Check for
updates

Sociodemographic and Policy Factors Associated with the Transmission of COVID-19: Analyzing Longitudinal Contact Tracing Data from a Northern Chinese City

Han Liu · **Zai Liang** · **Shiyong Zhang** · **Lihua Liu**

Abstract To examine how sociodemographic characteristics and non-pharmaceutical interventions affect the transmission of COVID-19, we analyze patient profiles and contact tracing data from almost all cases in an outbreak in Shijiazhuang, China, from January to February 2021. Because of universal testing and digital tracing, the data are of high quality. Results from negative binomial models indicate that the counts of close contacts and secondary infections vary with the cases' age and occupation. Notably, cases under age 18 are causing an increased infection rate among their close contacts and leading to more within-neighborhood secondary infections than adults aged 18–49. Also, county-wide interventions and lockdown are found to be effective at containing the spread of COVID-19. These measures can reduce the number of close contacts that each case has and largely restrict the remaining infections to the case's neighborhood. These results suggest that transmission risks of COVID-19 are associated with the case's sociodemographic characteristics and can be reduced with interventions at the county level. Implications on mitigation measures and reopening plans are discussed.

H. Liu (✉) · Z. Liang (✉)
Department of Sociology, University at Albany, State University of New York, 1400 Washington Avenue, Albany, NY 12222, USA
e-mail: hliu24@albany.edu

Z. Liang
e-mail: zliang@albany.edu

Z. Liang
Department of Sociology, School of Humanities and Social Science of Xi'an Jiaotong University, Xi'an, Shaanxi Province, China

S. Zhang · L. Liu (✉)
Shijiazhuang Center for Disease Prevention and Control, Shijiazhuang, Hebei Province, China
e-mail: liulihua067@126.com

Since the COVID-19 pandemic started, researchers have tried to understand the social determinants of its contagion dynamics. Incorporating these factors into empirical research is considered important for both epidemiological models and policy discussions [1]. In the current literature, most empirical studies on COVID-19 transmission are based on comparisons across nations or sub-national units [2]. These ecological analyses have furthered our understanding of area-level factors driving the pandemic and informed policymakers about the efficacy of non-pharmaceutical interventions [3], but their findings may not apply to the individual-level epidemiological dynamics.

There is also a growing number of studies using individual-level contact tracing data to investigate COVID-19 transmission and evaluate the effectiveness of mitigation policies [4, 5]. These studies have laid the foundation for non-pharmaceutical interventions, but because of limited testing capacities and risks of infringing on privacy, surveillance data used in individual-level research usually have limited representativeness.

In this study, we try to fill this gap in the literature by analyzing patient profiles and contact tracing data from a COVID-19 outbreak in a northern Chinese city, Shijiazhuang, in early 2021. During the outbreak, the municipal government conducted universal testing for all residents and used geolocation data from telecommunication providers to assist contact tracing for both symptomatic and asymptomatic/pre-symptomatic cases. These testing and contact tracing efforts provide high-quality data that support accurate analysis of transmission dynamics at the individual level. Specifically, this study assesses sociodemographic factors associated with transmission risks of COVID-19 and evaluates the effectiveness of county-level interventions. In the face of slow vaccination roll-out in many developing countries [6], threats from new variants [7], and the urgent need to safely reopen schools and other social institutions [8, 9], findings from our analyses will not only advance our understanding of social determinants of COVID-19 transmission but also help the design of mitigation policies and reopening plans in countries around the globe.

Background

Transmission Heterogeneities of COVID-19

The transmission of infectious diseases, like COVID-19, is based on person-to-person contagion. Thus, people with more social contacts, once infected by the virus, are also more likely to transmit it to others. Social network research suggests that network size and structure vary with sociodemographic factors, like age [10, 11], gender [12, 13], and socioeconomic status [11, 14]. While the direction and magnitude of these cross-group differences are modified by how social networks are conceptualized and measured [15], this literature indicates that certain sociodemographic characteristics may facilitate the spread of germs by exposing the host to more social contacts. Therefore, in this study on COVID-19 transmission, we expect the patients' age, gender, and socioeconomic status to have an impact on their numbers of close contacts and secondary infections.

In addition to network size, the duration and closeness of social interactions may also vary by sociodemographic factors. Children, for example, tend to have prolonged exposure to each other when attending school in person. While in-person schooling plays an essential role in children's welfare and education, without adequate mitigation policies, it would lead to rapid transmission within schools and their surrounding communities [8, 16, 17]. Similarly, females typically have more contact with relatives than males [11, 18]. Compared to social interactions in the workplace, these kinship-based interactions tend to be closer and consequently pose higher transmission risks [4]. Socioeconomic status can also affect the dynamics of social interactions through conditions of employment and housing. People working or living in overcrowded settings may not be able to comply with social distancing and other public health guidelines [1]. Therefore, patients' age, gender, and socioeconomic status are also expected to be associated with the risk of causing secondary infections among their social contacts.

Fully evaluating transmission heterogeneities by sociodemographic characteristics would help the design of targeted and cost-effective interventions. However, the collection of high-quality contact-tracing data is costly and faces concerns over data protection and privacy. During the first wave of the pandemic, the capacities of testing and contact tracing were still limited, leaving public health authorities with no choice but to prioritize testing to symptomatic individuals. The ramping up of testing capacities and use of digital technologies, like smartphone applications, in contact tracing can partially address this data limitation. Compared to traditional manual contact tracing, digital tracing is faster and provides a more complete picture [19, 20]. However, it may also infringe on privacy [21]. Therefore, the uptake of government-implemented contact tracing applications

tends to be low and varies from country to country [22, 23].

Impacts of Non-pharmaceutical Interventions.

Because COVID-19 can be transmitted by asymptomatic/pre-symptomatic carriers, epidemic control is unrealistic if case isolation and quarantine of close contacts are the only measures in place [4]. Population-level interventions (e.g., lockdown) have been shown to be effective mitigation strategies [24–26], but they also come with high costs on economic activities [27], employment [28], and mental health [29–31]. These economic and social costs can be lowered with timely small-scale interventions targeted towards places with the highest risk of transmission. Mathematical modeling and empirical research on aggregated data have supported the effectiveness of these measures [32, 33], but to fully understand how they change the dynamics of COVID-19 transmission, we need more evidence from high-quality contact tracing data.

In China, starting from late February 2020, a three-level prevention and control plan was rolled out in provinces where the initial lockdown policies were lifted [34]. This plan requires local governments to dynamically assess the risk of COVID-19 in each county-level jurisdiction (hereafter county) according to two criteria: cumulative cases in the past 14 days and whether cluster cases exist. Counties, or sometimes neighborhoods, are classified as either low-risk, medium-risk, or high-risk. Once the risk level changes from low to medium, the local government would put locations associated with cases under lockdown, activate enhanced surveillance, and conduct disinfection. In high-risk areas, the following policies would be adopted: expanded lockdown, universal screening of symptoms, restrictions on social gatherings, business and school closure, and the activation of the public health emergency management system. Compared to larger-scale lockdowns or restrictions, these interventions at the county/neighborhood level have smaller impacts on economic and social well-being. In this study, we evaluate the effectiveness of these measures using individual-level contact tracing data. Findings from this part of our analysis will be informative for places facing emerging outbreaks of COVID-19 or other infectious diseases.

The Shijiazhuang Outbreak

Shijiazhuang is the capital city of Hebei Province in northern China. It administers 22 counties with a total population of about 11 million. After the first wave of the pandemic in early 2020, the city reported no case for months. However, on January 2nd, 2021, locally infected cases re-emerged, and an outbreak ensued. During the outbreak, two counties, Gaocheng and Xinle, were classified as high-risk areas, and a few neighborhoods in other counties were classified as medium-risk areas. Early interventions resulting from the designation of medium/high-risk areas effectively cleared local transmission in mid-February. In total, there were 1,033 infected residents reported in the city from January 2nd to February 14th.

The Shijiazhuang outbreak provides high-quality data for research on transmission heterogeneities and policy impacts because of two measures taken by the local government and public health authority. First, the city did three rounds of free COVID-19 testing for all its residents, each covering more than 10.25 million people [35, 36, 37]. Namely, these testing efforts have reached almost the entire 11 million population three times. In addition to these large-scale testing campaigns, all close contacts of positive cases are tested three times on the 1st, 2nd, and 14th days after they were identified as a contact. Therefore, epidemiological data collected during this outbreak would be able to include asymptomatic cases that could not be found otherwise. The second measure that facilitates the collection of epidemiological data is the use of geolocation data from mobile phones in contact tracing. China authorized public health departments to use big data in COVID-19 prevention and control in early 2020 [38]. During the Shijiazhuang outbreak, geolocation data from telecommunication providers were used to identify people who might have been exposed to positive cases. Trained epidemiological investigators would then reach out to these people and evaluate whether they should be treated as close contacts according to the guidelines issued by the National Health Commission [34]. This centralized approach of digital tracing does not require individuals to download and use any application on their phones. Therefore, it can provide a more complete picture than software-based tracing techniques.

Data and Methods

We use de-identified patient profiles and contact tracing data from the Shijiazhuang Center for Disease Prevention and Control. The patient profiles include basic sociodemographic information, date of symptom onset, and date of COVID-19 diagnosis with the polymerase chain reaction test. The contact tracing data include all close contacts for each case. For close contacts who also tested positive, their ID in the patient profile data is also available. The universal testing and extensive contact-tracing efforts provide high-quality data that are rarely available in other places. Our analysis includes 1,028 (99.52%) of the 1,033 locally transmitted cases. only 5 cases are dropped because of missing information.

Four outcomes are evaluated in our study: (1) the count of close contacts, including all people who had unprotected contact with the case within four days prior to symptom onset or positive COVID-19 test, whichever earlier; (2) secondary infections, namely all close contacts who also tested positive for COVID-19[1]; (3) secondary infections residing in the same neighborhood as the case, and; (4) secondary infections in other neighborhoods.

Sociodemographic characteristics evaluated in our analysis include gender (male versus female), age (0–17, 18–49, 50–64, and 65 and over), and occupation (peasants, other manual jobs, non-manual jobs, and not employed).[2] Mitigation policies are measured with the risk level of each case's residence at the time of COVID-19 diagnosis. Because only 43 cases were from medium-risk areas, we combine low and medium risks as the reference group. Also, considering that the contact tracing data cover all contacts within four days prior to symptom onset or diagnosis, we expect interventions to reduce transmission risks after 4 days. Thus, we treat 0–4 days into high-risk and more than

4 days into high-risk as different categories. In the analysis, we also control for the county of residence (Gaocheng and Xinle, both are high-risk areas, or others) and whether the case was symptomatic, ranging from coughing to having pneumonia or more severe symptoms, when tested positive [34].

Because all four outcomes are count variables, we use negative binomial regression models in our analysis. Also, in this study, we focus on the net effect of each explanatory variable adjusted for other covariates. Therefore, we estimate coefficients for all explanatory variables in the same model. In addition to analysis on all cases, as a sensitivity test, we replicate all models with cases from Gaocheng and Xinle, the two counties that had been classified as high-risk during the outbreak. By restricting our analysis to these two counties and controlling for county fixed effects, we are essentially comparing cases who tested positive after the transition into high risk with their counterparts from the same county but before the transition. In this way, we further purge out confounding effects of time-constant county-specific characteristics and thus strengthen the validity of estimated policy effects.

Results

Descriptive statistics of the Shijiazhuang cases are shown in Table 1. On average, each case has 23.0 close contacts and 1.3 secondary infections. Compared to findings from other parts of China [4, 5], the Shijiazhuang cases have reported more close contacts. In terms of sociodemographic background, the Shijiazhuang cases are mainly females (58.9%), working-age adults (39.8%), and peasants (61.1%). About three-quarters of them are from high-risk areas. Most cases are from Gaocheng county (83.1%). Also, more than a third of the cases were asymptomatic/pre-symptomatic, showing the strengths of universal testing in identifying these cases.

Figure 1 and Table S1 show the distribution of close contacts and secondary infections that each case has by gender, age, occupation, and risk level. Consistent with previous research [4], we observe an overdispersion in these distributions. While most cases have no more than five close contacts and less than two secondary infections, there are 45 cases (4.5%) linked with more than 100 contacts and 3 cases (0.3%) associated with more than 10 secondary infections. This overdispersion is consistent across all sociodemographic groups and

---

[1] In this study, we assume each index case is the source and his or her identified contacts who turned positive are secondary infections. However, due to the potentially long incubation period of COVID-19, it could be the other way around. We thank an anonymous reviewer for pointing out this possibility.

[2] Peasants include self-employed small farmers and agricultural laborers. Other manual jobs include skilled and unskilled manual occupations that are non-agricultural. Non-manual jobs include all clerical, professional, and managerial occupations. Not employed includes unemployed working-age adults as well as children and retired elders.

Table 1 Descriptive profile of COVID-19 cases in Shijiazhuang, China, January—February 2021

|	N	Mean/%	SD	IQR
Close contacts	1028	23	67.2	(2, 14)
Secondary infections	1028	1.3	1.8	(0, 2)
Gender				
Female	605	58.9		
Male	423	41.2		
Age				
0—17	209	20.3		
18—49	409	39.8		
50—64	249	24.2		
65+	161	15.7		
Occupation				
Peasant	628	61.1		
Other manual	73	7.1		
Non-manual	54	5.3		
Not employed	273	26.6		
Risk level when tested positive				
Low/medium	267	26		
High, 0–4 days	277	27		
High, > 4 days	484	47.1		
County				
Gaocheng	854	83.1		
Xinle	71	6.9		
Other	103	10		
Symptomatic when tested positive				
Asymptomatic/ pre-symptomatic	368	35.8		
Symptomatic	660	64.2		

Standard deviation and inter-quartile range are presented for continuous variables only

risk levels, but there are also noticeable differences across groups. Females tend to have more close contacts than males, but the distribution of secondary infections does not change much with gender (Fig. 1A). In terms of age (Fig. 1B), cases aged 65 or older appear to have fewer close contacts and secondary infections compared to younger adults (18–64). Children, though with fewer close contacts than adults, are leading to more secondary infections. Additional analysis indicates that 32.1% of children's secondary infections are also under age 18 and the rest 67.9% are adults. Occupation also appears to affect the transmission risks (Fig. 1C). Non-manual jobs tend to increase the number of close contacts, but not secondary infections, indicating that most of their contacts might have very

brief exposures or adopted social distancing during the interactions. Importantly, the descriptive results show that intervention policies have an appreciable impact on containing transmission (Fig. 1D). Cases from high-risk areas have fewer close contacts and secondary infections than those from low/medium-risk areas.

Incident rate ratios (IRR) from negative binomial models are shown in Tables 2 and 3. According to Models 1 and 2 in Table 2, there are no gender differences in transmission risks, but age is associated with both close contacts and secondary infections. Cases aged 65 or older have 33% fewer close contacts and 27% fewer secondary infections than those aged 18–49. Occupation is also correlated with the number of close contacts. Compared to peasants, working on other manual jobs reduces the number of close contacts by 48% while having a non-manual job leads to 87% more contacts. However, these differences in the number of close contacts are not causing significant differences in secondary infections. Results from negative binomial models also confirm the effectiveness of interventions in high-risk areas. Within the first four days of implementation, these measures reduce the number of close contacts associated with an average case by 41% and secondary infections by 39%. After four days into high-risk, the numbers of close contacts and secondary infections per case are further reduced to 25% and 51% of their respective levels in low/medium-risk areas.

In model 3, we add the logged number of close contacts as an offset on the basis of model 2 to estimate the rate of close contacts that turn into secondary infections. This specification adjusts for the population at risk, allowing us to interpret the outcome as a rate [39]. Estimating the rate requires at least one close contact. Thus, fifty-four cases with no close contact are dropped in this model. The results show that cases' gender and occupation are not significantly correlated with the infection rate among their close contacts, but age makes a remarkable difference. The close contacts of children (ages 0–17) are 81% more likely to be infected than the contacts of adults aged 18–49. As for the effects of interventions in high-risk areas, the measures do not have an immediate impact on the rate of close contacts getting infected. However, after four days into high-risk, the rate increases to 155% of the level in low/medium-risk areas. This increase in the infection rate among close contacts is essentially consistent with results from Models 1 and 2. Although the number of close contacts (Model 1) gets reduced by
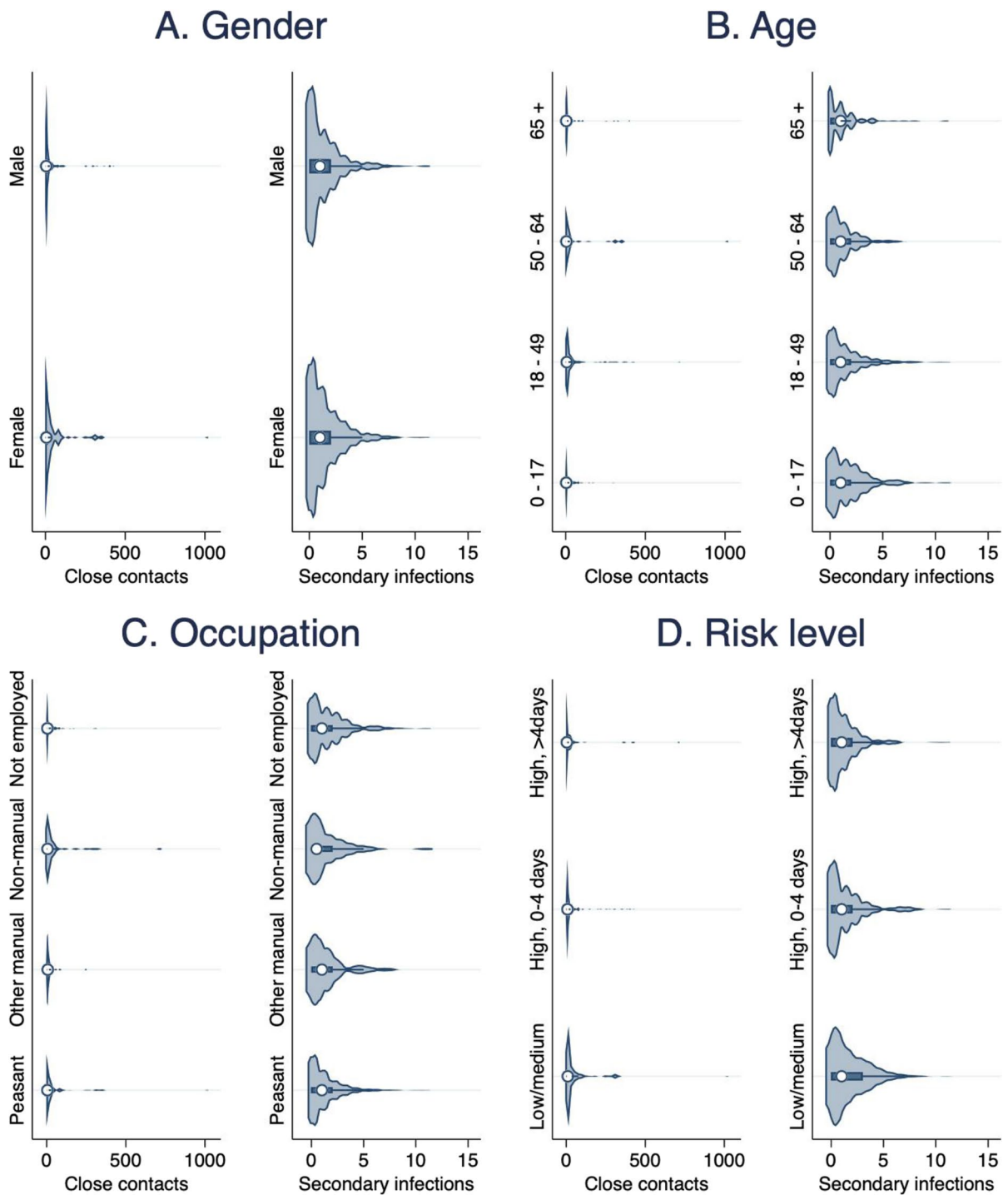
Fig. 1 Violin plots representing the distribution of close contacts and secondary infections by gender, age, occupation, and risk level of COVID-19 cases in Shijiazhuang, China, January—February 2021

Table 2 Negative binomial models predicting close contacts and secondary infections of COVID-19 cases in Shijiazhuang, China, January—February 2021

	Model 1 Close contacts IRR (95% CI)	Model 2 Secondary infections IRR (95% CI)	Model 3 Infection rate IRR (95% CI)
Male	0.91	0.87	0.99
	[0.76,1.10]	[0.74,1.03]	[0.80,1.22]
Age (ref. = 18–49)			
0—17	0.82	1.42	1.81*
	[0.55,1.23]	[0.99,2.05]	[1.13,2.90]
50—64	0.99	0.80	1.05
	[0.77,1.28]	[0.64,1.00]	[0.79,1.40]
65 +	0.67**	0.73*	1.26
	[0.50,0.88]	[0.56,0.95]	[0.90,1.76]
Occupation (ref. = peasant)			
Other manual	0.52***	1.13	1.17
	[0.36,0.76]	[0.81,1.57]	[0.78,1.77]
Non-manual	1.87**	1.03	0.69
	[1.22,2.88]	[0.70,1.53]	[0.41,1.16]
Not employed	0.78	1.03	1.01
	[0.53,1.15]	[0.73,1.46]	[0.64,1.59]
Risk level (ref. = low/medium)			
High, 0–4 days	0.59***	0.71*	1.09
	[0.44,0.79]	[0.54,0.95]	[0.75,1.58]
High, > 4 days	0.25***	0.51***	1.55*
	[0.18,0.34]	[0.37,0.68]	[1.04,2.30]
County (ref. = Gaocheng)			
Xinle	0.37***	0.95	1.85*
	[0.25,0.55]	[0.66,1.37]	[1.13,3.03]
Other	0.98	0.77	0.76
	[0.66,1.46]	[0.53,1.11]	[0.47,1.23]
Symptomatic	1.16	1.40***	1.51***
	[0.92,1.46]	[1.15,1.69]	[1.19,1.93]
ln(close contacts)			1
Constant	50.79***	1.70***	0.09***
	[37.12,69.50]	[1.29,2.25]	[0.06,0.13]
alpha	1.91	0.90	1.40
	[1.76,2.06]	[0.74,1.09]	[1.21,1.61]
N	1028	1028	974

Model 1 estimates effects of sociodemographic and policy factors on the count of close contacts with no offset variable

Model 2 estimates effects of sociodemographic and policy factors on the count of secondary infections with no offset variable

Model 3 estimates effects of sociodemographic and policy factors on the count of secondary infections with the number of close contacts as the offset variable. Therefore, the outcome should be interpreted as infection rates

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, two-tailed tests

75% after four days into high-risk, the corresponding reduction in secondary infections (Model 2) is much smaller (49%).

In Table 3, we estimate separate models for secondary infections from the same neighborhood as the case (Model 4) and those from other neighborhoods (Model 5). The results suggest that sociodemographic factors only affect transmission risks within neighborhoods. Children are associated with 64% more secondary infections in their neighborhoods than adults aged 18–49.

Table 3 Negative binomial models predicting secondary infections in the same neighborhood and other neighborhoods of COVID-19 cases in Shijiazhuang, China, January—February 2021

	Model 4 Secondary infections, same neighborhood IRR (95% CI)	Model 5 Secondary infections, other neighborhoods IRR (95% CI)
Male	0.89	0.82
	[0.74,1.07]	[0.57,1.17]
Age (ref. = 18–49)		
0—17	1.64*	1.22
	[1.09,2.47]	[0.57,2.65]
50—64	0.87	0.66
	[0.68,1.11]	[0.40,1.08]
65 +	0.77	0.61
	[0.58,1.03]	[0.35,1.07]
Occupation (ref. = peasant)		
Other manual	1.23	0.86
	[0.86,1.76]	[0.42,1.78]
Non-manual	0.60*	1.77
	[0.37,0.98]	[0.83,3.81]
Not employed	0.93	1.14
	[0.62,1.38]	[0.56,2.34]
Risk level (ref. = low/medium)		
High, 0–4 days	0.80	0.56*
	[0.58,1.10]	[0.32,0.98]
High, > 4 days	0.69*	0.22***
	[0.49,0.96]	[0.12,0.40]
County (ref. = Gaocheng)		
Xinle	0.86	1.11
	[0.57,1.31]	[0.54,2.27]
Other	0.67	0.86
	[0.44,1.02]	[0.41,1.79]
Symptomatic	1.41**	1.39
	[1.13,1.75]	[0.93,2.08]
Constant	1.03	0.70
	[0.75,1.41]	[0.39,1.23]
alpha	0.98	4.30
	[0.78,1.23]	[3.31,5.58]
N	1028	1028

Model 4 estimates effects of sociodemographic and policy factors on the count of secondary infections from the same neighborhood with no offset variable

Model 5 estimates effects of sociodemographic and policy factors on the count of secondary infections from other neighborhoods with no offset variable

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, two-tailed tests

Also, non-manual workers, compared to peasants, have 40% fewer secondary cases from the same neighborhood, probably because interactions among neighbors are more frequent in rural areas than in urban areas. The impact of county-wide interventions, in contrast, appears to increase with spatial and social distance. For within-neighborhood infections, interventions in high-risk areas have no significant impact in the first four days. However, after this initial period of time, the transmission is reduced by 31%. For transmission across neighborhoods, the measures lead to a significant 44% drop in secondary cases right after the designation of high-risk areas. After the first four days, the effect size further increases to a reduction of 78%.

Tables S2 and S3 show results from our sensitivity tests based on cases from Gaocheng and Xinle only. According to the tables, only one of the 10 policy coefficients shows a change in statistical significance. In Model 5, where we estimate the impacts on secondary infections from other neighborhoods, the coefficient of 0–4 days into high-risk is significant with all cases (Table 3), but no longer significant after we restrict the analysis to Gaocheng and Xinle. Despite this change in statistical significance, the difference in IRR is negligible (0.57 versus 0.56), confirming the pattern we found with all cases.

Conclusions and Discussion

By analyzing high-quality contact tracing data, this study highlights two important findings. First, the numbers of close contacts and secondary infections vary with sociodemographic characteristics. Compared to cases with non-manual jobs, peasants and urban manual workers tend to have fewer close contacts, but they are causing more secondary infections within their neighborhoods. Cases aged 65 or older have fewer close contacts and secondary infections compared to their working-age counterparts. Children, in contrast, are associated with an increased infection rate among their close contacts and a large number of secondary infections in their neighborhoods. The high infection rates among contacts of children might be due to two reasons. First, children are probably not as precautious as adults and may not take enough protective behaviors (e.g., proper mask-wearing, handwashing, or social distancing) when interacting with other people. Second, our additional analysis indicates that the majority of people infected by children are adults, suggesting that these secondary cases are likely to

be their parents or other household members. These contacts tend to have prolonged and frequent exposure to the case, which can increase the risk of infection. Future research with more detailed case–contact relationship data is needed as ascertaining children's role in the spread of COVID-19 is critical to the design of mitigation policies and reopening plans.

Second, the three-level prevention and control plan in China appears to be effective at containing the spread of COVID-19. Especially, we find that county-wide interventions adopted in high-risk areas, a combination of expanded lockdown, universal screening of symptoms, restriction on gatherings, and business and school closure, can drastically reduce infections by lowering the number of close contacts associated with each case. Also, these measures have a greater impact on transmission across neighborhoods than that within the same neighborhood. These results suggest that social interactions in high-risk areas are likely to be restricted to those with high frequency and closeness (e.g., among family/household members[3]), who also bear increased risks of infection if one of the participants carries the virus.

Because of limited information collected and reported in the patient profiles and contact tracing data, our results may have been confounded by several unobservable or unmeasured factors, like income or household size. Also, regarding the effects of county-wide interventions, although only two counties were classified as high-risk during the outbreak, other counties also implemented mobility restrictions for short periods of time. These spontaneous reactions might have lowered infections in low/medium-risk areas and led to an underestimation of the reduction in transmission risks seen in high-risk areas. However, sensitivity tests indicate that the policy effects do not change much after we restrict the analysis to counties that actually experienced the policy transition.

The generalizability of our findings is restricted by some conditions specific to the Shijiazhuang outbreak. First, the city had been reporting zero cases until local infections reappeared in January 2021. Thus, the local residents probably did not take adequate protective measures before they noticed the outbreak. This lack of precaution might have facilitated the spread of the virus and resulted in an increased

---

[3] Because the patient profiles do not have detailed address information for rural cases, which account for more than 60% of cases covered in our analysis, we are not able to evaluate within-household transmission in this study.

number of secondary infections. Second, the initial local transmissions in Shijiazhuang are thought to be associated with village wedding parties [40]. These events exposed their attendees, mainly peasants, to more social contacts than they usually have and might have an impact on our estimation of transmission heterogeneities across occupational groups. Third, by the time of the outbreak, vaccines were only available to health care workers and other occupational groups with high infection risks [41]. Therefore, only a few cases covered in our analysis were eligible for inoculation. Considering that vaccines can reduce the viral shedding of COVID-19 cases, we expect transmission rates to lower with the rollout of vaccination campaigns. While we are not able to take into account vaccination in this study, given the stark disparity gap between vaccination programs in different countries, our findings are still relevant for places with low vaccination rates, including most countries in Africa and some Eurasian countries [42]. Last, a large proportion of cases in the Shijiazhuang outbreak are from rural areas, so our findings may not be generalizable to countries with a small agricultural industry. Nevertheless, given that about 45% of the world population still resides in rural areas [43], our findings may inform epidemiological interventions with considerable returns in population health.

Despite these limitations, this study sheds new light on the prevention and control of COVID-19 as well as other infectious diseases. Our analysis indicates that timely non-pharmaceutical interventions, including restrictions on gatherings and school closure, can effectively contain further infections via contact reduction. Especially, when implemented in small areas with the highest caseload, these measures can drastically reduce infections with relatively small economic and social costs. Also, the sociodemographic heterogeneities in transmission risks revealed in this study can guide the epidemic control and reopening plans. Our data show that peasants and manual workers have more within-neighborhood secondary infections than those with non-manual jobs, suggesting that small-scale interventions would be more effective in less developed areas with larger rural populations. As for age, we find that the close contacts of children have a higher infection rate than contacts of adults. This finding echoes concerns about reopening schools recently expressed by epidemiologists [44]. While children have a low probability of having severe symptoms after being infected by COVID-19, they can seed the spread in the larger society by infecting their household members and other adults living in their neighborhoods. These adults can then transmit the disease to their own social contacts. As children return to in-person classes, countries, including China and the United States, reported spikes in children's cases and outbreaks in schools [45, 46]. Future studies on how to control within-school infections are therefore urgently needed.

References

1. Galanis G, Hanieh A. Incorporating Social Determinants of Health into Modelling of COVID-19 and other Infectious Diseases: A Baseline Socio-economic Compartmental Model. *Soc Sci Med*. 2021;274:113794.

2. Fatima M, O'Keefe KJ, Wei W, Arshad S, Gruebner O. Geospatial Analysis of COVID-19: a Scoping Review. *Int J Environ Res Public Health*. 2021;18:2336.

3. Perra N. Non-Pharmaceutical Interventions During the COVID-19 Pandemic: a Review. *Phys Rep*. 2021;913:1–52.

4. Sun K, Wang W, Gao L, Wang Y, Luo K, Ren L, et al. Transmission Heterogeneities, Kinetics, and Controllability of SARS-CoV-2. *Science*. 2021;371:eabe2424.

5. Yang Z. Analysis of Dynamic Contact Network of Patients with COVID-19 in Shaanxi Province of China. *Sci Rep*. 2021;11:4889.

6. Mathieu E, Ritchie H, Ortiz-Ospina E, Roser M, Hasell J, Appel C et al. A Global Database of COVID-19 Vaccinations. *Nat Hum Behav*. 2021 ; Available from: http://www.nature.com/articles/s41562-021-01122-8. Accessed 9 July 2021

7. Mahase E. Delta Variant: What is Happening with Transmission, Hospital Admissions, and Restrictions? *BMJ*. 2021;373:n1513.

8. Ladhani SN, the sKIDs Investigation Team. Children and COVID-19 in Schools. *Science*. 2021;374:680–2.

9. Lordan R, FitzGerald GA, Grosser T. Reopening Schools during COVID-19. *Science*. 2020;369:1146–1146.

10. Hill RA, Dunbar RIM. Social Network Size in Humans. *Hum Nat*. 2003;14:53–72.

11. Marsden PV. Core Discussion Networks of Americans. *Am Sociol Rev*. 1987;52:122.

12. Dickens WJ, Perlman D. *Friendship over the Life-cycle*. In: Duck S, Gilmour R, editors. Pers Relatsh. New York, NY: Academic Press; 1981. p. 91–122.

13. Dunbar RI, Spoors M. Social Networks, Support Cliques, and Kinship. *Hum Nat Springer*. 1995;6:273–90.

14. Belle DE. The Impact of Poverty on Social Networks and Supports. *Marriage Fam Rev*. 1983;5:89–103.

15. Bernard HR, Johnsen EC, Killworth PD, McCarty C, Shelley GA, Robinson S. Comparing Four Different Methods for Measuring Personal Social Networks. *Soc Netw*. 1990;12:179–215.

16. Lessler J, Grabowski MK, Grantz KH, Badillo-Goicoechea E, Metcalf CJE, Lupton-Smith C et al. Household COVID-19 Risk and In-person Schooling. 2021;372:1092–7.

17. Stein-Zamir C, Abramson N, Shoob H, Libal E, Bitan M, Cardash T et al. A Large COVID-19 Outbreak in a High School 10 Days After Schools' Reopening, Israel, May 2020. Eurosurveillance. 2020;25. Available from: https://www.eurosurveillance.org/content/https://doi.org/10.2807/1560-7917.ES.2020.25.29.2001352.

18. Ridgeway CL, Smith-Lovin L. The Gender System and Interaction. *Annu Rev Sociol*. 1999;25:191–216.

19. Anglemyer A, Moore TH, Parker L, Chambers T, Grady A, Chiu K et al. Digital Contact Tracing Technologies in Epidemics: A Rapid Review. Cochrane Public Health Group, editor. Cochrane Database Syst Rev. 2020 [cited 2021 Jul 11]; Available from: http://doi.wiley.com/https://doi.org/10.1002/14651858.CD013699.

20. Ferretti L, Wymant C, Kendall M, Zhao L, Nurtay A, Abeler-Dörner L, et al. Quantifying SARS-CoV-2 Transmission Suggests Epidemic Control with Digital Contact Tracing. *Science*. 2020;368:eabb6936.

21. Whitelaw S, Mamas MA, Topol E, Van Spall HGC. Applications of Digital Technology in COVID-19 Pandemic Planning and Response. *Lancet Digit Health*. 2020;2:e435–40.

22. Barbaschow A. Australian Committee Calls for Independent Review of COVIDSafe App . ZDNet. 2020. Available from: https://www.zdnet.com/article/australian-committee-calls-for-independent-review-of-covidsafe-app/. Accessed 18 July 2021.

23. Findlay S, Palma S, Milne R. Coronavirus Contact-tracing Apps Struggle to Make an Impact. *Financ Times*. 2020. Available from: https://www.ft.com/content/21e438a6-32f2-43b9-b843-61b819a427aa. accessed 8 July 2021.

24. Cowling BJ, Ali ST, Ng TWY, Tsang TK, Li JCM, Fong MW, et al. Impact Assessment of Non-Pharmaceutical Interventions Against Coronavirus Disease 2019 and Influenza in Hong Kong: an Observational Study. *Lancet Public Health*. 2020;5:e279–88.

25. Pan A, Liu L, Wang C, Guo H, Hao X, Wang Q, et al. Association of Public Health Interventions With the Epidemiology of the COVID-19 Outbreak in Wuhan, China. *JAMA*. 2020;323:1915.

26. Park SW, Sun K, Viboud C, Grenfell BT, Dushoff J. Potential Role of Social Distancing in Mitigating Spread of Coronavirus Disease, South Korea. *Emerg Infect Dis*. 2020;26:2697–700.

27. Mandel A, Veetil V. The Economic Cost of COVID Lockdowns: an Out-of-Equilibrium Analysis. *Econ Disasters Clim Change*. 2020;4:431–51.

28. Hoehn-Velasco L, Silverio-Murillo A, Balmori de la Miyar JR. The Long Downturn: The Impact of the Great Lockdown on Formal Employment. *J Econ Bus*. 2021;105983.

29. Bierman A, Schieman S. Social Estrangement and Psychological Distress before and during the COVID-19 Pandemic: patterns of Change in Canadian Workers. *J Health Soc Behav*. 2020;61:398–417.

30. Chen X, Zou Y, Gao H. Role of Neighborhood Social Support in Stress Coping and Psychological Wellbeing During the COVID-19 Pandemic: Evidence from Hubei, China. *Health Place*. 2021;69:102532.

31. Czeisler MÉ, Howard ME, Rajaratnam SMW. Mental Health During the COVID-19 Pandemic: challenges, Populations at Risk, Implications, and Opportunities. *Am J Health Promot*. 2021;35:301–11.

32. Davies NG, Barnard RC, Jarvis CI, Russell TW, Semple MG, Jit M, et al. Association of Tiered Restrictions and a Second Lockdown with COVID-19 Deaths and Hospital Admissions in England: a Modelling Study. *Lancet Infect Dis Elsevier*. 2021;21:482–92.

33. Ge J, He D, Lin Z, Zhu H, Zhuang Z. Four-tier Response System and Spatial Propagation of COVID-19 in China by a Network Model. *Math Biosci*. 2020;330:108484 (**Elsevier**).

34. National Health Commission of the PRC, Bureau of Disease Prevention and Control. Prevention and Control Plan of Coronavirus Disease 2019 (Seventh Edition). 2020. Available from: http://www.gov.cn/xinwen/2020-09/15/5543680/files/931ddbb00e134025bbb86739c8ed68ee.pdf. Accessed 18 July 2021.

35. Feng W, Du Y, Zhao D, Wang J, Zhao W, Wang Y. 8 New Confirmed Cases of New Coronary Pneumonia in Shijiazhuang City from 0:00 to 18:00 on the 22nd; 3 New Cases of Newly Confirmed Cases in Beijing Daxing from 0:00 to 24:00 on the 21st [Internet]. Newsdirectory3. 2021 [cited 2021 Jul 12]. Available from: https://www.newsdirectory3.com/8-new-confirmedcases-of-new-coronary-pneumonia-in-shijiazhuang-city-from-000-to-1800-on-the-22nd-3-newcases-of-newly-confirmed-cases-in-beijing-daxing-from-000-to-2400-on-the-21st_infected/

36. Global Times. Shijiazhuang Completes Sweeping COVID-19 Tests with 354 Positive Results [Internet]. Glob. Times. 2021 [cited 2021 Jul 12]. Available from: https://www.globaltimes.cn/page/202101/1212288.shtml

37. Zhang T. Shijiazhuang's Second Round of Nucleic Acid Testing Found 247 Positive Cases [Internet]. Teller Rep. 2021 [cited 2021 Jul 12]. Available from: https://www.tellerreport.com/life/2021-01-15-%0A-shijiazhuang-s-second-round-of-nucleicacid-testing-found-247-positive-cases%0A-.rJ-WT1z1yu.html

38. Zhong R, Creemers R, Webster G. Translation: Chinese Authorities Emphasize Data Privacy and Big Data Analysis in Coronavirus Response. *New Am*. 2020. Available from: https://www.newamerica.org/cybersecurity-initiative/digichina/blog/translation-chinese-authorities-emphasize-data-privacy-and-big-data-analysis-coronavirus-response/. Accessed 18 July 2021.

39. Cameron AC, Trivedi PK. *Microeconometrics Using Stata*. College Station, Texas: Stata Press; 2009.

40. Myers SL. Facing New Outbreaks, China Places Over 22 Million on Lockdown. NY Times. 2021 Jan 13; Available from: https://www.nytimes.com/2021/01/13/world/asia/china-covid-lockdown.html. Accessed 18 July 2021.

41. Xinhua. China Vaccinating Key Groups Against Covid-19. Xinhuanet. 2021. Available from: http://www.xinhuanet.com/english/2021-01/04/c_139641409.htm. Accessed 18 July 2021.

42. Holder J. *Tracking Coronavirus Vaccinations Around the World*. NY Times. 2022.

43. UN Department of Economic and Social Affairs. *World Urbanization Prospects: The 2018 Revision*. New York, NY: United Nations; 2018.

44. Miller CC, Quealy K, Sanger-Katz M. 723 Epidemiologists on When and How the U.S. Can Fully Return to Normal. N Y Times. 2021 May 15; Available from: https://www.nytimes.com/2021/01/13/world/asia/china-covid-lockdown.html. Accessed 18 July 2021.

45. Leatherby L, Walker AS. *Least Vaccinated States Lead Spike in Children's Cases, Leaving Some Hospitals Stretched*. New York: NY Times; 2021.

46. Woo R, Liu R, Huang K, Qiu S, Shen Y, Gao L. COVID-19 Cases in Southeast China More than Double as Delta Spreads. *Reuters*. 2021. Available from: https://www.reuters.com/world/china/china-repor ts-92-new-covid-19-cases-sept-13-vs-49-day-ago-2021-09-14/. Accessed 18 July 2021.

47. Schuetz AN, Hemarajata P, Mehta N, Campbell S, Mitchell S, Palavecino E et al. When Should Asymptomatic Persons Be Tested for COVID-19? *J Clin Microbiol*. 2020;59. Available from: https://journals.asm.org/doi/10.1128/JCM.02563-20. Accessed 18 July 2021.

48. Wu JT, Leung K, Bushman M, Kishore N, Niehus R, de Salazar PM, et al. Estimating Clinical Severity of COVID-19 from the Transmission Dynamics in Wuhan, China. *Nat Med*. 2020;26:506–10.

49. Ahmed N, Michelin RA, Xue W, Ruj S, Malaney R, Kanhere SS, et al. A Survey of COVID-19 Contact Tracing Apps. *IEEE Access*. 2020;8:134577–601.

50. Zhang J, Litvinova M, Liang Y, Wang Y, Wang W, Zhao S, et al. Changes in Contact Patterns Shape the Dynamics of the COVID-19 Outbreak in China. *Science*. 2020;368:1481–6.