wearing were universally and strictly implemented across the country, and therefore did not show statistically significant correlation with the effectiveness of local epidemic containment (Table 1). We did not include the item work resumption in this analysis, as this occurred mainly at the end or post-epidemic, which could be relevant for possible second wave but not the primary epidemic.



**Fig. 3 Visualizing simulations without control measures and the real-world confirmed situation from a provincial perspective.** (A). The number of simulated cases in different provinces of Mainland China excluding Wuhan without control measures was graded as of February 11, 2020. Considering that the first-generation cases were imported from the epicenter, we provided a secondary view of the population flow size from Wuhan to the different provinces of Mainland China from the start of the Spring Festival on January 10 to the Wuhan travel ban on January 24. (B). We provided a provincial view of the confirmed cases of different provinces reported by China National Health Commission as of February 11.



**Fig. 4 The effect of province-level responses in Mainland China excluding Wuhan.** Simulated and confirmed case numbers were plotted for each province, and the effectiveness of local epidemic control was calculated as percentage. Because the epicenter Wuhan is the capital for Hubei province. Thus, the case number of Wuhan was excluded for calculating Hubei province. Data on confirmed cases were reported by China National Health Commission.



**Fig. 5 The responses of different provinces in Mainland China to COVID-19 epidemic.** The local epidemic management systems of different provinces were characterized into six dimensions, and the dates on the horizontal axis was determined by the median time of each control measure in all provinces. Each province was assigned a rating based on the chronological order in which that measure was implemented, and then classified equally into three levels based on the timeliness of implementation. For the first five control measures, the earlier they were implemented, the better. But for the last dimension, the level scale was reversed: the later people return to work, the better. The work resumption dimension is expressed in terms of the return-to-work rate. At the same time point, the lower the return-to-work rate, the stronger intensity of work resumption dimension. Data/information main source: local government documents, announcements and press conferences.

**Table 1: Correlation analysis between epidemic responses and reduced COVID-19 cases in different provinces of Mainland China.** \* P<0.05; \*\* P<0.01; tested by the Kendall's tau-b correlation coefficient method.

Control measures	Correlation coefficient	P-value
Level 1 emergency response	0.566**	P<0.01
Traffic restriction	0.072	0.62
Mask wearing claim	0.126	0.38
Community prevention	0.384**	P<0.01
Big data tool	0.285*	P<0.05

## Discussion

Effective mitigation of COVID-19 pandemic requires deep understanding of transmission dynamics and control measures for both epicenters and emerging regions primarily with imported cases. However, research tends to focus on epicenters but disregard other regions. In fact, emerging regions with imported cases are relatively easy to be contained, but neglectance bears high risk of growing into new epicenters.

Classical compartmental mathematical models are excellent for studying epidemics in relatively homogeneous settings (Chinazzi et al., 2020; Giordano et al., 2020; Salje et al., 2020). The COVID-19 epidemics in other regions of China outside Wuhan, however, involve initial case importation and subsequent local transmission and control measure implementation (Ji, Ma, Peppelenbosch, & Pan, 2020). An innovative aspect of this study is that we constructed a two-tier SLIHR model to accommodate this complexity and heterogeneity. The first tier considered an open SLIHR model of latent time lag, incomplete isolation, heterogeneous contacts, and exported cases from Wuhan to other regions in Mainland China. The second tier captured the spread of COVID-19 in multiple closed and unassociated regions based on the pre-imported cases. In our model, we adopted epidemiological data of case numbers reported by China National Health Commission. There was debate on the accuracy of reporting case number at early stage of the outbreak in Wuhan (Wu et al., 2020), but we mainly focus on the epidemics outside the epicenter. Their reported data are highly accurate, because identification of imported cases and contact tracing were rigorously implemented in all provinces of Mainland China. Furthermore, we did not consider the death cases, because data reporting on death was scarce and inconsistent at early stage of the epidemic. More importantly, we focus on regions outside Wuhan where death rates were extremely low in general but with huge variations across different provinces (Ji et al., 2020).

Exporting cases from Wuhan to other regions of China is a heterogamous process mainly determined by population migration. We have incorporated the large population flow network data in our model. These data were retrieved from the "Baidu Migration" big data platform which is based on the user's mobile device geographic location changes and reflects the size of population regional inflow or outflow. There are many factors affecting the level and pattern of population flow, including geographic locations, transportation connections, socioeconomic status, population characteristics, social values and cultural norms. In Mainland China, there are 22 provinces, five autonomous regions and four municipalities, which are all at provincial level but have their own distinct features. For example, the Tibet Autonomous Region (Xizang) only had one imported case and it was immediately contained. This is clearly attributed to the geographic, population and cultural distinctions of the region that has minimal population movement between the epicenter. In contrast, Zhejiang province

which is several hundred kilometers away from Wuhan had one of the highest numbers of imported cases. This was mainly related to advanced economic development that many people from this province are running businesses in Wuhan, and returned back for Chinese New Year festival. The high level of case importation has compromised their effectiveness of epidemic control, although Zhejiang province has implemented heavy measures with ample resources available (Qian et al., 2020).

Although the general policy of control measures was ordered by the central government, it is operated and coordinated at provincial level locally. This is the second level attributing to the heterogeneity of effectiveness in epidemic control among different provinces. For example, Heilongjiang, the northernmost province of the Northeast region, had limited number of imported cases, but their epidemic grew substantially. As we estimated, Heilongjiang has the lowest effectiveness in controlling COVID-19 epidemic.

In this study, we in-depth analyzed the association of six dimensions of control measures with the effectiveness of epidemic control in different provinces. We did not observe significant correlation between traffic restrictions or mask wearing claim with effectiveness of local epidemic control. Travel restriction is essential for limiting COVID-19 spread. But the central government has already imposed strict lockdown across the country, which may explain that this policy dimension from local government may not have additional effect. Whether wearing face masks for the general population can protect against SARS-CoV-2 has long been debated, especially in the western world (Lazzarino, Steptoe, Hamer, & Michie, 2020). It now becomes clear that people wear masks not only protect themselves, but also protect others by limiting spread of respiratory droplets. Many countries, including from western world, have required or advised their citizens to wear masks in public places (Cheng, Lam, & Leung, 2020). In contrast, the Chinese population is highly aware of the protective values of wearing face masks (Wang et al., 2020), and they spontaneously and universally adopted this measure even before the request from government. This may explain why the dimension of mask wearing claim from local government did not have additional impact in our study.

It is not surprising that Level 1 emergency response is significantly associated with epidemic control effectiveness. We also found community prevention and application of big data tools are significant factors. Both require resources, expertise and advanced economic status. The economic status and growth in China have geographic imbalances, and our results indicate that this appears to have an effect on local epidemic response. We call the authorizes to pay attention to this regional inequality and to ensure equal access to resources, advanced tools and technologies for enhancing outbreak preparedness across the country.

In summary, we developed a multi-regional, hierarchical-tier SLIHR model that is capable of recapitulating the complexity and heterogeneity of COVID-19 epidemics in China. We revealed insight into how COVID-19 was spread from the epicenter to other regions of Mainland China, and characterized the key control measures that contributed to the effectiveness of local epidemic containment. These findings bear important implications for many countries or regions to understand and better respond to their local epidemics associated with this COVID-19 pandemic.

## References

- Chen, S., Yang, J., Yang, W., Wang, C., & Barnighausen, T. (2020). COVID-19 control in China during mass population movements at New Year. *Lancet*, 395(10226), 764-766. doi:S0140-6736(20)30421-9 [pii]10.1016/S0140-6736(20)30421-9
- Cheng, K. K., Lam, T. H., & Leung, C. C. (2020). Wearing face masks in the community during the COVID-19 pandemic: altruism and solidarity. *Lancet*. doi: S0140-6736(20)30918-1 [pii]10.1016/S0140-6736(20)30918-1
- Chinazzi, M., Davis, J. T., Ajelli, M., Gioannini, C., Litvinova, M., Merler, S., . . . Vespignani, A. (2020). The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science*, 368(6489), 395-400. doi: science.aba9757 [pii]10.1126/science.aba9757
- Giordano, G., Blanchini, F., Bruno, R., Colaneri, P., Di Filippo, A., Di Matteo, A., & Colaneri, M. (2020). Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. *Nat Med*. doi:10.1038/s41591-020-0883-7.1038/s41591-020-0883-7 [pii]
- Green, M. S. (2020). Did the hesitancy in declaring COVID-19 a pandemic reflect a need to redefine the term? *Lancet*, 395(10229), 1034-1035. doi: S0140-6736(20)30630-9 [pii]10.1016/S0140-6736(20)30630-9
- Guan, W. J., Ni, Z. Y., Hu, Y., Liang, W. H., Ou, C. Q., He, J. X., . . . China Medical Treatment Expert Group for, C. (2020). Clinical Characteristics of Coronavirus Disease 2019 in China. *N Engl J Med*, 382(18), 1708-1720. doi:10.1056/NEJMoa2002032
- Ji, Y., Ma, Z., Peppelenbosch, M. P., & Pan, Q. (2020). Potential association between COVID-19 mortality and health-care resource availability. *Lancet Glob Health*, 8(4), e480. doi: S2214-109X(20)30068-1 [pii]10.1016/S2214-109X(20)30068-1
- Lazzarino, A. I., Steptoe, A., Hamer, M., & Michie, S. (2020). Covid-19: Important potential side effects of wearing face masks that we should bear in mind. *BMJ*, 369, m2003. doi:10.1136/bmj.m2003
- Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., . . . Feng, Z. (2020). Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. *N Engl J Med*, 382(13), 1199-1207. doi:10.1056/NEJMoa2001316
- Qian, G. Q., Yang, N. B., Ding, F., Ma, A. H. Y., Wang, Z. Y., Shen, Y. F., . . . Chen, X. M. (2020). Epidemiologic and Clinical Characteristics of 91 Hospitalized Patients with COVID-19 in Zhejiang, China: A retrospective, multi-centre case series. *QJM*. doi:5809152 [pii]10.1093/qjmed/hcaa089
- Salje, H., Tran Kiem, C., Lefrancq, N., Courtejoie, N., Bosetti, P., Paireau, J., . . . Cauchemez, S. (2020). Estimating the burden of SARS-CoV-2 in France. *Science*. doi: science.abc3517 [pii]10.1126/science.abc3517



- Sun, H., Ning, R., Tao, Y., Yu, C., Deng, X., Zhao, C., . . . Tang, F. (2020). Comparison of clinical and microbiological diagnoses for older adults with COVID-19 in Wuhan: a retrospective study. *Aging Clin Exp Res*, 32(9), 1889-1895. doi:10.1007/s40520-020-01647-4 10.1007/s40520-020-01647-4 [pii]
- The, L. (2020). COVID-19: too little, too late? *Lancet*, 395(10226), 755. doi: S0140-6736(20)30522-5 [pii]10.1016/S0140-6736(20)30522-5
- Tian, H., Liu, Y., Li, Y., Wu, C. H., Chen, B., Kraemer, M. U. G., . . . Dye, C. (2020). An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China. *Science*, 368(6491), 638-642. doi: science. abb6105 [pii]10.1126/science.abb6105
- Tolles, J., & Luong, T. (2020). Modeling Epidemics with Compartmental Models. *JAMA*. doi:2766672 [pii]10.1001/jama.2020.8420
- Wang, Y., Tian, H., Zhang, L., Zhang, M., Guo, D., Wu, W., . . . MacIntyre, C. R. (2020). Reduction of secondary transmission of SARS-CoV-2 in households by face mask use, disinfection and social distancing: a cohort study in Beijing, China. *BMJ Glob Health*, 5(5). doi: bmjgh-2020-002794 [pii]10.1136/bmjgh-2020-002794
- Wu, J. T., Leung, K., & Leung, G. M. (2020). Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *Lancet*, 395(10225), 689-697. doi: S0140-6736(20)30260-9 [pii]10.1016/S0140-6736(20)30260-9
- Xing, Y., Wong, G. W. K., Ni, W., Hu, X., & Xing, Q. (2020). Rapid Response to an Outbreak in Qingdao, China. *N Engl J Med*, 383(23), e129. doi:10.1056/NEJMc2032361

## Supplementary Materials

### Material and Methods

#### The structure of multi-regional, hierarchical-tier model

The multi-regional, hierarchical-tier model has been characterized by comparing with the classical SEIR epidemiological models as follows:

a) Open system: Wuhan, where COVID-19 was initially known in Mainland China, is one of the largest transportation hubs in China. It cannot be ignored the large-scale population outflow from Wuhan before the Wuhan travel ban, which changed the susceptible base and the number of actively and latently infected individuals in the early epidemic. At this time, the model could not be assumed as a closed endogenous model with the constant population.

b) Population flow network: There are advanced transportation networks in Mainland China, so the scale of population movements is enormous. In addition, the traffic peak during Chinese New Year has also become the accelerator of virus transmission. In the free spreading stage before control, Wuhan and its neighbors form a star network structure with traffic connections, where the actively infected individuals are imported, and the active or latent patients outflowed from Wuhan in one direction<sup>(1)</sup>. The initial values of the epidemiological transmission models applied to each region outside Wuhan vary according to the closeness to Wuhan, which cannot be treated equally.

c) Incomplete isolation: Epidemic control and patient treatment require substantial medical resources. As the outbreak was unexpected, the supply of medical resources in early outbreak of Wuhan could not meet the demand, in which caused the active patients could not be hospitalized in time for isolation and became the virus disseminators. Li et al. found that at the beginning of the epidemic, 27% of infected patients were seen within 2 days of the illness onset, but 89% of infected patients were not admitted until at least 5 days of the illness onset. However, the classic SEIR model does not differentiate between individuals who are hospitalized and those who are not.

d) Heterogeneous interaction: Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) was transmitted through human contact. The disease transmission rate depends on the number of contact links and the probability of infection<sup>(2)</sup>. However, the real-world interpersonal contact network is not homogeneous, rather than close to community structure. The nodes (individuals) in the network are grouped, the nodes in the group are closely connected (high-frequency contact), and the nodes between the groups are sparsely connected (low-frequency contact)<sup>(3)</sup>. Studies have shown that in the disease transmission network, the probability of infection of close neighbors (members of the same family) is much higher than that of general neighbors (daily contact acquaintances)<sup>(2)</sup>. Therefore, it is necessary to distinguish the risks of individuals exposure to the disease in the susceptible

compartment of the model, while the classical SEIR model does not specifically treat heterogeneous contacts.

e) Time-varying cure rate: The cure rate of infected individuals is increasing with improving pathological understanding and clinical experience. The cure rate is an important underlying parameter in the stability analysis of the differential equation of epidemiological transmission dynamics, which affects the epidemic trend. But the classical SEIR model is an autonomous system, which means the system does not contain any time-varying parameters.

Here, the dead compartment was not set in our model. SARS-CoV-2 is more communicable but less pathogenic compared with severe acute respiratory syndrome coronavirus (SARS-CoV). There were significant regional differences and time differences in mortality rates. Moreover, it was not possible to know the mortality of unrecognized infections in the early outbreak, and the mortality rate would continue to change until all the infected individuals recover or dead. Virus with lower health threats at the individual level could cause high risk of interpersonal transmission<sup>(4)</sup>. This study focused on the description of COVID-19 transmission chain and how to cut off the transmission. Therefore, the death compartment was not considered in the model.

Overall, this improved model is expected to portray the true pathways of transmission of COVID-19 pandemic. The COVID-19 transmission structure in Mainland China could be decomposed into two subsections. The open SLIHR model estimated Wuhan epidemic situation and outflow of active or latent patients before the Wuhan travel ban, considering the incomplete isolation and heterogeneous contacts. The multi-regional description of viral spread dynamics was based on population flow network, which was used to estimate the COVID-19 spread in each province in Mainland China.

### **An open SLIHR model considering incomplete isolation and heterogeneous interaction**

In order to provide a more accurate dynamic description of COVID-19 epicenter Wuhan in the early free spread stage before the travel ban, the open SLIHR model with heterogeneous interaction were set as follows:

SARS-CoV-2 transmitted in an open system, and the total population of Wuhan varies with the cross-regional inflow and outflow of the population. The distribution probability of the susceptible, the latently infected and un-isolated infected individuals with symptoms were assumed equal probability distribution in the outflow population and the resident population. And isolated infected admitted to hospital individuals and the recovered ones could not flow out of Wuhan.

The disease transmission rate  $\beta$  of un-isolated infected individuals depends on the infection rate  $\lambda$  and contact links  $K$ , and it is generally considered that  $\beta = \lambda K$  <sup>(2)</sup>. Heterogeneous mix of susceptible individuals exposed to the infected, with different transmission rates. In other

words, those who are more frequently contact with the actively infected are more likely to be infected<sup>(5)</sup>. Our model assumed that susceptible individuals in contact with infected individuals were divided into two subpopulations according to the frequency of interaction. One was the susceptible subpopulation of high-frequency interaction, that was from the same household. And the other was the susceptible subpopulation of low-frequency interaction, that was general acquaintances (colleague relationships, classmate relationships, gathering events).

Latently infected individuals have the ability to infect susceptible individuals<sup>(4,5)</sup>, but there is no clear medical evidence yet to validate the gap between the ability of latently and actively infected individuals to infect. We hypothesized that the latently infected individuals had a slightly weaker infectious capacity than the actively infected ones. The latently infected individuals share the same subpopulation of high-frequency links as the actively infected ones. However, because the onset of illness limited mobility of the actively infected individuals, there were differences in the subpopulation of low-frequency links between the actively infected and latently infected individuals.

Constrained by medical resources at the early stage of the COVID-19 outbreak in Wuhan, there was a time lag between the illness onset and admission to isolation or treatment<sup>(6)</sup>. The actively infected individuals were admitted to isolation in hospital with a certain probability<sup>(7)</sup>, assuming that the actively infected individuals who were admitted to isolation in hospital without infectious conditions, while the actively infected individuals who were non-hospitalized with infectious capacity and conditions. The number of reported confirmed cases that could be traced was assumed to be equal to the part of actively infected individuals who had been confirmed and admitted to hospital, Admittedly, this situation had been considerably alleviated with the replenishment of medical resources in the control stage of the COVID-19 epidemic in Wuhan.

By optimizing the openness of the system, heterogeneous interaction, and isolation constrained. The improved model could be described by the following equations:

$$\begin{aligned}
 S(t+1) &= S(t) - \beta_1 I(t) \frac{S(t)}{N(t)} - \beta_2 L(t) \frac{S(t)}{N(t)} - s(t) \\
 L(t+1) &= L(t) + \beta_1 I(t) \frac{S(t)}{N(t)} + \beta_2 L(t) \frac{S(t)}{N(t)} - \alpha L(t) - e(t) \\
 I(t+1) &= I(t) + \alpha L(t) - \eta I(t) - i(t) \\
 H(t+1) &= H(t) + \eta I(t) - \gamma H(t) \\
 R(t+1) &= R(t) + \gamma H(t) \\
 s(t) &= -S(t) \frac{N(t+1) - N(t)}{N(t)} \\
 l(t) &= -L(t) \frac{N(t+1) - N(t)}{N(t)} \\
 i(t) &= -I(t) \frac{N(t+1) - N(t)}{N(t)}
 \end{aligned} \tag{1}$$

The total individuals, the susceptible, the latently infected, the actively infected, the hospitalized isolated, and the recovered during the free transmission phase were respectively  $N(t)$ ,  $S(t)$ ,  $L(t)$ ,  $I(t)$ ,  $H(t)$ ,  $R(t)$ ,  $D(t)$  in period  $t$ . And  $N(t)$  is exogenous. The outflows of the susceptible, the exposed and infected individuals without insolation were respectively  $s(t)$ ,  $l(t)$ ,  $i(t)$ . The COVID-19 transmission rates of the infected and the exposed individuals are  $\beta_1$  and  $\beta_2$ , respectively. The number of high-frequency interaction (same household members) links between infected individuals and the susceptible ones was  $K_1$ , and the infected probability is  $\lambda_1$ , which was the same for the latently infected subpopulation. The number of low-frequency interaction (general acquaintances) links between infected individuals and the susceptible ones was  $K_2^I$ , while that was  $K_1^L$  for the latently infected subpopulation, and the infection probability was  $\lambda_2$ . So, it could be summarized as  $\beta_1 = \lambda_1 \cdot K_1 + \lambda_2 \cdot K_2^I$ , and  $\beta_2 = c\lambda_1 \cdot K_1 + c\lambda_2 \cdot K_2^L$ .  $C$  was the magnitude coefficient of the infected probability of the latently infected individuals relative to that of the actively infected ones. The latently infected would be converted to the infected with a probability  $\alpha$ . A proportion  $\eta$  of actively infected individuals was admitted to hospital and isolated. The probability of recovery among actively infected individuals admitted to hospital was  $\gamma(t)$ .

For the basic reproductive number  $R_0$ , it means that an infected person will spread the disease to  $R_0$  individuals on average without any external intervention. Generally,  $R_0 = \beta D$  <sup>(8)</sup>, which means, transmission from one infected individual to  $R_0$  susceptible individuals within  $D$  days of the infection period. The following equations are satisfied:

$$R_0 = \beta_1 D^I + \beta_2 D^L \quad (2)$$

$$D = D^L + D^I \quad (3)$$

The model was applied to the free spread stage before the implementation of control measures, with a  $R_0$  value of 2.2, the mean latent period  $D^L$  of 5.2. And the mean of transmission period  $D$  was set to 9.1 with reference to early studies on Wuhan epidemic (6, 8). Then, the mean infection period  $D^I$  was calculated as 3.9. Here, the infection period was set as a time window between the illness onset with the ability to transmit and loss of the ability to transmit by isolation.

For the disease transmission rate  $\beta$ ,  $\beta = \lambda K$  <sup>(2)</sup>. An actively or latently infected individual was exposed to mean  $K$  susceptible individuals per day, and the susceptible individual was exposed to  $\lambda$  probability of being infected, and an infected or latent individual transmits the disease to  $\beta$  susceptible individuals per day. If the heterogeneity of the exposure frequency was considered, there are:

$$R_0 = (\lambda_1 K_1 + \lambda_2 K_2^I) D^I + (c\lambda_1 \cdot K_1 + c\lambda_2 \cdot K_2^L) D^L \quad (4)$$

The number of high frequency interaction links  $K_1$  between infected individuals and the susceptible ones was set to the number of other members from the same household. Based

on the number of households and the permanent residents in the 2018 Statistical Yearbook of Wuhan City<sup>(9)</sup>, the average household size in Wuhan was 2.7. So,  $K_1$  was calculated as 1.7, which was same for the latent subpopulation. The susceptible individuals with high-frequency interaction were approached every day throughout the transmission period.

Based on the reported daily data from the China National Health Commission<sup>(10)</sup>, we obtained the total links  $K^{tot}$  was 10.2 during the infection period using the ratio of new daily traced close contacts to new daily confirmed cases. The average total number of the susceptible individuals with low-frequency interaction approached by per infected person was calculated to be 8.5 by  $K^{tot} - K_1$ . During the latent period, the normal daily activities were normally carried out for the latently infected without symptoms, and the number of low-frequency links  $K_2^L$  between the latently infected and susceptible individuals was calculated to be 1.6 by  $K_2^L = \frac{K^{tot} - K_1}{D^L}$ . However, those who were actively infected during the infection period are confined by their physical conditions and stay at home. In our model, the average number of low-frequency contact links  $K_2^I$  between the latently infected and susceptible individuals was set at 0, and the magnitude of the latent infection rate relative to the active infection rate  $c$  was set at 0.8.

Susceptible individuals with high-frequency interaction are more likely to be infected than those who are exposed to low-frequency interaction. We set the infection rate of high-frequency interaction was 1.46 times higher than that of low-frequency interaction individuals with reference to a study of SARS-CoV (5),  $\lambda_1 = 1.46\lambda_2$ .

A portion  $\alpha$  of latently infected individuals would convert to the actively infected, which was calculated as 0.19 by  $\alpha = \frac{1}{D^L}$  (2). The proportion  $\eta$  of actively infected individuals were admitted to hospital for isolation treatment, which was determined by the study of China Centers for Disease Control (CDC) on the difference between the epidemic curve of the illness onset and diagnosis of 70,000 confirmed cases<sup>(8)</sup>, and the result of Li et.al about the statistics of patients who visit medical clinic before January 22, so  $\eta$  valued 0.128. The probability  $\gamma(t)$  of recovery among actively infected individuals admitted to hospital was calculated as the ratio of the number of new daily cures to the cumulative number of confirmed cases two days earlier. It hovered around 0.9% before February 7, and then the cure rate increased significantly according to the reported data from National Health Commission (Supplementary Table S1).

The total dynamic population of Wuhan  $N(t)$  was exogenous, which was determined by real-world population movement data from “Baidu Migration” Big Data Platform<sup>(11)</sup>. For initial values in model, we set the initial value of the latently infected  $L(1)$  to 0, the actively infected  $I(1)$  to 1, the hospitalized infected  $H(1)$  to 0, and the recovered  $R(1)$  to 0.

### A multi-regional SLIHR model based on Wuhan population outflow network

As of January 25, all provinces except Xizang Province had launched Level 1 emergency response, which means that the COVID-19 was intervened in the control stage. At this control stage, the COVID-19 had spread to multiple regions all over Mainland China, and the SARS-CoV-2 was confined to several separate closed systems under traffic restrictions. At the same time, the treatment level of COVID-19 had been improving with the increase in viral awareness and clinical cases, and the cure rate parameter reflected the time-varying characteristics. These characteristics of COVID-19 transmission during the control stage made the model of the free spread stage not applicable to the control stage. Therefore, there is a demand for a multi-stage, multi-layered, multi-regional model, so as to simulate the COVID-19 dynamic epidemic in all regions of Mainland China.

The COVID-19 epidemic governance reflected the characteristics of local control, and they various from each other. Interregional movement was difficult under traffic restrictions, and each region could be considered as a separate closed system, which means the COVID-19 spreading within the region and incapable of acting on other regions. And we assumed that the transmission mechanism of COVID-19 was consistent across regions, in other words, the epidemiological parameters were the same.

Different from the classical SEIR model, the cure rate of the actively infected to the recovered was a function of time  $f'(t)$ , so the model was no longer an autonomous system. Medical resources would meet the need for timely access to isolation for diseased infected patients except Hubei Province, so the control stage model did not distinguish un-isolated infected compartment and hospitalized infected compartment.

The multi-regional epidemic model applied to the control stage was as follows:

$$\begin{aligned}
 S'(j, t+1) &= S'(j, t) - \beta I'(j, t) \frac{S'(j, t)}{N'(j)} - \beta L(j, t) \frac{S'(j, t)}{N'(j)} \\
 L'(j, t+1) &= L'(j, t) + \beta I'(j, t) \frac{S'(j, t)}{N'(j)} + \beta L(j, t) \frac{S'(j, t)}{N'(j)} - \alpha L'(j, t) \quad (5) \\
 I'(j, t+1) &= I'(j, t) + \alpha L'(j, t) - f'(t) I'(j, t) \\
 R'(j, t+1) &= R'(j, t) + f'(t) I'(j, t)
 \end{aligned}$$

Here,  $j$  represented region, and  $t$  represented time. The transmission parameters referred to studies on Mainland China outside Wuhan<sup>(12, 13)</sup>, with a  $R_0$  value of 2.57 (90% CI 2.37-2.78). The cure rate as a function of time was fitted from the reported data reported by the National Health Commission in China.

Initial values of the susceptible  $S'(j, t)$ , the latently infected  $L'(j, t)$ , the actively infected  $I'(j, t)$  and the recovered  $R'(j, t)$  in each region was based on the estimated number of latently and actively infected individuals outflowing from Wuhan during the free spread stage.

$$\begin{aligned}
S'(j,1) &= N(j) - L'(j,1) - I'(j,1) \\
L'(j,1) &= \rho_j A(T) \\
I'(j,1) &= \rho_j B(T) \\
R'(j,1) &= 0 \\
A(t+1) &= A(t) + \beta' A(t) + \beta' B(t) - \alpha A(t) + e(t) \\
B(t+1) &= B(t) + \alpha A(t) + i(t)
\end{aligned} \tag{6}$$

Based on the daily outflow of the latently infected  $e(t)$  and the actively infected  $i(t)$  in Wuhan from the free spread stage model (Eq.1), the total number of the latently infected  $A(T)$  and the actively infected  $B(T)$  in all regions of Mainland China except Wuhan on Jan 24 could be calculated. The regional share parameters  $\rho_j$  were obtained from the proportion of the population flowing out from Wuhan to each region. Ultimately, the number of latently infected individuals  $A(T)$  in all regions of Mainland China except Wuhan was 1,323 and that of the actively infected  $B(T)$  was 795 by Jan 24.

### Population flow

The population flow network was estimated based on the "Baidu Migration" Big Data Platform, a database that reflects the size of the regional inflow or outflow of population based on the geographic change of the users' mobile devices. According to transportation of Wuhan in 2019 Chinese New Year, 14.69 million passengers were sent by railway, highway and air, the share parameter of the "Baidu Migration" scale index is estimated as 34.8 thousand. According to the first confirmed case notified by Wuhan Health Committee, and therefore we set December 08, 2019 as the starting point of the simulation. Combined with the "Baidu Migration" data, we estimated the total population in Wuhan as of January 24, 2020  $N(48)$  to be 8.93 million, which was basically in line with the household population of 9.06 million published in 2019 Wuhan Statistics Bulletin<sup>(14)</sup>. From the start of 2020 Chinese New Year transportation on Jan 10 until the complete Wuhan travel ban on Jan 24, the population outflow from Wuhan was estimated as 3.362 million. If we considered the period from December 08, 2019 when the first confirmed case onset to January 24 when Wuhan travel ban by all vehicles, it was estimated that the outflow of population from Wuhan would be 5.07million.

### Provincial control effect

After January 24, due to movement restriction across provinces, the local spread of COVID-19 was the main channel, as described by the second stage of our model. COVID-19 spread in each province was independent of each other, and the control measures present heterogeneity. By means of simulating probable actively infected individuals without intervention and then comparing them with real-world reported confirmed cases, the heterogeneity in the effectiveness of different responses across provinces could be



quantitatively evaluated. The data of real-world confirmed cases in each of these provinces were reported by the China National Health Commission. We analyzed the reduced number and proportion of probable actively infected individuals compared to the real-world reported accumulative confirmed cases due to province responses in Mainland China outside Wuhan. Here, we focused on the transmission of infectious diseases, so simulated actively infected individuals were chosen as comparison objects in view of low cure rate in the simulation period. In order to avoid inconsistent data caliber of the confirmed cases data after February 12 due to the temporary inclusion of the “clinically diagnosed” cases without RT-PCR confirmation, our analysis period was from January 24 to February 11.

### **Provincial responses**

The effect of local responses to epidemic depends on the timeliness and enforcement, especially for this emergency infectious disease, timely measures are more significant. Here we measured provincial responses according to the time of each measure implementation, and put enforcement on hold since difficult quantification. To some extent, government policy making embodied tendentiousness to different provincial responses, which influenced sufficient enforcement. We described epidemic responses in each province in six dimensions and classified control measures into three levels based on the timeliness of policy implementation. All data/information were comprehensively collected from public provincial government documents, announcements and press conferences, except for a few data about “big data tool” was from news reports and work resumption rate referred to special reports on epidemic impact tracking in Information Website of Development Research Center of the State Council<sup>(15)</sup>.

### Supplementary Tables

(The supplementary Table 1 and 2 can be found with the link : <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8014041/>)

Supplementary Table S3: Parameters of early epidemic model in Wuhan. Parameters applied to the open SLIHR model considering incomplete isolation and heterogeneous interaction in the free spread stage. The probability of recovery after February 7 was calculated by exponential fitting function  $0.0045 \times e^{(t-15) \times 0.0786}$  based on the reported recovered individuals by China National Health Commission.

Parameter	Definition	Mean	Sources
$R_0$	basic reproductive number	2.2	(14,6)
$\beta_1$	transmission rate of the actively infected	0.204	
$D^I$	mean infection period	3.9	(6,8,15)
$\beta_2$	transmission rate of latently infected	0.271	
$D^E$	mean latent period	5.2	
$K_1$	high-frequency interaction links	1.7	
$K_2^I$	low-frequency interaction links between the actively infected and susceptible individuals	0	
$K_2^E$	low-frequency interaction links between the latently infected and susceptible individuals	1.6	(5, 8, 9, 16)
$\lambda_1$	infected probability of high-frequency interaction links	0.08	
$\lambda_2$	infected probability of low-frequency interaction links	0.12	
$c$	coefficient of the latent infection rate relative to the active infection rate	0.8	
$\alpha$	latent-to-active conversion portion	0.19	(6)
$\eta$	proportion of actively infected individuals admitted to hospital for isolation	0.128	(8)
$\gamma$	probability of recovery before February 7	0.009	(16)

## References

1. J. T. Wu, K. Leung, G. M. Leung, Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *The Lancet* 395, 689-697 (2020). doi:10.1016/S0140-6736(20)30260-9
2. M. Small, C. K. Tse, Small world and scale free model of transmission of SARS. *International Journal of Bifurcation and Chaos* 15, 1745-1755 (2005). doi:10.1142/S0218127405012776
3. M. Girvan, M. E. Newman, Community structure in social and biological networks. *Proceedings of the national academy of sciences* 99, 7821-7826 (2002). doi:10.1073/pnas.122653799
4. N. Zhu *et al.*, A novel coronavirus from patients with pneumonia in China, 2019. *New England Journal of Medicine* 382, 727-733 (2020). doi:10.1056/NEJMoa2001017
5. N. Jia, L. Tsui, Epidemic modelling using SARS as a case study. *North American Actuarial Journal* 9, 28-42 (2005). doi:10.1080/10920277.2005.10596223
6. Q. Li *et al.*, Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. *New England Journal of Medicine* 382, 1199-1207 (2020). doi:10.1056/NEJMoa2001316
7. I. Dorigatti *et al.*, Report 4: Severity of 2019-novel coronavirus (nCoV). <https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/Imperial-College-2019-nCoV-severity-10-02-2020.pdf>, (2020).
8. The Novel Coronavirus Pneumonia Emergency Response Epidemiology Team, The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19)—China, 2020. *China CDC Weekly* 2, 113-122 (2020). doi:10.46234/ccdcw2020.032
9. Wuhan Municipal Bureau of Statistics. <http://tjj.wuhan.gov.cn/tjfw/tjnj/> (2020).
10. China National Health Commission, Novel coronavirus pneumonia notification. [http://www.nhc.gov.cn/xcs/yqtb/list\\_gzbd\\_2.shtml](http://www.nhc.gov.cn/xcs/yqtb/list_gzbd_2.shtml), (2020).
11. Baidu Online Network Technology (Beijing) Co.,Ltd, Baidu Map Smart. <https://qianxi.baidu.com>, (2020).
12. M. Chinazzi *et al.*, The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science* 368, 395-400 (2020). doi:10.1126/science.aba9757
13. H. Tian *et al.*, An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China. *Science* 368, 638-642 (2020). doi:10.1126/science.abb6105
14. W. M. B. o. Statistics, Statistical Bulletin of Wuhan City. <http://tjj.wuhan.gov.cn/tjfw/tjgb/>, (2020).
15. Development Research Center of the State Council. Council, Special Reports on Epidemic Impact Tracking. <http://www.drcnet.com.cn/www/int/Leaf.aspx?uid=9101&version=integrated&leafid=26493>, (2020).
16. China National Health Commission, Epidemic situation report. [http://www.nhc.gov.cn/xcs/yqtb/list\\_gzbd.shtml](http://www.nhc.gov.cn/xcs/yqtb/list_gzbd.shtml), (2020).



# Chapter 10

**Distinct effectiveness in containing COVID-19 epidemic:  
comparative analysis of two cities in China  
by mathematical modeling**

Yunpeng Ji, Pengfei Li, Qinyue Zheng, Zhongren Ma, Qiuwei Pan

PLOS Global Public Health. 2021.1(11): e0000043



## Abstract

For better preparing future epidemic/pandemic, important lessons can be learned from how different parts of China responded to the early COVID-19 epidemic. In this study, we comparatively analyzed the effectiveness and investigated the mechanistic insight of two highly representative cities of China in containing this epidemic by mathematical modeling. Epidemiological data of Wuhan and Wenzhou was collected from local health commission, media reports and scientific literature. We used a deterministic, compartmental SEIR model to simulate the epidemic. Specific control measures were integrated into the model, and the model was calibrated to the recorded number of hospitalized cases. In the epicenter Wuhan, the estimated number of unisolated or unidentified cases approached 5000 before the date of city closure. By implementing quarantine, a 40% reduction of within-population contact was achieved initially, and continuously increased up to 70%. The expansion of emergency units has greatly reduced the mean duration from disease onset to hospital admission from 10 to 3.2 days. In contrast, Wenzhou is characterized as an emerging region with large number of primarily imported cases. Quick response effectively reduced the duration from onset to hospital admission from 20 to 6 days. This resulted in reduction of  $R$  values from initial 2.3 to 1.6, then to 1.1. A 40% reduction of contact through within-population quarantine further decreased  $R$  values until below 1 (0.5; 95% CI: 0.4-0.65). Quarantine contributes to 37% and reduction of duration from onset to hospital admission accounts for 63% to the effectiveness in Wenzhou. In Wuhan, these two strategies contribute to 54% and 46%, respectively. Thus, control measures combining reduction of duration from disease onset to hospital admission and within-population quarantine are effective for both epicenters and settings primarily with imported cases.

**Keywords:** COVID-19 epidemic, Wuhan, Wenzhou, Comparative analysis, Mathematical modeling

## Introduction

The global population is constantly facing the threats of emerging infectious diseases. Different countries and regions often react differently in response to outbreaks, whereas the right early response is essential for containing the outbreak, thereby avoiding large epidemic or pandemic. Because of their disparities in culture, socioeconomic status, and types of government, the implementation of control measures can vary dramatically among different countries.

For better preparing future epidemic/pandemic, important experiences and lessons could be learned from how the COVID-19 epidemic was responded from different regions of China, in particular at their different stages. COVID-19 was sparked in December, 2019 from Wuhan, the capital of Hubei province, with a population of 11 million<sup>[1]</sup>. Due to initial delay of necessary actions, the local government missed the first chance of containing the epidemic. Until January 23, the central government imposed heavy control measures, including city lockdown, travel ban, and within-population quarantine. The total confirmed cases have finally climbed up to nearly 50,000 cases in Wuhan, but was controlled subsequently.

COVID-19 outbreak was coincided with a massive population migration, because of the Chinese lunar new year holiday<sup>[2]</sup>. This has led to the rapid spread across China. Of particular relevance is Wenzhou, a prefecture-level city of Zhejiang province with a total population of 9 million. It is 900 km away from Wuhan, and about 170,000 Wenzhou businesspeople are working there<sup>[3]</sup>. The first case in Wenzhou was identified on January 21, who returned from Wuhan. It quickly became the highest figure of COVID-19 cases for any city outside of Hubei province, and most of these cases were imported from the epicenter<sup>[3]</sup>. However, it took only about 46 days for this city to fully contain the epidemic, otherwise bearing high risk of growing into a new epicenter.

The distinct experiences of these two cities are mirroring what has happened in many parts of the world, either as epicenters or regions primarily with imported case. In this study, we aim to comprehensively compare the effectiveness of containing COVID-19 between Wuhan and Wenzhou by mathematical modeling, and to provide mechanistic understanding of how to effectively contain the epidemic at early stage.

## Methods

### Data collection

We systematically collected the epidemical data on SARS-CoV-2 transmission in Wuhan and Wenzhou. For Wuhan (Supplementary Table S1), data regarding onset distribution of early



identified cases before January 15, and number of cumulative and hospitalized cases were collected from the Health Commission of Hubei Province and previous studies [4-6]. For Wenzhou (Table 1), we collected data of all identified cases recorded by the Health Commissions of both Zhejiang Province and Wenzhou city<sup>[3]</sup>.

## Model structure and control measures

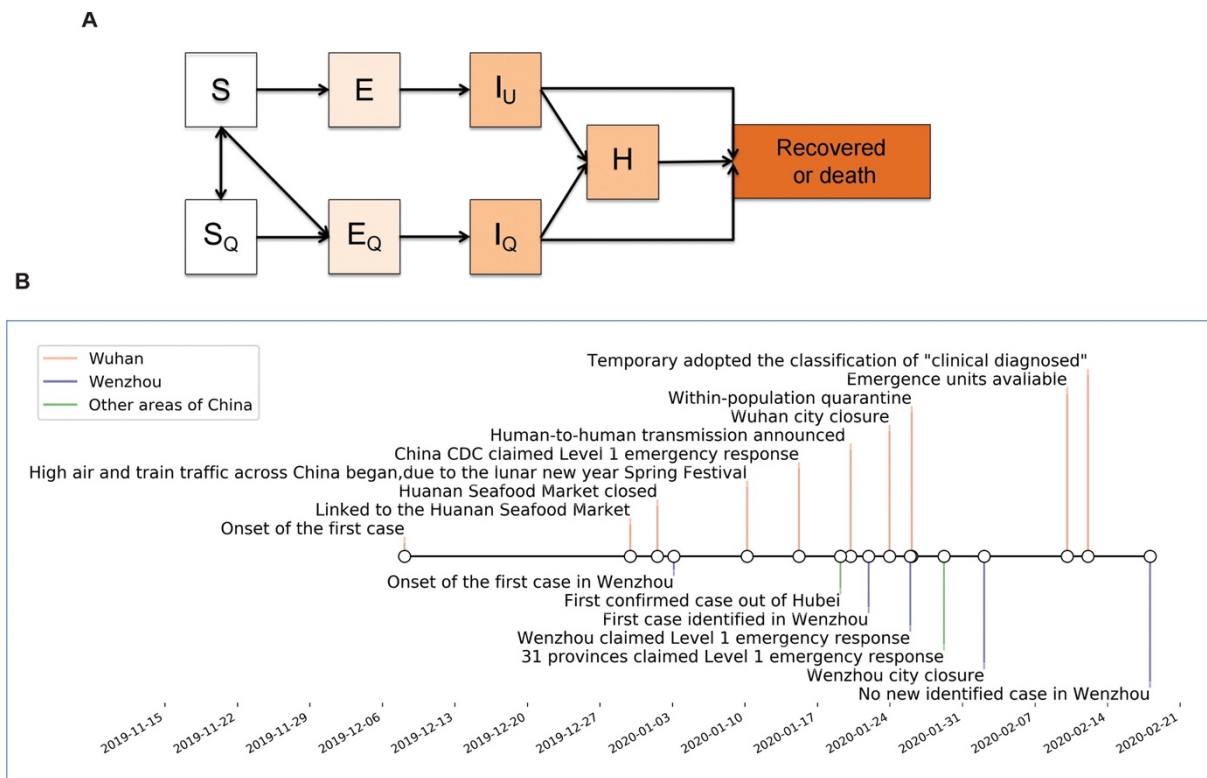
We focused on two major control measures, isolation of symptomatic cases and reduction of within-population contact rates. To estimate the impact of the measures, we constructed a deterministic, compartmental model for SARS-CoV-2 transmission, in which a standard susceptible-exposed-infectious-recovered (SEIR) structure was modified to accommodate quarantine and isolation (Figure 1A). The model was fitted to data on hospitalized cases ( $H(t)$ ) in Wuhan and Wenzhou from early January to February 2020. The model incorporates data on city population size, the current knowledge of natural history of SARS-CoV-2, and other relevant parameters.

Briefly, we assumed that all the citizens are initially susceptible ( $S$ ), and a fraction  $q$  of all persons infected by an infectious case is successfully quarantined ( $E_Q$ ), and a fraction  $q$  of all persons contacted but not infected by an infectious case is also quarantined ( $S_Q$ ). The infectious compartment is composed of cases who are unidentified or unisolated ( $I_U$ ), those who develop from  $E_Q$  ( $I_Q$ ), and those who have been hospitalized ( $H$ ). We assumed that cases in  $I_Q$  and  $H$  are successfully quarantined and would not infect other susceptible people.

We assumed that an infectious individual has a mean of  $c$  potentially infectious contacts per day, that susceptible contacts are infected with probability  $\beta$ , and that the number of contacts was independent of population density. We further assumed that individuals are hospitalized and correspondingly isolated at a fixed rate per day after becoming infectious, and that isolated individuals are no longer at risk of transmitting the virus. Infected individuals become noninfectious by dying, recovering, or being hospitalized, and the mean duration of infectiousness is  $D$  days.

We used least-square fitting to look for the model trajectory that best matches the epidemic time series and to estimate the parameters. Specifically, we fit the hospitalized number of cases given by equation  $H(t)$  to the hospitalize number of case notifications. Sets of realizations of the best-fit curve  $H(t)$  were generated using parametric bootstrap [7], in which 200 realizations were made. Each realization of the cumulative number of case notifications  $H_i(t)$  ( $i = 1, 2, \dots, m$ ) is generated as follows: for each observation  $C(t)$  for  $t = 2, 3, \dots, n$  days generates a new observation  $H_i^*(t)$  for  $t \geq 2$  ( $H_i^*(1) = C(1)$ ) that is sampled from a Poisson distribution with mean  $H(t) - (t - 1)$  (the daily increment in  $H(t)$  from day  $t - 1$  to

day  $t$ ). The corresponding realization of the cumulative number of influenza notifications is given by  $H_i(t) = \sum_{j=1}^t H_i^*(t)$ , where  $t=1, 2, 3, \dots, n$ . Detailed methods for estimating  $R_0$  and  $R_t$  were described in Supplementary file.



**Figure 1.** (A) Mathematical model for COVID-19 transmission. Susceptible individuals are infected by unidentified individuals, and become infectious themselves after an interval of incubation. Infectious individuals lose infectiousness by death, recovery, or successful isolation. No new births or deaths of other causes are considered. When quarantine started, a proportion,  $q$ , of new infections is quarantined before they become infectious. The same proportion ( $q$ ) of susceptible individuals who have contacts with infectious persons but were not infected are also quarantined. Susceptible individuals are released from quarantine after 14 days. For simplicity, we assume that quarantined individuals are isolated and not infectious. (B) Timeline of important events occurred during COVID-19 outbreaks in Wuhan and Wenzhou.

### Estimating control measures in Wuhan and Wenzhou

Control measures were introduced with different time points (Figure 1B). In the model of Wuhan epidemic, we started on January 15, on which China CDC claimed to execute level 1 emergence. We set two time points. One is around January 27 when within-population quarantine was implemented, which would reduce the within-population contact rate. Another is around February 10, since then the newly built emergency medical units were operational. It mainly accelerates isolation of the unidentified cases. To evaluate the measures, we arbitrarily delayed the time point by 7 days and analyzed the effect. Seeds of the model

were from the early 370 identified cases in Wuhan, because those were firstly proved to be associated with human-to-human transmission<sup>[4]</sup>.

For Wenzhou, the sources of SARS-CoV-2 transmission were all imported cases with travel history to Wuhan or other areas of China. We started the model on the date of onset of the first imported case (January 3), and took all imported cases as model seeds. The first time point was set on January 21 as identification and hospitalization of cases were started. The second and third time points were set based on the reduction of mean observed duration from onset to hospital admission (January 27) and the implementation of travel restriction within the city (early February). Because of the lack of information, we assumed that early mean duration from onset to hospital admission (before January 27) is close to that in Wuhan. Finally, we also performed sensitivity analysis based on the mean interval incubations (Supplementary Figure S1).

## Results

### Epidemiological feature of COVID-19 in Wenzhou

We systematically collected the epidemiologic information of in total 504 cases which were identified in Wenzhou by March 5 (Table 1, Figure 1B). We first analyzed the data based on the distribution of onset date. Because this information was missing for the early 31 cases identified from January 21 to 26, we assumed that the onset of these cases follows a Gaussian distribution and re-ranged the distribution of the 31 cases between January 7 and 17 (Figure 2A). The duration from onset to hospital admission was continuously decreased over the periods from January 3 to 20 (reported 20 days; Table 1), January 21 to 27 (assumed 12 days; Table 1), and then became stable as observed 6 days in average (Figure 2B), until there was no new case identified. Based on the available information of imported cases, the average incubation interval was estimated as 5.5 (95% CI, 4.8-6.2) days (Figure 2C).

### Modeling control measures in Wenzhou

How rapid an epidemic can spread largely depends on the reproductive number. The basic reproductive number,  $R_0$ , is defined as the number of secondary infections generated by one primary case in a totally susceptible population. The effective reproductive number,  $R$ , measures the number of secondary cases generated by an index case and the declines because of the reduction of susceptibility in the population and the use of effective control measures.  $R$  must be below 1 to stop an outbreak. The epidemic of Wenzhou was mainly caused by imported cases with travel history of Wuhan or other areas of China, which accounts for 43.7% (Figure 2A). The value of  $R_0$  is estimated to be 2.3 (95% CI, 2.2-2.5).

**Table 1. Parameters used in the quarantine simulation of Wenzhou epidemic.**

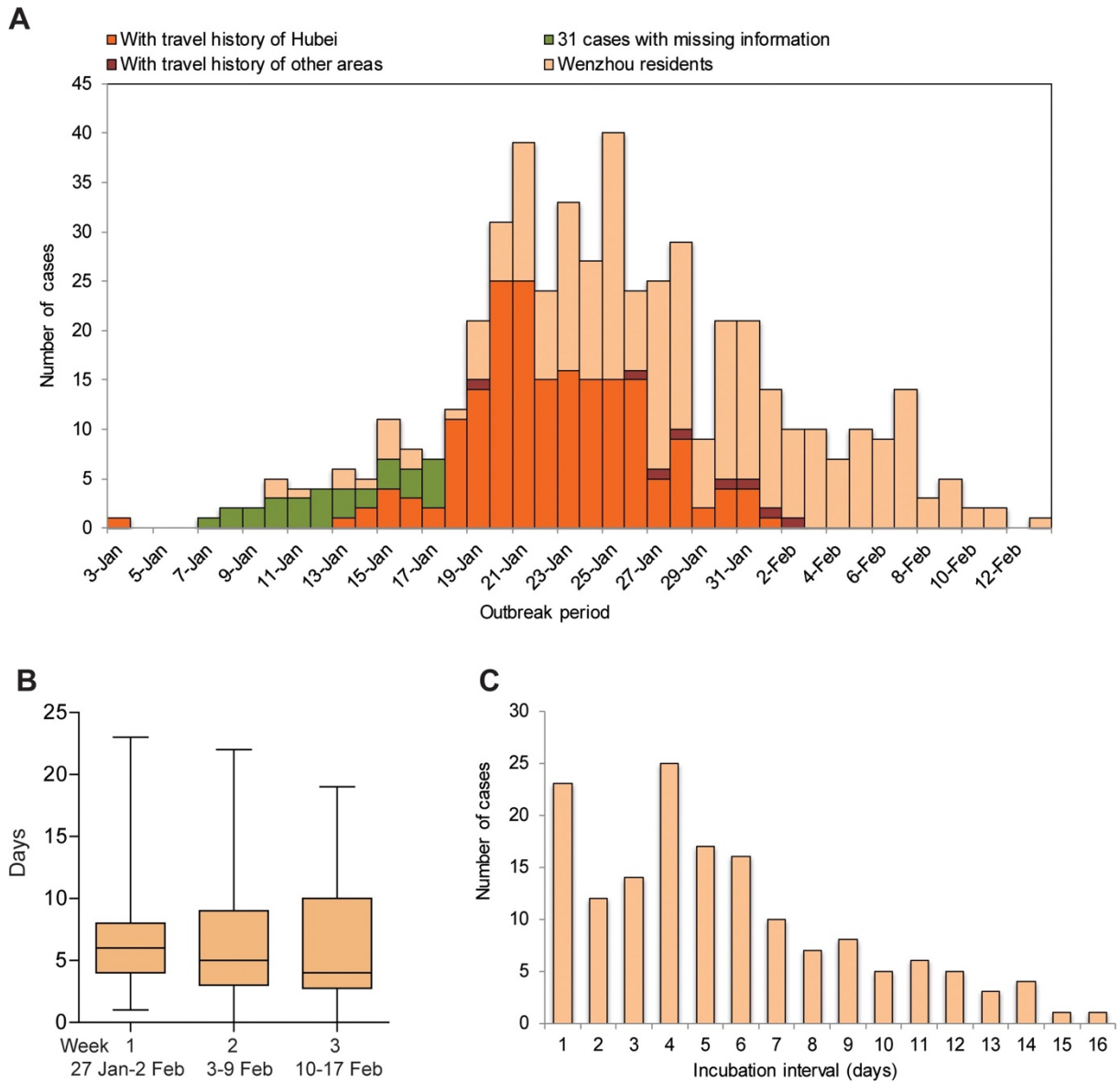
Parameter	Symbol	Time	Baseline value	Estimated value	Reference
Contact rate	$c$	3-20 Jan	10	-	Assumption
		21-26 Jan	10	-	Assumption
		27-31 Jan	10	-	Assumption
		1-17 Feb	-	6	Estimated
Mean duration from onset to hospital admission	$i$	3-20 Jan	20 days (18-21)	-	the Health Commission of Wenzhou
		21-26 Jan	12 days (10-14)	-	*Li, et al. 2020, NEJM
		27-31 Jan	6.1 days (5-8)	-	the Health Commission of Wenzhou
Mean Recovery rate	$\nu$	1-17 Feb	6.2 days (5-8)	-	the Health Commission of Wenzhou
		-	0.045 (0.04-0.066)	-	the Health Commission of Wenzhou
Mean interval of incubation	$p$	-	5.5 days (3-7)	-	the Health Commission of Wenzhou
Case fatality rate	$m$	-	0	-	the Health Commission of Wenzhou
Duration of quarantine	$r$	-	14 days	-	WHO recommendation

\*Note: Li, Q., et al. (2020). Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. *N Engl J Med.* Mar 26;382(13):1199-1207.

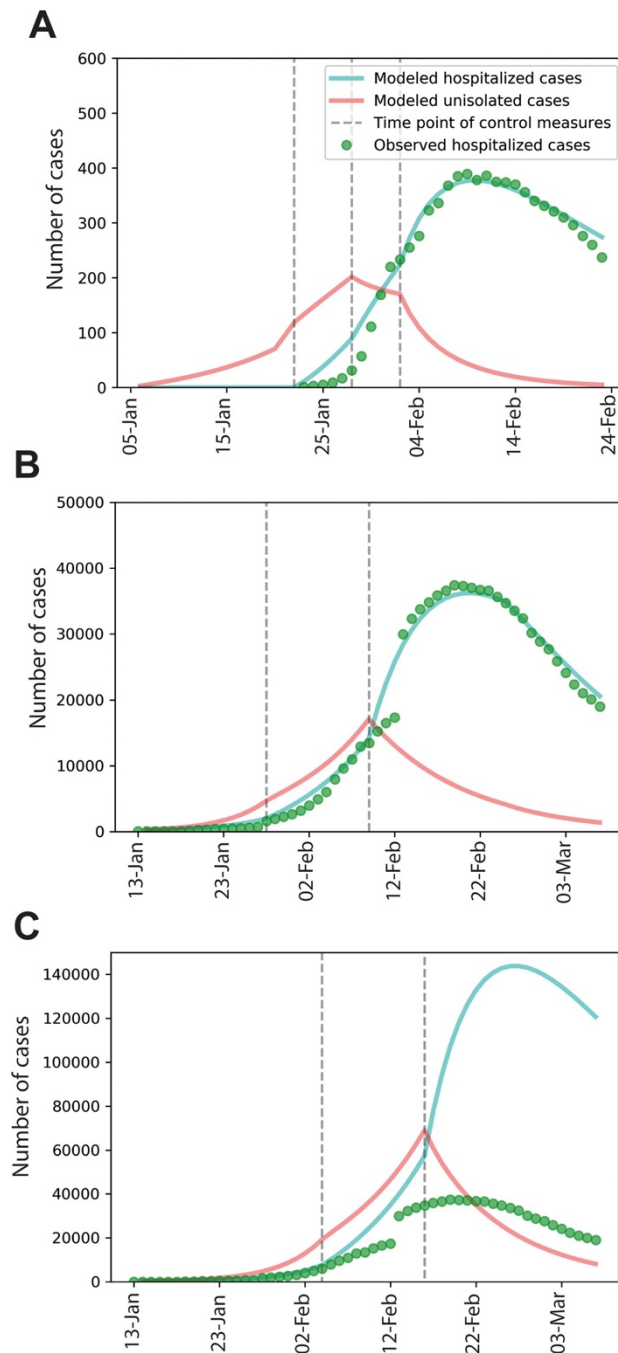
Our mathematical model has well-reproduced the COVID-19 epidemic with several key epidemic patterns identified (Figure 3). The unidentified cases were initially composed of only imported cases, but as the transmission proceeding more secondary cases from the local were gradually included. Importantly, the model shows that the epidemic has been fully contained and the number of unidentified cases is close to zero by February 17. This perfectly matches the real-world situation that no new cases were further identified. Since the first case identified on January 21, mean duration from disease onset to hospital admission were continuously decreased. This is important for preventing further spread by infected cases. Correspondingly, the modeled  $R$  values were reduced to 1.6 and 1.1 (Figure 4A). Finally, it reached 0.5 after implementation of within-city quarantine, which has dramatically reduced average daily contacts from initial 10 to 6.

### Modeling epidemiology and control measures in Wuhan

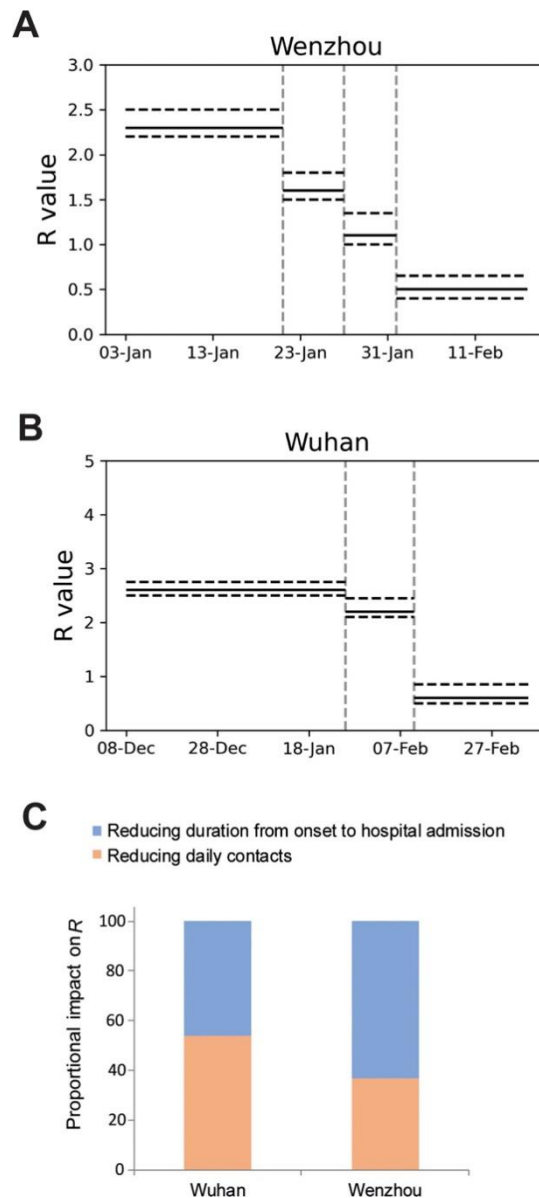
In the model of Wuhan epidemic, we have systematically collected the epidemiological data (Supplementary Table S1; Figure 1B). We found that both within-population quarantine and the use of emergency unit are effective (Figure 3B). After the first time point (January 26), modeled within-population contact rate was decreased from initial 10 to 6, and then 3.5 after the second time point (February 10). A seven-day delay of implementing these two measures would exacerbate the epidemic, and indirectly increase the burden of quarantine (Figure 3C). Our model estimated 983 unidentified or unisolated cases by March 5, 2020. According to the current effectiveness, it requires additional 40 days to contain all these unidentified cases.



**Figure 2. Epidemical features of COVID-19 in Wenzhou.** (A) Distribution of disease onset of all the identified cases. Because specific information of the early 31 imported cases is not available; we assumed that date of onset of all imported cases follows a Gaussian distribution, and arbitrarily ranged them in the period from January 7 to 17. (B) Information on duration from onset to hospital admission of 467 cases from January 27 to February 17. (C) Distribution of interval of incubation of 157 imported cases.



**Figure 3. Simulation of COVID-19 epidemics in Wenzhou and Wuhan.** (A) Modeled hospitalized and unidentified cases in Wenzhou from January 3 to February 21. The first dash line is on January 21 when the first case was identified. The second dash line is on January 27 when mean duration from disease onset to hospital admission reduced from 12 days to observed 6.1 days. The last dash line is on February 2 when a within-city quarantine started. (B) Modeled hospitalized and unidentified cases in Wuhan from January 15 to March 5. The first dash line is on January 26 when a within-city quarantine started. The second dash line is on February 10 when temporary emergency units were operational. (C) Modeled hospitalized and unidentified cases in Wuhan with a seven-day delay of the two control measures of (B).



**Figure 4. Estimating effective reproductive number  $R$ .** Estimated reproduction number for the period of the epidemic in Wenzhou (A) and Wuhan (B). Confidence intervals are also shown. (C) Proportion of impact of reducing contact rate and duration from onset to hospital admission to  $R$  values.

According to our modeling, prior to applying of effective control measures, the estimated number of unidentified or unisolated cases was always higher than that of hospitalized cases. Within-population quarantine started on January 26 was proven to be effective, reducing  $R$  from the estimated initial value of 2.6 to 2.2 (Figure 4B). After February 10, many of these hidden cases were isolated at the new emergency units temporarily built for this epidemic. Correspondingly, the modeled mean duration from onset to hospital was reduced from 10 to 3.2 days after February 10. Modeled  $R$  value within this period is about 0.6 (95% CI, 0.5-0.8) (Figure 4B), indicating the feasibility of containing the epidemic.

## Comparative analyses between Wuhan and Wenzhou

The initiation, development and outcomes of the epidemics in these two cities are evidently distinct. Before adopting within-population quarantine and travel restriction, the epidemic has already developed for more than one month (about 49 days) in Wuhan (Figure 2B). In addition to 630 hospitalized cases, the number of the unidentified cases was estimated to be around 5,000 by January 26 (Figure 3B).

By striking contrast, the response period in Wenzhou is much shorter. Before adopting quarantine, the estimated number of unidentified cases is 180 by February 2. This relatively low number allowed subsequent control by rapid allocation of medical resources per capita from the local authorities. Effectiveness of quarantine and reduction of the period for disease onset to hospital admission was compared in the construct of  $R$  (Figure 4C). Quarantine contributes to 37% and reduction of duration from onset to hospital admission accounts for 63% to the effectiveness in Wenzhou. In Wuhan, Quarantine contributes to 54% and the other accounts for 46%, respectively (Figure 4C).

## Discussion

The transmission dynamics of a virus is primarily determined by the effective reproductive number. These  $R$  values often consistently evolve attributing to viral adaption, susceptibility of the targeting population, environmental changes, potential SSE and implementation of control measures. Although the basic reproductive number  $R_0$  of SARS-CoV-2 (2.6) appears even lower than that of SARS-CoV (2.9) <sup>[8]</sup>, the speed and scale of COVID-19 spreading greatly surpass SARS and MERS <sup>[9]</sup>. The drastic escalation of the epidemic in Wuhan mainly attributed to the non-responsiveness of the local authorities at the early stage. Incredibly, a massive annual potluck banquet for 40,000 families was continued to be held in Wuhan on January 18, which very likely exacerbated the outbreak. Mass gatherings could impose high risks of super-spreading event (SSE) <sup>[10]</sup>. We suspected that SSE likely occurred in Wuhan at the early stage. However, we do not have actual data to confirm the occurrence of SSE, and therefore it is not specifically emphasized in our model.

The turning point for coping with the epidemic in Wuhan was the implementation of vigorous public health measures directly ordered by the central government. From January 23, the central government began to implement heavy control measures, including city lockdown, travel ban, and within-population quarantine. The implantation of travel ban of Wuhan was already too late to have major impact on the spread in Mainland China, although it was estimated to have substantial impact at the international scale<sup>[11]</sup>. In our analysis, we emphasized two important factors that determine the outcomes in containing the epidemic



including effectiveness of quarantine, and the duration from disease onset to hospital admission. A major challenge for Wuhan was that the numbers of confirmed and suspected COVID-19 cases were already too high at the time of implementing measures. This has plunged the local healthcare system into crisis, and all the resources have to exclusively dedicate to contain this outbreak. Since late January and early February, the central government has mobilized strong medical forces and emergency medical supplies, and new emergency units were built in very short-term. These efforts have leveraged the levels of quarantine, and shortened the duration from disease onset to hospital admission that prevented spreading by infected cases.

One important reason for the wide and quick spread of COVID-19 to other parts of China is that the Wuhan outbreak coincided with the Chinese lunar new year holiday. It is a massive population migration with estimated 3 billion individual trips to take place<sup>[2]</sup>. Wuhan is a large hub connecting different parts of China through railways and an international airport, and five million people have left Wuhan before the travel ban<sup>[12]</sup>. A heavily affected city in other part of China is Wenzhou, because there are an estimated 170,000 Wenzhou businesspeople working in Wuhan<sup>[3]</sup>. Because of the returnees from Wuhan, Wenzhou had the highest figure of imported COVID-19 cases for any city outside Hubei province<sup>[3]</sup>. However, it only took 46 days to fully contain the epidemic, otherwise bearing high risk of growing into a new epicenter. Based on our analysis, the key to achieve this outcome is attributed to the quick response from the local authorities and the general public. The government has rapidly implemented restrictive measures including city closure, travel ban, effective isolation and quarantine, and rapid case identification and hospital admission. The public is highly aware of the contagiousness of SARS-CoV-2, and spontaneously adopt social distancing measures such as cancellation of public gatherings and holiday visits, and minimizing outdoor activities. This resulted in continuous decrease of  $R$  values to 1.6, 1.1 and 0.5, and no new case was further reported in Wenzhou.

Given the nature of retrospectively analyzing the outcomes of the COVID-19 epidemics, our findings are not of completely unexpected. However, we provided quantitative analysis of the impact of reducing duration from disease onset to hospital admission and within-population quarantine. Importantly, we have comparatively taken two real-world examples with unique characteristics as a COVID-19 epicenter and a setting with mainly imported cases. These findings would add new insight in understanding control measures at early stages of the COVID-19 epidemics, and shall serve as an importance reference for quick response to future new outbreaks/epidemics with full consideration of the real-world situations.

Because of the disparities in culture, socioeconomic status and types of government, the awareness, preparedness and responsiveness towards emerging threats vary dramatically

among different countries<sup>[9,13-17]</sup>. Our results highlight the essential of early, quick and adequate response with implementation of the right control measures. Delayed response will soon plunge the healthcare system into crisis, as seen in Wuhan and other countries like Italy, Spain, UK and USA, even with well-developed healthcare systems. Availability of healthcare is critical for controlling the epidemic but also for minimizing severe patient outcomes<sup>[18]</sup>. As seen in Wenzhou, rapid response has avoided to emerge as a new epicenter, and soon contained the epidemic. On the other hand, understanding and cooperation from the general public are also extremely important for the effectiveness of these control measures<sup>[19]</sup>.

In summary, we comparatively analyzed the epidemiology, estimated the transmission dynamics, and assessed the effectiveness of control measures between Wuhan and Wenzhou by monitoring the dynamics of the effective reproductive number  $R$ . Our findings suggest that combination of reducing the interval from disease onset to hospital admission and stringent within-population quarantine are effective for both epicenters and regions mainly caused by imported infections. These two cities in China represent real-world examples that can serve as important lessons for improving preparedness and early response to future emerging epidemic/pandemic.

## References

1. Zheng Q, Wang X, Bao C, Ji Y, Liu H, Meng Q, et al. A multi-regional, hierarchical-tier mathematical model of the spread and control of COVID-19 epidemics from epicentre to adjacent regions. *Transbound Emerg Dis*. 2021. Epub 2021/02/05. doi: 10.1111/tbed.14019. PubMed PMID: 33539678; PubMed Central PMCID: PMC8014041.
2. Chen S, Yang J, Yang W, Wang C, Barnighausen T. COVID-19 control in China during mass population movements at New Year. *Lancet*. 2020;395(10226):764-6. Epub 2020/02/28. doi: S0140-6736(20)30421-9 [pii]10.1016/S0140-6736(20)30421-9. PubMed PMID: 32105609; PubMed Central PMCID: PMC7159085.
3. Virnik K, Rosati M, Medvedev A, Scanlan A, Walsh G, Dayton F, et al. website of wenzhou health commission. *PLoS One*. 2020;15(3): e0228163. Epub 2020/03/05. doi: 10.1371/journal.pone.0228163 PONE-D-19-35126 [pii]. PubMed PMID: 32130229.
4. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. *N Engl J Med*. 2020;382(13):1199-207. Epub 2020/01/30. doi: 10.1056/NEJMoa2001316. PubMed PMID: 31995857; PubMed Central PMCID: PMC7121484.
5. Muller GS, Faccin CS, Silva DR, Dalcin PTR. website of Hubei health commission. *J Bras Pneumol*. 2020;46(2):e20180419. Epub 2020/03/05. doi: S1806-37132020000200204 [pii]10.36416/1806-3756/e20180419. PubMed PMID: 32130332.
6. Special Expert Group for Control of the Epidemic of Novel Coronavirus Pneumonia of the Chinese Preventive Medicine A. [An update on the epidemiological characteristics of novel coronavirus pneumonia COVID-19]. *Zhonghua Liu Xing Bing Xue Za Zhi*. 2020;41(2):139-44. Epub 2020/02/15. doi: 10.3760/cma.j.issn.0254-6450.2020.02.002. PubMed PMID: 32057211.
7. Efron B TR. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Stat Sci*. 1986; 1: 54-75. doi: Doi 10.1097/00001648-199205000-00015. PubMed PMID: WOS: A1992HT91200016.

8. Peak CM, Childs LM, Grad YH, Buckee CO. Comparing nonpharmaceutical interventions for containing emerging epidemics. *Proc Natl Acad Sci U S A*. 2017;114(15):4023-8. Epub 2017/03/30. doi: 10.1073/pnas.1616438114. PubMed PMID: 28351976; PubMed Central PMCID: PMCPMC5393248.
9. Li Y, Xia L. WHO website. *AJR Am J Roentgenol*. 2020;1-7. Epub 2020/03/05. doi: 10.2214/AJR.20.22954. PubMed PMID: 32130038.
10. Ebrahim SH, Memish ZA. COVID-19: preparing for superspreader potential among Umrah pilgrims to Saudi Arabia. *Lancet*. 2020;395(10227): e48. Epub 2020/03/03. doi: S0140-6736(20)30466-9 [pii]10.1016/S0140-6736(20)30466-9. PubMed PMID: 32113506; PubMed Central PMCID: PMC7158937.
11. Chinazzi M, Davis JT, Ajelli M, Gioannini C, Litvinova M, Merler S, et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science*. 2020;368(6489):395-400. Epub 2020/03/08. doi: science. aba9757 [pii]10.1126/science.aba9757. PubMed PMID: 32144116; PubMed Central PMCID: PMC7164386.
12. Peeri NC, Shrestha N, Rahman MS, Zaki R, Tan Z, Bibi S, et al. The SARS, MERS and novel coronavirus (COVID-19) epidemics, the newest and biggest global health threats: what lessons have we learned? *Int J Epidemiol*. 2020;49(3):717-26. Epub 2020/02/23. doi: 5748175 [pii]10.1093/ije/dyaa033. PubMed PMID: 32086938; PubMed Central PMCID: PMC7197734.
13. Boldog P, Tekeli T, Vizi Z, Denes A, Bartha FA, Rost G. Risk Assessment of Novel Coronavirus COVID-19 Outbreaks Outside China. *J Clin Med*. 2020;9(2). Epub 2020/02/26. doi: 10.3390/jcm9020571. PubMed PMID: 32093043.
14. Bordi L, Nicastrì E, Scorzolini L, Di Caro A, Capobianchi MR, Castilletti C, et al. Differential diagnosis of illness in patients under investigation for the novel coronavirus (SARS-CoV-2), Italy, February 2020. *Euro Surveill*. 2020;25(8). Epub 2020/03/05. doi: 10.2807/1560-7917.ES.2020.25.8.2000170. PubMed PMID: 32127123.
15. Bernard Stoecklin S, Rolland P, Silue Y, Mailles A, Campese C, Simondon A, et al. First cases of coronavirus disease 2019 (COVID-19) in France: surveillance, investigations and control measures, January 2020. *Euro Surveill*. 2020;25(6). Epub 2020/02/20. doi: 10.2807/1560-7917.ES.2020.25.6.2000094. PubMed PMID: 32070465.
16. Pullano G, Pinotti F, Valdano E, Boelle PY, Poletto C, Colizza V. Novel coronavirus (2019-nCoV) early-stage importation risk to Europe, January 2020. *Euro Surveill*. 2020;25(4). Epub 2020/02/06. doi: 10.2807/1560-7917.ES.2020.25.4.2000057. PubMed PMID: 32019667; PubMed Central PMCID: PMCPMC7001240.
17. Zheng Q, Wang X, Bao C, Ma Z, Pan Q. Mathematical modelling and projecting the second wave of COVID-19 pandemic in Europe. *J Epidemiol Community Health*. 2021. Epub 2021/02/18. doi: jech-2020-215400 [pii]10.1136/jech-2020-215400. PubMed PMID: 33593851; PubMed Central PMCID: PMC8142418.
18. Ji Y, Ma Z, Peppelenbosch MP, Pan Q. Potential association between COVID-19 mortality and health-care resource availability. *Lancet Glob Health*. 2020;8(4): e480. Epub 2020/02/29. doi: S2214-109X(20)30068-1 [pii]10.1016/S2214-109X(20)30068-1. PubMed PMID: 32109372; PubMed Central PMCID: PMC7128131.
19. The L. COVID-19: too little, too late? *Lancet*. 2020;395(10226):755. Epub 2020/03/09. doi: S0140-6736(20)30522-5 [pii]10.1016/S0140-6736(20)30522-5. PubMed PMID: 32145772.

## Supplementary information

### 1. Transmission model for the effect of quarantine

The model depicted in Figure 2 consists of the following ordinary differential equations (Lipsitch, Cohen et al. 2003).

$$dS/dt = -cbI_u X/N + rS_q \quad (1)$$

$$dS_q/dt = qc(1 - b)I_u S/N - rS_q \quad (2)$$

$$dE/dt = -pE + cb(1 - q)I_u S/N \quad (3)$$

$$dE_q/dt = qcbI_u S/N - pE_q \quad (4)$$

$$dI_u/dt = pE - (v + m + i)I_u \quad (5)$$

$$dH/dt = iso(I_u + I_q) - (v + m)H \quad (6)$$

$$dI_q/dt = pE_q - (v + m + i)I_q \quad (7)$$

$$dR/dt = v(I_u + H + I_q) \quad (8)$$

$$dD/dt = m(I_u + H + I_q) \quad (9)$$

The model (including equations (1) to (9)) is a modification of the classical SEIR model, involving susceptible, infected but not yet infectious, infectious, and recovered/immune individuals in compartments  $S$ ,  $E$ ,  $I$  and  $R$  respectively. At time  $t=0$ , there are  $N$  people in the population, of whom certain number of individuals are exposed but not yet infectious ( $E$ ), and all the rest are susceptible ( $S$ ). In our model, the  $I$  compartment is composed of cases who are not isolated ( $I_u$ ) and those who have been isolated ( $H$ ).  $c$  is the number of daily contacts per person, and  $b$  is the probability of transmission per contact between a susceptible and an infectious person. We separated the parameters  $c$  and  $b$ , rather than using transmission parameter,  $\beta$ . We thus assume that each infectious person makes  $c$  contacts per day, of whom a proportion  $b$  is infected if the infectious person is undetected. We attained  $b$  based on the function  $R = \beta \times \text{mean duration of infectiousness}$ .  $1/r$  is the mean duration of quarantine for susceptible persons who are suspected to be infected but actually not ( $S_q$ ), they would flow back to susceptible ( $S$ ).  $1/p$  is the mean time for progression from latent to infectious.  $v$  is the per capita recovery rate,  $m$  is the per capita death rate, and  $i$  is the mean daily rate at which infectious cases are detected and isolated. In our model, we calculated  $i$  as the mean rate of cases from onset to hospital admission.

We made a simplified assumption that isolated cases ( $I_d$ ) do not infect anyone, and the rate of isolation  $i$  should be thought of as a rate of “effective” isolation. Under these assumptions, the mean duration of infectiousness is  $1/(v + m + i)$ . We estimated parameters by using least-square fitting to look for the model trajectory that best matches the epidemic time series. Specifically, we fit the hospitalized number of cases given by equation  $H(t)$  to the hospitalize number of case notifications (Figure 4A and B).

For epidemic in Wenzhou, we modified equation (5) as:

$$dI_u/dt = pE - (v + m + i)I_u + import \quad (5'),$$

where *import* is mean number of onset of imported cases with travel history of Hubei province or other areas (Figure 3A). At time  $t=0$ , we assumed that all the individuals in the model are susceptible, and imported infectious cases enter the model as rate of *import*.

## 2. Estimating $R_0$

We used a time-dependent method to compute  $R_0$  (Obadia, Thomas et al. 2012; Wallinga, Jacco et al. 2004). Briefly, the probability  $p_{ij}$  that case  $i$  was infected by case  $j$  is presented as

$$p_{ij} = \frac{N_{ij}w(t_i-t_j)}{\sum_{i \neq k} N_{ik}w(t_i-t_k)}$$

Where  $t_i$  is the onset time of case  $i$ , and  $k$  denotes the index of other case that are primary case of case  $i$ .  $w(t)$  is the generation interval, or the serial interval, which is the time from symptom onset of a primary case to that of its secondary case. As generation interval was not directly observed in this study, we assumed that the timing of transmission events is not skewed during the early term of the infectious period, and approximately calculated the interval as sum of the average incubation period and half of the average infectious period (Fine, Paul et al 2003).  $N_{ij}$  is the number of case  $i$  infected by case  $j$ , and  $N_i$  is the overall number of case  $i$ . The reproduction number for case  $j$  is the sum over all case  $i$ , weighted by the probability that case  $i$  is infected by case  $j$ :

$$R_j = \sum_i p_{ij},$$

and reproduction number over all cases with the same date of onset is

$$R_t = \frac{1}{N_t} \sum_{\{t_j=t\}} R_j.$$

Combined with equation derivation above, we calculated the basic reproduction number  $R_0$  by determining the initial exponential disease growth phase with the goodness-of-fit test statistic (Lipsitch, Marc et al. 2003), and obtained the confidence interval by model simulation.

## 3. Estimating time-dependent $R$

We used a simple method to calculate the effective reproduction number  $R_t$  with real-time data (Contreras, Sebastián et al. 2020). Briefly, in a standard *SEIR* model, four compartments were established, namely, susceptible individuals  $S$ , exposed individuals  $E$ , infected individuals  $I$ , and removed individuals  $R$ . The dynamics of the *SEIR* model are thus represented by:

$$\dot{S} = -\frac{\beta(t)SI}{N} \quad (1)$$

$$\dot{E} = \frac{\beta(t)SI}{N} - \sigma(t)E \quad (2)$$

$$\dot{I} = \sigma(t)E - \gamma(t)I - \mu(t)I \quad (3)$$

$$\dot{R} = \gamma(t)I \quad (4).$$

We combined equations (1), (2) and (3), and attained:

$$\frac{dI}{dS} + \frac{dE}{dS} = -1 + \frac{(\gamma(t)+\mu(t))N}{\beta(t)S} \quad (5).$$

We thus calculate  $R_t$  as the ratio between the time-dependent infection, and combination of recovery rates and case fatality rate,  $\beta(t)$ ,  $\gamma(t) + \mu(t)$  (Chen, Yi-Cheng et al. 2020), multiplied by the proportion of susceptible individuals in the whole population ( $\frac{S}{N}$ ):

$$R_t(t) = \frac{\beta(t)N}{(\gamma(t) + \mu(t))S} \quad (6),$$

and equation (5) can be re-written as:

$$\frac{dI}{dS} + \frac{dE}{dS} = -1 + \frac{1}{R_t(t)} \quad (7).$$

Equation (7) can be discretized in an interval  $[t_{i-1}, t_i]$  where we can assume that  $R_t(t) = R_t(t_i)$  is constant:

$$R_t(t_i) = \frac{1}{\frac{\Delta_i I}{\Delta_i S} + \frac{\Delta_i E}{\Delta_i S} + 1} \quad (8).$$

Extending the classic SEIR model to consider also exposed individuals and deaths, a disease states balance dictates the discrete differences to follow  $\Delta_i S + \Delta_i E + \Delta_i I + \Delta_i R + \Delta_i D = 0$ . Then, Equation (8) takes the final form applying the chain rule:

$$R_t(t_i) = \frac{1}{1 - \frac{\Delta_i E + \Delta_i I}{\Delta_i E + \Delta_i I + \Delta_i R + \Delta_i D}} \Leftrightarrow R_t(t_i) = \frac{\Delta_i E + \Delta_i I}{\Delta_i R + \Delta_i D} + 1 \quad (9).$$

Based on equation (9), we estimated  $R_t$  with real-time data by model simulations.

Table S1. Parameters used in the quarantine simulation of Wuhan epidemic

Parameter	Symbol	Time	Baseline value	Estimated value	Reference
Contact rate	$c$	15-26 Jan	10	-	Assumption
		27 Jan-9 Feb	-	6	Estimated
		10 Feb-6 Mar	-	3	Estimated
Mean duration from onset to hospital admission	$i$	15-26 Jan	10 days (9-12)	-	(Li, Guan et al. 2020)
		27 Jan-9 Feb	-	10 days (9-12)	Estimated
		10 Feb-5 Mar	-	3.2 days (3-3.8)	Estimated
Mean Recovery rate	$\nu$	-	0.0237 (0.02-0.041)	-	the Health Commission of Wuhan
Mean interval of incubation	$\rho$	-	4 days (3.8-7)	-	(Special Expert Group for Control of the Epidemic of Novel Coronavirus Pneumonia of the Chinese Preventive Medicine 2020)
Mean case fatality rate	$m$	-	0.0324 (0.031-0.044)	-	(Special Expert Group for Control of the Epidemic of Novel Coronavirus Pneumonia of the Chinese Preventive Medicine 2020)

#### 4. Sensitivity of mean interval incubation

Based on most recent studies on COVID-19, mean interval of incubation is around 3 to 4 days, but its range is large, and there is clinical case that the largest interval could reach 24 days. Therefore we analyzed the sensitivity of interval of incubation to the simulation (Figure S1). It shows that even mean interval of incubation is reduced to 2 days, the epidemic would be stopped although a higher medical burden will be imposed.

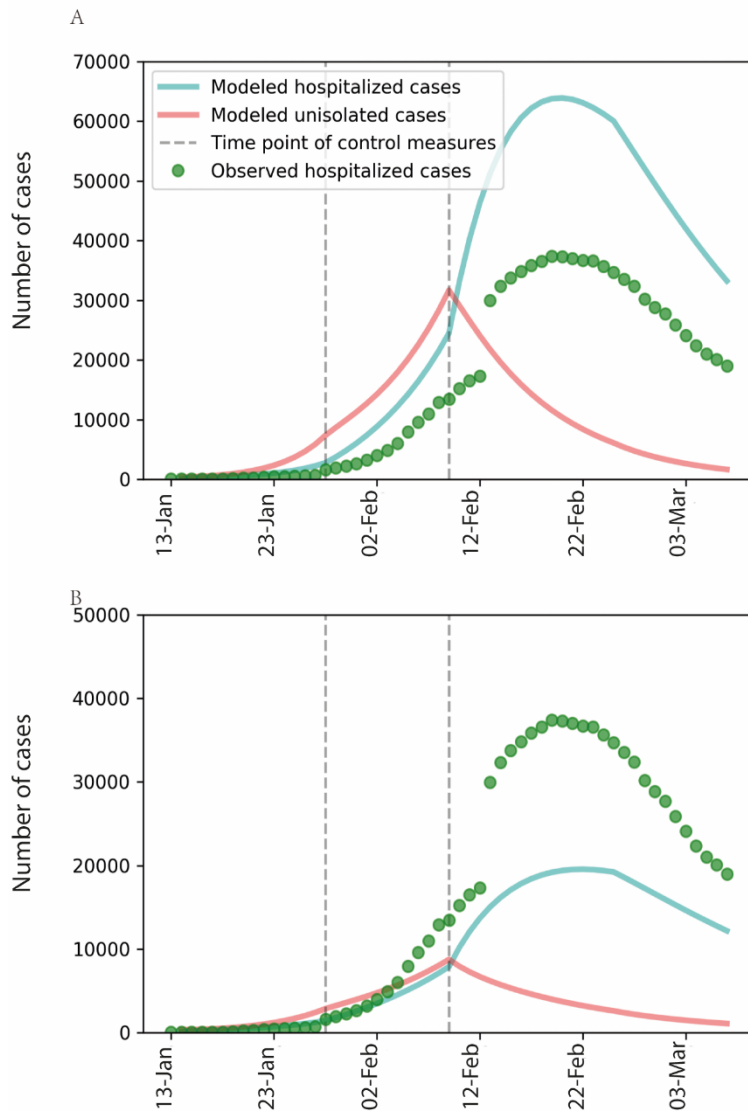


Figure S1. Sensitivity analysis of mean interval of incubation to the simulation of Wuhan epidemic. (A) Mean interval of incubation is assumed to 2 days. (B) Mean interval of incubation is assumed to be 6 days.



## References

- Obadia, T., et al. (2012). "The R0 package: a toolbox to estimate reproduction numbers for epidemic outbreaks." *BMC Med Inform Decis Mak*12:147.
- Wallinga, J., et al. (2004). "Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures." *Am J Epidemiol* 160 (6):509-516.
- Fine, PE. (2003). "The interval between successive cases of an infectious disease." *Am J Epidemiol*158(11):1039-47.
- Lipsitch, M., et al. (2003). "Transmission dynamics and control of severe acute respiratory syndrome." *Science* 300(5627): 1966-1970.
- Contreras, S., et al. (2020). "Real-Time Estimation of Rt for Supporting Public-Health Policies Against COVID-19." *Front Public Health* 8:556689
- Chen, YC., et al. (2020). "A Time-Dependent SIR Model for COVID-19 With Undetectable Infected Persons." *Transactions on Network Science and Engineering*7(4): 3279-3294
- Li, Q., et al. (2020). "Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia." *N Engl J Med*.
- Lipsitch, M., et al. (2003). "Transmission dynamics and control of severe acute respiratory syndrome." *Science* 300(5627): 1966-1970.
- Special Expert Group for Control of the Epidemic of Novel Coronavirus Pneumonia of the Chinese Preventive Medicine, A. (2020). "[An update on the epidemiological characteristics of novel coronavirus pneumoniaCOVID-19]." *Zhonghua Liu Xing Bing Xue Za Zhi* 41(2): 139-144.



# **Chapter 11**

## **Summary and Discussion**



Epidemiological research serves as a key pillar in understanding how pathogens transmit and their transmission consequences in the human society. The susceptibility and vulnerability of certain populations to particular viral infections are determined by many elements, which can mutually interact as well. It is clear that different types of pathogens are widely co-circulating among human population. Some of them may co-infect a same individual that exacerbates the clinical outcome. In this thesis, I attempted to understand the epidemiological and clinical features of viral infections, with focus on HPV, HEV and SARS-CoV-2, from several perspectives.

### **HPV prevalence in Chinese women and strategies to mitigate the burden**

For HPV, it is believed that most women in China acquire the infection once or twice during the lifetime<sup>[1]</sup>. Although HPV infection is usually self-limiting, failure of viral clearance with resultant persistent infection can lead to progression of precancerous lesions in the deep cell layers. According to its oncogenic potential, HPV is classified as either being low-risk HPV (LR) genotype or high-risk (HR) HPV genotype<sup>[2]</sup>. Considering that a significant proportion of persistent infections are difficult to diagnose, and thus eventually accelerate the progression of existing lesions to full blown cancer, in **Chapter 2**, I investigated the prevalence and genotype distribution of HPV in women from Inner Mongolia, China. I analyzed cytopathology status of the collected cell samples, and mapped its association between age and the HPV genotype distribution. I found that genotypes 16, 58, and 52 are most prevalent in that geographical region, and importantly, I observed a potential ethnic disparity of HPV infection between Han and Mongolian populations. In Mongolian women, the prevalence of HPV is significantly higher, and HPV31 and multiple-genotype infection are more common in comparison to Han women. These findings are partially consistent with a previous study reporting that HPV31 is the second most prevalent genotype in Mongolian women living in Ulaanbaatar, the capital of outer-Mongolia. Interestingly, the DRB\*1501 allele of the HLR class II gene has been reported to be associated with high susceptibility to HPV infection in the Han population of Inner Mongolia<sup>[3,4]</sup>, but strikingly has also been related to protection from cervical cancer in the Han of Xinjiang. As Xinjiang, similar to Inner Mongolia, is an important autonomous region but mainly composes Uygur, Kazak, and Hui ethnic groups, it is tempting to postulate that frequent cultural and genetic recombination or exchange may essentially affect the susceptibility of a population to viral infection. I thus continued HPV research in **Chapter 3**, but now focusing on the co-infection with *C. trachomatis*, another sexually transmitted pathogen in Chinese women. I found that the rate of HPV and *C. trachomatis* mono-infection and co-infection were 36.0%, 14.3% and 4.8%, respectively. Co-infection dramatically increases the risk for abnormal cytology (OR = 11.6; 95% CI, 7.29–18.6), and may further result in severe cervical pathogenesis and even carcinogenesis. The genotype HPV66 is surprisingly more prevalent than other high-risk genotypes in co-infection cases. I speculated a synergistic pathogenic effect by the co-infection, as *C. trachomatis* mono-

infection is usually asymptomatic. Co-infection may disrupt the immune system of hosts and lead to local inflammation<sup>[5,6]</sup>, which subsequently triggers the production of reactive oxygen species (ROS) and oxidative DNA damage<sup>[7]</sup>, and accelerates the progression of HPV invasion and related carcinogenesis<sup>[8]</sup>. A recent study has suggested that *C. trachomatis* may interact with several cancer-associated proteins that play a role in carcinogenesis<sup>[9]</sup>. For instance, in cultured cells *C. trachomatis* infection can lead to multinucleation that associates with chromosome instability and further promotes carcinogenesis<sup>[10]</sup>. A similar study shows that *C. trachomatis* plays a role in modifying PP2A signaling to suppress Ataxia-telangiectasia mutated activation which is responsible for high-fidelity repair of DNA double-strand breaks<sup>[11]</sup>. *C. trachomatis* may also compete with other DNA bind proteins, such as CpG binding proteins, chromodomain helicase DNA protein 5, and DDB2 protein, which involve regulations of cancer initiation and growth, and thereby alter normal intracellular environment consequently increasing the risk of cervical cancer. These evidences suggest that *C. trachomatis* may play a direct role in cervical carcinogenesis and this effect should be further explored in the setting of co-infection with HPV.

Mathematical modeling is a powerful tool to understand virus transmission and to estimate the burden of caused diseases in large populations. In **Chapter 4**, I used a SEIR compartmental model to estimate the incidence of HPV-related cervical cancer in China on the basis of the current Chinese cancer screening policy, and predicted its future trends toward 2030. In the modeling, I considered the role of aging in the occurrence of highly progressed lesions, and divided the modeled population by gender as to allow a heterosexual HPV transmission between females and males. I found that the rates of cervical cancer would gradually decrease from 2018 (14.1 cases per 100,000 women) to 2025 (13.4 cases per 100,000 women) by the current screening project. Increasing the coverage of cytological testing and introducing primary HPV screening would further prevent 223 new cases and 276 new cases (by screening 10 million women) on average per year, respectively, between 2017 and 2021, and 346 new cases and 425 new cases per year between 2022 and 2025. The 5-year-interval repeated screenings from 2017 till 2030 would substantially reduce the rate of new cases by 27.9%. I observed that the benefit of vaccination, when targeted on both girls and boys aged 12 years from 2017 to 2030, is more apparent in age groups younger than 34 years, and it assumes that maximized effect of vaccination would take time to cover the whole populations with a wider age range. This analysis represents a substantial expansion of the efforts of previous chapters to quantify the HPV-caused cervical cancer disease burden, and a better response to the vision of the "Healthy China 2030" issued by Chinese central government and the outlook of WHO to globally eliminate cervical cancer by achieving an incidence of lower than 4 per 100,000 women by 2100<sup>[12,13]</sup>. In my modeling, HPV vaccination has been shown to be effective for cervical cancer control and prevention. Although HPV vaccines are currently available in China,

the implementation remains at its infancy. Thus, a key issue emerges for the regional policy makers of public health as how to increase the accessibility and affordability of HPV vaccine in China and other resource-limited regions. It is also notable that the acceptance of HPV vaccine, relevant to education background and cervical cancer awareness, further influences the wide application of HPV vaccination. For instance, a recent meta-analysis shows that the estimated awareness rate of HPV vaccine among Chinese college students was 40.27%, which is relatively lower compared with European countries<sup>[14]</sup>. Another questionnaire in western China suggests that most female citizens (about 71.1%) lack basic knowledge about HPV, although at least half are willing to take the HPV vaccine<sup>[15]</sup>.

### **HEV zoonotic transmission and HEV infection in Chinese pregnant women**

Unlike HPV, HEV seldom transmits via a human-to-human route. There are various transmission routes, such as via contaminated water, contaminated food, contact with infected animals or blood transfusion, but this is associated with specific genotypes of HEV. Swine is recognized as the main reservoir of zoonotic HEV genotype 3 and 4. Consumption of pork-related food products contaminated with genotype 3 HEV is thought to be the main source of HEV infection in many European countries. In this thesis, I focused on the transmission and epidemiological features of zoonotic HEV strains. In **Chapter 5**, I estimated HEV prevalence in domestic pigs and wild boars at a global scale by conducting a systematic review and meta-analysis. I found that nearly 60% of domestic pigs and 27% of wild boars are HEV seropositive. Importantly, about 13% of domestic pigs and 9.5% of wild pigs are HEV RNA positive, indicating active infection. Of note, 10% of commercial pork products are HEV RNA positive. It has been reported that blood of HEV-viremic pigs from slaughterhouses may contaminate a whole pork supply chain<sup>[16]</sup>. Genotype 3 is predominately circulating in pigs, and the prevalence of HEV RNA in pigs and retail pork livers is similar among Canada, the USA and Europe<sup>[17]</sup>. HEV incidence and prevalence in human population varies geographically. From 1990 to 2017, the number of HEV cases was increased in low and low-middle socio-demographic index regions, but also in some developed areas such as Oceania and western Europe<sup>[18]</sup>. Importantly, in Western Europe the growth of HEV cases mainly concentrates on populations above 40 years of age, which implies the possible involvement of specific HEV genotypes and elements such as environmental, life-style, socioeconomic factors.

In **Chapter 6**, I focused on the transmission of zoonotic HEV in representative European countries. I modeled how individuals acquire HEV infection by consuming contaminated pork-derived food. I found that foodborne infection is associated with the levels of active HEV in contaminated food and the habit of pork food consumption. It is estimated that orally ingested amount of HEV at which the probability of infection equals 50% is  $8.1 \times 10^6$  (95% CI  $2.4 \times 10^6$ - $2.0 \times 10^7$ ) viral genomes. This level is lower than those tested in swine samples

collected in or near farming areas, and this implies that workers of pig farms or individuals living nearby may have a higher risk of acquiring HEV infection. The first HEV vaccine named Hecolin was approved in China in 2012<sup>[19]</sup>. A large clinical trial that included over 100,000 healthy participants demonstrated a 95.5% efficacy over 19 months<sup>[20]</sup>. Long-term follow up of a 4.5-year period showed a vaccine efficacy of 86.8%, and 87% of those who received three doses of the vaccine maintained anti-HEV antibodies<sup>[21]</sup>. In the real-world, it is not feasible to widely implement this vaccine in the general population. Therefore, I made an assumption that specifically in high-risk population of foodborne HEV infection in Germany can be targeted for vaccination, and found under these conditions a clear beneficial effect of reducing the infection by my modeling. Since there is no HEV vaccine available for pigs, I theoretically assumed two vaccination strategies targeting at reduction of the susceptibility of uninfected animals, and reduction of the contagiousness of infectious animals, respectively. I illustrated the effectiveness of applying vaccine in pigs on the risk of HEV transmission to humans in Germany with the mathematical models. Considering that modern pig industry may largely shape the landscape of HEV epidemiology in human population in Western countries, it deserves more attention to evaluate active interventions including vaccination and to examine the safety in the food chain. A study from Switzerland has provided proof-of-concept that implementing public health measures to control pork-related food products can reduce the risk of HEV transmission to humans<sup>[22]</sup>.

In China, genotype 1 and 2 HEV is dominating<sup>[23,24]</sup>, which is thought to be more pathogenic than genotype 3. In **Chapter 7**, I screened both viral RNA and antibodies of HEV in a Chinese pregnant women population. The seroprevalence of HEV (6.62%) in pregnant population is slightly higher than that of the non-pregnant control cohort (4.5%), but whether pregnant women are more susceptible to HEV infection remains to be further clarified. Importantly, pregnant women with recent/ongoing HEV infection have slightly higher ALT level and appear to be associated with adverse maternal and neonatal outcomes. However, the number of positive cases is too small to draw firm conclusions. Severe maternal and neonatal outcomes usually occur in pregnant women infected with genotype 1 or 2 HEV from developing countries, but not from industrialized countries with genotype 3 prevalence. Thus, the burden of HEV infection in Chinese pregnant women may be distinct from resource-limited regions and western countries. Relatively mild symptoms in HEV-infected pregnant women have been observed in similar studies. A study in Shanghai, China showed that most of the hospitalized HEV-infected pregnant women (85.2%, 127/149) present with no obvious clinical symptoms and the disease outcomes were generally benign with no liver failure or maternal mortality<sup>[25]</sup>. Slightly elevated ALT levels were also detected in most of the anti-HEV sero-positive women in Ghana<sup>[26]</sup>. However, the neonatal outcomes of maternal infection might be adverse. Preterm birth, followed by premature rupture of membranes, fetal distress, low birth weight,



threatened preterm labor, neonatal asphyxia, stillbirth and threatened abortion have been observed in around 43% infants with maternal infection<sup>[25]</sup>. Therefore, future studies are required to delineate the horizontal epidemic of pregnant HEV infection in Inner Mongolia of China.

### **Modeling the spread and control of COVID-19**

During my PhD, the COVID-19 pandemic caused by SARS-CoV-2 infection emerged from Wuhan, China and subsequently spread over the globe. Different countries and regions often react differently in response to outbreaks, whereas the right early response is essential for containing the outbreak, thereby avoiding large epidemic or pandemic. Because of their disparities in culture, socioeconomic status, and types of government, the implementation of control measures can vary tremendously among different countries. Further, the disease burden caused by SARS-CoV-2 varies tremendous among different populations. Thus, in **Chapter 9**, I first analyzed the fatality rate when comparing the epicenter Wuhan to other regions of mainland China at the early stage of the pandemic. As recorded by the Chinese Center for Disease Control and Prevention, by Feb 16, 2020, there had been 70641 confirmed cases and 1772 deaths due to COVID-19, with an average mortality of about 2.5%<sup>[27]</sup>. However, in-depth analysis of these data shows clear disparities in mortality rates between Wuhan (>3%), different regions of Hubei province (about 2.9% on average), and across the other provinces of China (about 0.7% on average). I postulated this is likely to be related to the rapid escalation in the number of infections in the epicenter, which resulted in insufficiency of health-care resources, thereby negatively affecting patient outcomes, while this was not the situation for the other parts of China. In assuming that average levels of health care are similar throughout China, higher numbers of infections in a given population can be considered an indirect indicator of a heavier health-care burden. Plotting mortality against the incidence of COVID-19 (cumulative number of confirmed cases since the start of the outbreak, per 10000 population) showed a significant positive correlation, suggesting that mortality is correlated with health-care burden. Hospital capacity has been being one of main concerns in tackling with the spread of COVID-19, and events that disease outbreak in the peak period will overwhelm hospital capacity has been repeatedly reported. A modeling study further estimated the impact of inadequacy of critical care when regional disease breaks up and hold that the number of ICU beds needs to double if no effective social distancing policies such as self-isolation, school closure or limiting public transportation were taken<sup>[28]</sup>.

However, Wenzhou, a city of China, seemed to set an example of demonstrating adequate management of health care resources leading to rapid disease control. In early 2020, the COVID-19 outbreak coincided with a massive population migration, and led to the rapid disease spread across China. Wenzhou quickly reached the highest number of COVID-19 cases

for any city outside of Hubei province, and most of these cases were imported from the epicenter (with a total of 504 imported cases). However, it took only about 46 days for this city to fully contain the epidemic, preventing the city to grow into a new epicenter. In **Chapter 11**, I thus modeled how epidemics was rapidly contained in Wenzhou. I found that specific measures, including immediate hospital admission, disease quarantine and isolation, and lock down of most districts, greatly decreased the exposure of the susceptible to the disease. These results highlight the essential of early, quick and adequate response with implementation of the right control measures. Delayed response will soon plunge the healthcare system into crisis, as seen in Wuhan and other countries like Italy, Spain, UK and USA, even in the context of well-developed healthcare systems. Availability of healthcare is critical for controlling the epidemic but also for minimizing severe patient outcomes<sup>[29]</sup>. The distinct experiences of Wenzhou and Wuhan are mirroring what has happened in many parts of the world, either as epicenters or regions primarily with imported case.

In **Chapter 10**, I modeled the transmission of SARS-CoV-2 from the epicenter to other Chinese provinces in early 2020. This process I mathematically interpreted using the metapopulation transmission theory, namely, an initial infectious population (here I referred to Wuhan) can transmit virus to more susceptible populations (here I referred to other provinces of China). A recent modeling study also uses the metapopulation model to estimate the course of the epidemic in America, in which the simulation was performed at a county-scale and estimate critical, time-varying epidemiological properties underpinning the dynamics of SARS-CoV-2<sup>[30]</sup>. In the modeling, I found that two events were essential to block viral transmission within population migration. One is the immediate lock-down of the city of Wuhan, and another is intense quarantine of suspected case importers. I also noticed that disease transmission was closely associated with the size and speed of the population migrations among cities, and transportation hubs and trade ports considered in this study are still key objectives in disease controls today. Populated neighboring provinces having convenient transportation links with the epidemic regions should be given adequate concerns<sup>[31]</sup>, and potential infection threats in destination and directions of population migrations from inside and outside of nation should be pre-elucidated. China is among the few countries that still has locked down its boundaries, by pressure that the disease will rear its ugly head always exists as the country maintains the largest susceptible pool globally, although nearly the entire population has received at least two-doses of vaccine. While small scale epidemics caused by imported cases occasionally occur, the Chinese government is fully conscious of the dynamics of the virus spread and has learnt to rapidly isolate the epidemic area to enable other areas to perform normal economic activities and continue the routine daily life.

In summary, this thesis has provided new insight in better understanding viral infectious disease with focus on specific populations and vulnerable settings. I achieved this by

prioritizing research on three representative pathogenic viruses, namely, HPV, HEV and SARS-CoV-2. These viruses have distinct transmission routes and affect different human populations. By taking multidimensional approaches, including epidemiological methodologies and mathematical modelling, this thesis has provided comprehensive perspectives in understanding virus transmission, disease burden, and preventive measures. I hope these knowledges would help to combating the current threats by viral diseases and better preparing future epidemic/pandemic.

## References

1. Yang L, Huangpu XM, Zhang SW, Lu FZ, Sun XD, Sun J, Mu R, Li LD, Qiao YL. [Changes of mortality rate for cervical cancer during 1970's and 1990's periods in China]. *Zhongguo Yi Xue Ke Xue Yuan Xue Bao*. 2003 Aug;25(4):386-90. Chinese. PMID: 12974079.
2. Muñoz N, Bosch FX, de Sanjosé S, Herrero R, Castellsagué X, Shah KV, Snijders PJ, Meijer CJ; International Agency for Research on Cancer Multicenter Cervical Cancer Study Group. Epidemiologic classification of human papillomavirus types associated with cervical cancer. *N Engl J Med*. 2003 Feb 6;348(6):518-27. doi: 10.1056/NEJMoa021641. PMID: 12571259.
3. Xie J, Wang L, Wan G, Ma X, Qiao H. Study on the association between HPV16-positive cervical cancer and polymorphism of HLA-DRB1, HLA-DQB1 in Inner Mongolian women. *J Report Med*. 2007; 16: 91–95.
4. Sun Q, Qi C, Yang A, et al. A study on the correlations of human leucocyte antigens-DRB1\* 1501, DQB1\* 0602 alleles with HPV16 infection and invasive squamous cell carcinoma of cervix in Xinjiang Uigur and Han women. *J Shihezi Univ*. 2009; 27: 133-138.
5. Brunham RC, Rey-Ladino J. Immunology of Chlamydia infection: implications for a Chlamydia trachomatis vaccine. *Nat Rev Immunol*. 2005 Feb;5(2):149-61. doi: 10.1038/nri1551. PMID: 15688042.
6. Paavonen J. Chlamydia trachomatis infections of the female genital tract: state of the art. *Ann Med*. 2012 Feb;44(1):18-28. doi: 10.3109/07853890.2010.546365. Epub 2011 Feb 1. PMID: 21284529.
7. Yang X, Siddique A, Khan AA, Wang Q, Malik A, Jan AT, Rudayni HA, Chaudhary AA, Khan S. Chlamydia Trachomatis Infection: Their potential implication in the Etiology of Cervical Cancer. *J Cancer*. 2021 Jun 11;12(16):4891-4900. doi: 10.7150/jca.58582. PMID: 34234859; PMCID: PMC8247366.
8. IARC Working Group on the Evaluation of Carcinogenic Risk to Humans. Biological Agents. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, No. 100B. Lyon, France: International Agency for Research on Cancer, 2012.
9. Khan AA, A Abuderman A, Ashraf MT, Khan Z. Protein-protein interactions of HPV-Chlamydia trachomatis-human and their potential in cervical cancer. *Future Microbiol*. 2020 May; 15: 509-520. doi: 10.2217/fmb-2019-0242. Epub 2020 Jun 1. PMID: 32476479.
10. Brown HM, Knowlton AE, Grieshaber SS. Chlamydial infection induces host cytokinesis failure at abscission. *Cell Microbiol*. 2012 Oct;14(10):1554-67. doi: 10.1111/j.1462-5822.2012.01820.x. Epub 2012 Jun 19. PMID: 22646503; PMCID: PMC3443326.
11. Mi Y, Gurusurthy RK, Zadora PK, Meyer TF, Chumduri C. Chlamydia trachomatis Inhibits Homologous Recombination Repair of DNA Breaks by Interfering with PP2A Signaling. *mBio*. 2018 Nov 6;9(6): e01465-18. doi: 10.1128/mBio.01465-18. PMID: 30401777; PMCID: PMC6222135.
12. Xinhua News Agency. The CPC Central Committee and the State Council jointly released an outline for the "Healthy China 2030" initiative. [http://www.xinhuanet.com//politics/2016-10/25/c\\_1119785867.htm](http://www.xinhuanet.com//politics/2016-10/25/c_1119785867.htm), 2016-10-25.
13. Garland SM, Giuliano A, Brotherton J, Moscicki AB, Stanley M, Kaufmann AM, Bhatla N, Sankaranarayanan R, Palefsky JM, de Sanjose S; IPVS. IPVS statement moving towards elimination of cervical cancer as a public health problem. *Papillomavirus Res*. 2018 Jun; 5: 87-88. doi: 10.1016/j.pvr.2018.02.003. Epub 2018 Feb 27. PMID: 29499389; PMCID: PMC5887016.
14. Yin G, Zhang Y, Chen C, Ren H, Guo B, Zhang M. Have you ever heard of Human Papillomavirus (HPV) vaccine? The awareness of HPV vaccine for college students in China based on meta-analysis. *Hum Vaccin Immunother*. 2021 Aug 3;17(8):2736-2747. doi: 10.1080/21645515.2021.1899731. Epub 2021 Mar 31. PMID: 33787459; PMCID: PMC8475627.
15. He J, He L. Knowledge of HPV and acceptability of HPV vaccine among women in western China: a cross-sectional survey. *BMC Womens Health*. 2018 Jul 27;18(1):130. doi: 10.1186/s12905-018-0619-8. PMID: 30053844; PMCID: PMC6063014.
16. Sooryanarain H, Heffron CL, Hill DE, Fredericks J, Rosenthal BM, Werre SR, Opriessnig T, Meng XJ.

- Hepatitis E Virus in Pigs from Slaughterhouses, United States, 2017-2019. *Emerg Infect Dis.* 2020 Feb;26(2):354-357. doi: 10.3201/eid2602.191348. PMID: 31961315; PMCID: PMC6986846.
17. Wilhelm B, Fazil A, Rajić A, Houde A, McEwen SA. Risk Profile of Hepatitis E Virus from Pigs or Pork in Canada. *Transbound Emerg Dis.* 2017 Dec;64(6):1694-1708. doi: 10.1111/tbed.12582. Epub 2016 Oct 7. PMID: 27718330.
  18. Jing W, Liu J, Liu M. The global trends and regional differences in incidence of HEV infection from 1990 to 2017 and implications for HEV prevention. *Liver Int.* 2021 Jan;41(1):58-69. doi: 10.1111/liv.14686. PMID: 33025620.
  19. Park SB. Hepatitis E vaccine debuts. *Nature.* 2012 Nov 1;491(7422):21-2. doi: 10.1038/491021a. PMID: 23128204.
  20. Zhu FC, Zhang J, Zhang XF, Zhou C, Wang ZZ, Huang SJ, Wang H, Yang CL, Jiang HM, Cai JP, Wang YJ, Ai X, Hu YM, Tang Q, Yao X, Yan Q, Xian YL, Wu T, Li YM, Miao J, Ng MH, Shih JW, Xia NS. Efficacy and safety of a recombinant hepatitis E vaccine in healthy adults: a large-scale, randomised, double-blind placebo-controlled, phase 3 trial. *Lancet.* 2010 Sep 11;376(9744):895-902. doi: 10.1016/S0140-6736(10)61030-6. Epub 2010 Aug 20. PMID: 20728932.
  21. Long-term efficacy of a hepatitis E vaccine. *N Engl J Med.* 2015 Apr 9;372(15):1478. doi: 10.1056/NEJMx150008. Erratum for: *N Engl J Med.* 2015 Mar 5;372(10):914-22. PMID: 25853767.
  22. Ripellino, P.; Pianezzi, E.; Martinetti, G.; Zehnder, C.; Mathis, B.; Giannini, P.; Forrer, N.; Merlani, G.; Dalton, H.R.; Petrini, O.; et al. Control of Raw Pork Liver Sausage Production Can Reduce the Prevalence of HEV Infection. *Pathogens* 2021, 10, 107. <https://doi.org/10.3390/pathogens10020107>.
  23. Wang Y, Liu H, Liu S, Yang C, Jiang Y, Wang S, Liu A, Peppelenbosch MP, Kamar N, Pan Q, Zhao J. Incidence, predictors and prognosis of genotype 4 hepatitis E related liver failure: A tertiary nested case-control study. *Liver Int.* 2019 Dec;39(12):2291-2300. doi: 10.1111/liv.14221. Epub 2019 Sep 3. PMID: 31436371.
  24. Wang Y, Chen G, Pan Q, Zhao J. Chronic Hepatitis E in a Renal Transplant Recipient: The First Report of Genotype 4 Hepatitis E Virus Caused Chronic Infection in Organ Recipient. *Gastroenterology.* 2018 Mar;154(4):1199-1201. doi: 10.1053/j.gastro.2017.12.028. Epub 2018 Feb 9. PMID: 29432746.
  25. Zhang F, Wang J, Cheng J, Zhang X, He Q, Zhaochao L, Shu J, Yan L, Wang L, Wang L, Zhang J. Clinical features of sporadic hepatitis E virus infection in pregnant women in Shanghai, China. *J Infect.* 2022 Jan;84(1):64-70. doi: 10.1016/j.jinf.2021.11.004. Epub 2021 Nov 9. PMID: 34767838.
  26. Obiri-Yeboah D, Asante Awuku Y, Adu J, Pappoe F, Obboh E, Nsiah P, Amoako-Sakyi D, Simpore J. Sero-prevalence and risk factors for hepatitis E virus infection among pregnant women in the Cape Coast Metropolis, Ghana. *PLoS One.* 2018 Jan 25;13(1): e0191685. doi: 10.1371/journal.pone.0191685. PMID: 29370271; PMCID: PMC5784989.
  27. Chinese Center for Disease Control and Prevention. Distribution of new coronavirus pneumonia. <http://2019ncov.chinacdc.cn/2019-nCoV/> (accessed Feb 16, 2020).
  28. Moghadas SM, Shoukat A, Fitzpatrick MC, Wells CR, Sah P, Pandey A, Sachs JD, Wang Z, Meyers LA, Singer BH, Galvani AP. Projecting hospital utilization during the COVID-19 outbreaks in the United States. *Proc Natl Acad Sci U S A.* 2020 Apr 21;117(16):9122-9126. doi: 10.1073/pnas.2004064117. Epub 2020 Apr 3. PMID: 32245814; PMCID: PMC7183199.
  29. Ji Y, Ma Z, Peppelenbosch MP, Pan Q. Potential association between COVID-19 mortality and health-care resource availability. *Lancet Glob Health.* 2020;8(4): e480. Epub 2020/02/29. doi: S2214-109X (20)30068-1 [pii]10.1016/S2214-109X (20)30068-1. PubMed PMID: 32109372; PubMed Central PMCID: PMC7128131.
  30. Sen P, Yamana TK, Kandula S, Galanti M, Shaman J. Burden and characteristics of COVID-19 in the United States during 2020. *Nature.* 2021 Oct;598(7880):338-341. doi: 10.1038/s41586-021-03914-4. Epub 2021 Aug 26. PMID: 34438440.
  31. Hu Z, Wu Y, Su M, Xie L, Zhang A, Lin X, Nie Y. Population migration, spread of COVID-19, and epidemic prevention and control: empirical evidence from China. *BMC Public Health.* 2021 Mar 17;21(1):529. doi: 10.1186/s12889-021-10605-2. PMID: 33731053; PMCID: PMC7968569.



# **Chapter 12**

**Nederlandse Samenvatting**

**Dutch Summary**





Een virus is een minuscule stukje organische substantie dat zich uitsluitend kan vermenigvuldigen in de cellen van levende wezens. Wanneer een virus een levende cel binnendringt, zal deze gastheercel grote hoeveelheden kopieën van het oorspronkelijke virus gaan produceren. Virussen infecteren alle vormen van leven, van dieren, schimmels en planten tot micro-organismen en komen voor in alle ecosystemen. De vermenigvuldiging van virussen gaat gepaard met functieverlies en vaak zelfs het uiteen knappen van de geïnfecteerde cel en organismen beschermen zich dan ook tegen virussen met een grote variëteit van strategieën en mechanismen. Zowel de virusinfectie zelf als het antwoord van het lichaam op de infectie gaan gepaard met ziekte en zelfs sterfte. In dit proefschrift probeer ik te begrijpen welke processen de verspreiding van virussen in specifieke groepen van de bevolking bepalen. De keuzes die ik hierbij gemaakt heb, leg ik uit in **Hoofdstuk 1**.

In het eerste gedeelte van het proefschrift concentreer ik mij op Het humaan papillomavirus, ook wel HPV genoemd. Papillomavirussen kunnen abnormale celgroei van huid en slijmvliezen teweegbrengen en zijn de veroorzakers van wratten, terwijl seksueel overgedragen varianten als een belangrijke oorzaak van baarmoederhalskanker worden gezien (n 75% van de gevallen van baarmoederhalskanker wordt deze door bepaalde HPV-soorten veroorzaakt). Bij bestudering van de epidemiologie van de baarmoederhalskanker-geassocieerde vorm van HPV infectie wordt vooral geografie betrokken, maar in **hoofdstuk 2** van dit proefschrift laat ik zien dat in het zelfde gebied (Binnen-Mongolië) er zeer grote verschillen zijn in de prevalentie van het HPV virus tussen vrouwen uit verschillende etnisch-culturele groepen. Het lijkt er dan ook op dat de puur geografische benadering van de epidemiologie van HPV infectie bij vrouwen te simpel is. Ik heb deze resultaten wereldkundig gemaakt middels een publicatie in het zeer vooraanstaande wetenschappelijke vaktijdschrift *Journal of Medical Virology*. De Amerikaanse *Centers for Disease Control and Prevention* stellen dat het gebruik van condooms niet voldoende is om bescherming te bieden tegen HPV-gerelateerde baarmoederhalskanker, waar dit voor het beschermen tegen chlamydia besmetting wel voldoende zou zijn. In **hoofdstuk 3** echter laat ik zien dat er een sterke correlatie is tussen chlamydia besmetting en HPV besmetting, wat twijfel over deze stelling oproept. Ook deze resultaten wist ik te publiceren in het zeer vooraanstaande wetenschappelijke vaktijdschrift *Journal of Infectious Disease*. Een laatste HPV-gerelateerde studie presenteer ik in **hoofdstuk 4**. Hier onderzoek ik de effectiviteit van het bevolkingsonderzoek baarmoederhalskanker in China. In het verleden

waren er twijfels over de effectiviteit van dit programma, met name wat betreft afgelegen, minder ontwikkelde, gebieden in dit land. Ik presenteer echter bewijs dat dit programma effectief is. Deze studie heb ik onlangs aangeboden voor wetenschappelijke publicatie en hoop ik dus op die wijze binnenkort wereldkundig te maken.

In het tweede gedeelte van mijn dissertatie focus ik mij op het Hepatitis E virus (HEV), de belangrijkste oorzaak van acute virale leverontsteking ter wereld. In ontwikkelingslanden wordt dit virus vooral verspreid via de oraal-fecale route, met name via het water. In Westerse landen lijkt besmetting uit dierlijke reservoirs het belangrijkste te zijn. Van oudsher wordt gesuggereerd dat besmetting van mensen door het consumeren van varkensproducten hierbij een belangrijke rol speelt. In **hoofdstuk 5** probeer ik een inschatting te maken over hoe vaak varkensvlees nu eigenlijk een bron kan zijn voor Hepatitis E infectie op een wereldwijde schaal. De resultaten van deze analyse publiceerde ik in het vooraanstaande wetenschappelijke vaktijdschrift *One Health*. De conclusie was dat slecht gebakken varkensvlees zeker kan bijdragen aan de HEV infectiedruk, maar *an sich* niet voldoende is om verspreidingspatronen van deze ziekte volledig te verklaren en dat andere factoren belangrijk hierin ook geïdentificeerd moeten worden. In **hoofdstuk 6**, ook gepubliceerd in het vooraanstaande wetenschappelijke vaktijdschrift *One Health* werk ik dan uit hoe middels vaccinatie dit probleem het hoofd kan worden geboden. In **hoofdstuk 7** laat ik zien dat HEV infectie inderdaad tot (iets) slechtere zwangerschapsresultaten leidt, een resultaat wat ik publiceerde in het wetenschappelijk vaktijdschrift *Clinics and Research in Hepatology and Gastroenterology*. Alles bijelkaar leveren deze studies inzicht op hoe om te gaan met HEV infecties, met name in de ontwikkelde wereld.

Gedurende dit promotieonderzoek werd de wereld getroffen door COVID-19. Ik besloot de resterende tijd van mijn studies in Rotterdam te gebruiken aan begrijpen van de epidemiologie van COVID-19). Ik was een van de eersten die de gevolgen van een door COVID-19 overbelast medisch zorgsysteem benoemde en kwantificeerde, een analyse die in beknopte vorm verscheen in het wetenschappelijk toptijdschrift *Lancet Global Health* (**hoofdstuk 8**) en in meer uitgewerkte versie in wetenschappelijke vaktijdschrift *Transboundary and Emerging Diseases* (**hoofdstuk 9**), terwijl ik ook in ga op de relatieve effectiviteit van strategieën gericht op het beheersen van de epidemie (**hoofdstuk 10**), een

studie die verscheen in het wetenschappelijke toptijdschrift *PloS Global Health*. Een samenvattende discussie, waarin ik mijn bevindingen plaats in het corpus van de contemporaine biomedische literatuur is te vinden in **hoofdstuk 11**.

Alles beschouwend hoop ik met deze dissertatie te hebben bijgedragen aan het begrip van hoe virusziekten zich in de menselijke populatie verspreiden, wat de gevolgen daarvan zijn in relatie tot de capaciteit van gezondheidszorgsystemen, en wat de effectiviteit is van de verschillende mogelijke maatregelen (vaccinatie van de populatie of mogelijke zoönotische reservoirs, verminderen sociale contacten *etcetera*). Dit moet helpen toekomstige virale bedreigingen in toom te houden.



# **Appendix**

**Acknowledgements**

**Publications**

**PhD Portfolio**

**Curriculum Vitae**



## Acknowledgments

I started my PhD program at Erasmus-MC University Rotterdam in 2017, and my research interest and enthusiasm are well rewarded by this program. Both the research atmosphere at the Department of Gastroenterology and Hepatology and the kind and vigorous researchers I encountered there will provide good memories for the rest of my career. I thereby would send my greatest gratitude to all of you, my promotor, co-promoter, colleagues, my families, friends, and the Inner Mongolia Maternal and Child Care Hospital.

First of all, I would like to thank my co-promotor and daily supervisor Dr. Qiuwei (Abdullah) Pan. Thank you for your enormous encouragement and support during my PhD study. Thank you for your help and suggestions when I have difficult in life and in study. You are so gentle and passionate that I have learnt a lot from you on how to be a qualified researcher and a good individual. I still remember your well-intentioned criticism on each of my immature research proposal and patient revise on each tiny writing error in my manuscript. And I still remember your encouragement and help on my research of infectious disease modeling. To be honest, I am still quite young in this field. Thank you so much for supervising me.

To my promotor Prof. dr. Maikel P. Peppelenbosch, thank you so much for guiding me and for your enormous support at the routine group meeting and weekly MDL meeting. I really enjoy the group discussion that you attended, because you always come up with new inspiring ideas and help to solve our problems. I am so impressed by your creative thoughts and multi-disciplinary angle. I still remembered your great interest on my modeling job and am impressed by your affluent mathematical knowledge. You also show great kindness on cultivating us with critical thinking, which is essential to our future careers.

To Mr. Jacobus Hagoort, thank you for sharing the writing skills in the academic writing course of ten lessons. It is your warm words and patience that decrease my inner tense when I firstly attend a course with lots of class discussions. To all classmates in the writing course, thank you for your kindness that helps me learn to enjoy the communications in a multi-culture environment where different points of view are always welcome.

To Dr. Karl Brand, thank you for providing affluent contents and instances in R programing of gene expression. Your expertise enlarges my insight and helps me design more precise and efficient models. To Dr. Leticia G. Leon, thank you for sharing your knowledge on python programing. You teach me how to skillfully cope with the limits of computer languages. The study experiences gained from you are quite precious. To Mr. Carlo van der Zon, thank you for sharing your expertise in adobe professional software uses. I learn a lot of practical skills from you such as app production and database design.

To Dr. Changbo Qu, Dr. Sunrui Chen, Dr. Jiaye Liu, Shihao Ding, Dr. Wanlu Cao, Dr. Buyun Ma, Dr. Zhijiang Miao, Dr. Peifa Yu, Yang Li, Yunlong Li, Ling Wang, Pengfei Li, Yining Wang and Bingting Yu, my respectful lab fellows, thanks for sharing me many skills and experiences on Rotterdam study. We spend most of our time in the biweekly group meeting, MDL meeting as well as Molmed class. I am so glad to work with you and wish you all and your family healthy and happy.

To Wenshi, you are a brilliant young scientist and kind person. Thank you for your enormous helps on my living in Rotterdam. Your attitudes to research and life deeply impress me. To Junhong, my elder fellow, thank you for helping me adapt to the new research life. I still remember the scenes that we discuss many interesting things with a cup of Cappuccino in hand.

To Zhen Ping, my high school classmate, it is amazing that we meet again in Erasmus MC and happy to know your excellent performances here. I wish you and your family healthy and happy.

To my roommates in Na1013, Zhouhong Ge, Shan Li, Hank, Sophia, and other colleagues in this office. I am very happy for sharing the office with all of you. The chatting and discussion with you on Rotterdam livings and research everyday are a precious memory for me.

To all MDL colleagues: Raymond, Leonie, Naomi, Jan, Auke, Henk, Sonja, Marla, Andre, Hugo, Jaap, Lucia, Natasha, Kelly, Patrick, Anthonie, Marcel, Monique, Shanta, Petra, Lisanne, and everyone in our lab. We are a big family; I want to thank all of you for your help and support during my study in the MDL lab.

To Prof. Zhongren Ma, Prof. Hua Liu, Dr. Kan Chen, Dr Xiaoxia Ma, Dr. Jianhua Zhou, Yong Ye, Kai Zhang, thanks for your support and help for our research collaboration in Lanzhou.

To my colleagues in Inner Mongolia Maternal and Child Care Hospital: Zhifen Wang, Yueqi Jia, Guixiang Wang, Xiaoping Ji, Xiaohua Wang, Hong Dong, Bo Zhu, Yan Zhou, Xueyuan Zhou, Jie Wang, Dongxia Hou, thank you for your encouragement and support for my study abroad.

To my parents, thank you for everything you have done for me. They are presents of a life that I always remember and am unable to repay.

To my wife, Yang Bai, thank you for your support on my study. All the best wishes to our bright future. To my little daughter, 语(Yu), I hope you to be a happy person always.



## Publication list

1. Li P<sup>#</sup>, Ji Y<sup>#</sup>, Li Y, Ma Z, Pan Q. Estimating the global prevalence of hepatitis E virus in swine and pork products. *One Health*. 2021 Dec 14; 14: 100362. (#: contribute equally)
2. Ji Y<sup>#</sup>, Li P<sup>#</sup>, Jia Y, Wang X, Zheng Q, Peppelenbosch MP, Ma Z, Pan Q. Estimating the burden and modeling mitigation strategies of pork-related hepatitis E virus foodborne transmission in representative European countries. *One Health*. 2021 Nov 18; 13: 100350. (#: contribute equally)
3. Ji Y, Li P, Zheng Q, Ma Z, Pan Q. Distinct effectiveness in containing COVID-19 epidemic: comparative analysis of two cities in China by mathematical modeling. *PLOS Global Public Health*. 2021.1(11): e0000043.
4. Ma X<sup>#</sup>, Ji Y<sup>#</sup>, Jin L, Baloch Z, Zhang D, Wang Y, Pan Q, Ma Z. Prevalence and clinical features of hepatitis E virus infection in pregnant women: A large cohort study in Inner Mongolia, China. *Clinics and Research in Hepatology and Gastroenterology*, 2021 Jul;45(4):101536. (#: contribute equally)
5. Baloch Z, Ma Z, Ji Y, Ghanbari M, Pan Q, Aljabr W. Unique challenges to control the spread of COVID-19 in the Middle East. *Journal of Infection and Public Health*, 2020, 13(9):1247-1250.
6. Ji Y, Ma Z, Peppelenbosch M.P, Pan Q. Potential association between COVID-19 mortality and health-care resource availability. *Lancet Global Health*, 2020 Apr;8(4): e480. doi: 10.1016/S2214-109X (20)30068-1.
7. Ji Y, Ma X, Li Z, Peppelenbosch M.P, Ma Z, Pan Q. The Burden of Human Papillomavirus and Chlamydia trachomatis Coinfection in Women: A Large Cohort Study in Inner Mongolia, China. *The Journal of Infectious Diseases*, 2018. doi:10.1093/ infdis/jiy497.
8. Ji Y, Zhao C, Ma X, Peppelenbosch M.P, Ma Z, Pan Q. Outcome of a screening program for the prevention of neonatal early-onset group B Streptococcus infection: a population-based cohort study in Inner Mongolia, China, *Journal of Medical Microbiology*, 2019 May;68(5):803-811. doi: 10.1099/jmm.0.000976.
9. Ji Y, Ikram A, Ma Z, P. Peppelenbosch M.P, Pan Q. Co-occurrence of heterozygous mutations in COL1A1 and SERPINF1 in a high-risk pregnancy complicated by osteogenesis imperfect, *Journal of Genetics*. 2019 Jun;98(2). pii: 51.
10. Wang X<sup>#</sup>, Ji Y<sup>#</sup>, Li J, Dong H, Zhu B, Zhou Y, Wang J, Zhou X, Wang Y, Peppelenbosch M.P, Pan Q, Ji X, Liu D. Prevalence of human papillomavirus infection in women in the Autonomous Region of Inner Mongolia: A population-based study of a Chinese ethnic minority. *Journal of Medical Virology*, 2018, 90(1):148-56. (#: contribute equally)
11. Ji Y, Xiao J, Shen Y, Ma D, Li Z, Pu G, Li X, Huang L, Liu B, Ye H, Wang H. Cloning and Characterization of AabHLH1, a bHLH Transcription Factor that Positively Regulates Artemisinin Biosynthesis in *Artemisia annua*. *Plant Cell Physiology*, 2014, 55(9): 1592-1604.

12. Chen J, Fang H, Ji Y, Pu G, Guo Y, Huang L, Du Z, Liu B, Ye H, Li G, Wang H. Artemisinin Biosynthesis Enhancement in Transgenic *Artemisia annua* Plants by Downregulation of the  $\beta$ -Caryophyllene Synthase Gene. *Planta Medica*, 2011, 77: 1759-1765.
13. Guo Y, Ma L, Ji Y, Pu G, Liu B, Du Z, Li G, Ye H, Wang H. Isolation of the 5'-End of Plant Genes from Genomic DNA by TATA-Box-Based Degenerate Primers. *Molecular Biotechnology*, 2011, 47:152-156.
14. Li X, Ma D, Chen J, Pu G, Ji Y, Lei C, Du Z, Liu B, Ye H, Wang H. Biochemical characterization and identification of a cinnamyl alcohol dehydrogenase from *Artemisia annua*. *Plant Science*, 2012, 193-194:85-95.

## PhD Portfolio

<b>Name of PhD student</b>	Yunpeng Ji
<b>Department</b>	Gastroenterology and Hepatology, Erasmus MC- University Medical Center, Rotterdam
<b>PhD Period</b>	September 2017 – September 2021
<b>Promotor</b>	Prof. dr. Maikel P. Peppelenbosch
<b>Co-promotor</b>	Dr. Qiuwei Pan

### PhD training

---

#### Seminars

- 2019-2021, weekly MDL seminar program in experimental Gastroenterology and Hepatology (attending); (42 weeks/year; @1.5h) (ECTS, 4.5)
- 2019-2021, weekly MDL seminar program in experimental Gastroenterology and Hepatology (presenting); (preparation time 16h; 2 times/year) (ECTS, 2.3)
- 2019-2021, biweekly research group education(attending); (20 times/year; @1.5h) (ECTS, 2.1)
- 2019-2021, biweekly research group education(presenting); (preparation time 8h; 4 times/year) (ECTS, 2.3)

#### General Courses and Workshops

- 2020, The SNP Course XVII: SNPs and Human Diseases (ECTs, 2.0)
- 2020, The Course Biomedical Research Techniques (ECTs, 1.5)
- 2020, The Gene expression data analysis using R: How to make sense out of your RNA-Seq/microarray data (ECTs, 2.0)
- 2020, The FlowJo: High dimensional analysis of flowcytometry data (ECTs, 0.6)
- 2020, The OneNote Organize your digital information (ECTs, 0.3)
- 2020, The PowerApps: App in a day! (ECTs, 0.3)
- 2020, The course on Molecular Diagnostics (ECTs, 1.0)
- 2020, The Biomedical English Writing Course for MSc and PhD students (ECTs, 2.0)
- 2020, The Microscopic Image Analysis: From Theory to Practice (ECTs, 0.8)

- 2020, The Photoshop and Illustrator CC 2019 Workshop for PhD-students and other researchers (ECTs, 0.3)
- 2019, The Survival Analysis Course (ECTs, 0.6)
- 2019, The Workshop on Microsoft Access 2010: Basic (ECTs, 0.3)
- 2019, The Data analysis with Python (ECTs, 1.7)
- 2019, The Two-days Course Bayesians statistics and JASP: Basic & Advanced (ECTs, 0.6)
- 2019, The Course Basic and Translational Oncology (ECTs, 1.8)
- 2019, The SCORE PhD course: Stem Cells, Organoids and Regenerative Medicine (ECTs, 1.0)
- 2019, The Basic course on R (ECTs, 1.8)
- 2019, The Course Basic and Translational Endocrinology: “Emerging Technologies in Endocrine (ECTs, 2.0)
- 2019, The NGS in DNA Diagnostics Course (ECTS, 1.0)

#### **National and International Conferences**

---

- 2017, December 10<sup>th</sup>. The 12<sup>th</sup> Conference of National Genetic Counseling
- 2015, September. 22<sup>th</sup> Chinese Human Placenta Project, Guangzhou, China
- 2015, July. 9<sup>th</sup>. National two cancer screening training, Taicang, China

#### **Academic Awards**

---

- 2021, August. Provincial science foundation (¥60,000), China

## Curriculum Vitae

Yunpeng Ji was born on January 22<sup>th</sup>, 1985 in Hohhot, Inner Mongolia, China. He was raised and finished his primary, secondary, and high school in his hometown.

In 2004, he started his bachelor study in biological technology at Inner Mongolia University, China. Four years later, he formally started his master's research training at the Institute of Botany, the Chinese Academy of Sciences. Under the supervision of Prof. Hong Wang and Prof. Hechun Ye, he worked on transcriptional regulation of secondary metabolism pathways in *Artemisia annua* plant.

In 2018, with the support of the Inner Mongolia Maternal and Child Care Hospital, he got an opportunity to start his PhD research at the Department of Gastroenterology and Hepatology, Erasmus University Medical Center, Rotterdam, the Netherlands. Under the supervision of Prof. Maikel P. Peppelenbosch and Dr. Qiuwei (Abdullah) Pan, he focused on the modeling of infectious disease associated with hepatitis E virus and SARS-CoV-2.

