

RESEARCH

Open Access



# The role of vaccine status homophily in the COVID-19 pandemic: a cross-sectional survey with modelling

Elisha B. Are<sup>1,2\*</sup>, Kiffer G. Card<sup>2,3,4</sup> and Caroline Colijn<sup>1,2</sup>

## Abstract

**Background** Vaccine homophily describes non-heterogeneous vaccine uptake within contact networks. This study was performed to determine observable patterns of vaccine homophily, as well as the impact of vaccine homophily on disease transmission within and between vaccination groups under conditions of high and low vaccine efficacy.

**Methods** Residents of British Columbia, Canada, aged  $\geq 16$  years, were recruited via online advertisements between February and March 2022, and provided information about vaccination status, perceived vaccination status of household and non-household contacts, compliance with COVID-19 prevention guidelines, and history of COVID-19. A deterministic mathematical model was used to assess transmission dynamics between vaccine status groups under conditions of high and low vaccine efficacy.

**Results** Vaccine homophily was observed among those with 0, 2, or 3 doses of the vaccine. Greater homophily was observed among those who had more doses of the vaccine ( $p < 0.0001$ ). Those with fewer vaccine doses had larger contact networks ( $p < 0.0001$ ), were more likely to report prior COVID-19 ( $p < 0.0001$ ), and reported lower compliance with COVID-19 prevention guidelines ( $p < 0.0001$ ). Mathematical modelling showed that vaccine homophily plays a considerable role in epidemic growth under conditions of high and low vaccine efficacy. Furthermore, vaccine homophily contributes to a high force of infection among unvaccinated individuals under conditions of high vaccine efficacy, as well as to an elevated force of infection from unvaccinated to suboptimally vaccinated individuals under conditions of low vaccine efficacy.

**Interpretation** The uneven uptake of COVID-19 vaccines and the nature of the contact network in the population play important roles in shaping COVID-19 transmission dynamics.

**Keywords** COVID-19, Vaccine, Homophily, Contact network, Mathematical model, Transmission

## Introduction

COVID-19 is a respiratory illness caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which is transmitted predominantly via aerosols and droplets [1]. In high-income countries, the general population case fatality rate of COVID-19 is sufficiently high to necessitate widespread public health interventions and targeted protections for vulnerable populations, such as seniors and people who are immunocompromised [2].

\*Correspondence:

Elisha B. Are  
eare@sfu.ca

<sup>1</sup> Mathematics, Simon Fraser University, Burnaby, BC, Canada

<sup>2</sup> Pacific Institute On Pathogens, Pandemics and Society (PIPPS), Simon Fraser University, Burnaby, BC, Canada

<sup>3</sup> Faculty of Health Sciences, Simon Fraser University, Burnaby, BC, Canada

<sup>4</sup> Institute for Social Connection, Victoria, BC, Canada



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Fortunately, several safe and effective vaccines are available that can prevent severe COVID-19 and reduce mortality risk, although they have lower effectiveness against transmission than initially hoped [3]. At the individual level, the effectiveness of these vaccines wanes over time, and is subject to immune escape [4]. At the population level, the effectiveness of these vaccines is also dependent on their uptake within and across geographic regions and social networks [5, 6]. Of course, vaccine uptake is heterogeneous within any given population, and this heterogeneity may create disproportionate risk for SARS-CoV-2 transmission within and across communities.

Vaccine hesitancy is an important factor shaping vaccine uptake [7]. A 2014 systematic review documented a range of factors that influence vaccine hesitancy, including contextual influences (e.g., politics, government, religion, geographic patterns, media); individual and social group influences (e.g., beliefs, attitudes, knowledge, trust in healthcare systems and providers); and vaccine-specific issues (e.g., mode of administration and delivery, vaccination schedules, risk vs. benefit) [8]. The results showed that vaccine status tends to cluster with sociodemographic characteristics, such as age, socioeconomic status, race/ethnicity, and political orientation [8–10].

Homophily is a principle in sociology and mathematical modelling that describes the clustering of individual-level characteristics, such as vaccination status, within social networks [11]. Kadelka and McCombs [12] suggested that vaccine homophily may impact COVID-19 vaccine effectiveness given the potential for uneven vaccination uptake. Modelling studies have explored the impact of homophily in a range of contexts, and its impact on transmission dynamics is well documented [13, 14]. For instance, one modelling study argued that the mixing of vaccinated and unvaccinated groups contributes to a considerable risk of infection for the vaccinated group, occurring at a rate that is disproportionately higher than what would be expected based solely on the contact between the two groups [15]. However, these previous studies were not based on descriptive data regarding vaccine homophily. Broadly, empirical research related to vaccine homophily in the context of the COVID-19 pandemic has been limited. Therefore, it is important to describe COVID-19 vaccine homophily and its relationship to vaccination status to gain an improved understanding of COVID-19 transmission [16–19].

The present study was performed to characterize observable patterns of vaccine homophily and to determine the impact of vaccine homophily on COVID-19 transmission within and between vaccination groups under conditions of high and low vaccine efficacy. In doing so, this paper makes important contributions by connecting vaccination and contact heterogeneity, which

are two crucial determinants of transmission dynamics and assesses the impacts of these determinants under low and high vaccination efficacy scenarios.

## Methods

### Participant recruitment

Participants were recruited using paid Facebook and Instagram advertisements (Fig. 1) between February 16, 2022, and March 3, 2022, a period during which the average number of new COVID-19 cases in British Columbia, Canada was declining (7-Day Rolling Average: 865 on February 16, 487 on March 3) and the province continued to experience high numbers of Omicron variant infections [20].

### Data collection

After providing informed consent, potential participants recruited via Facebook and Instagram advertisements were screened for eligibility. The eligibility criteria restricted participation to individuals aged 16 years or older living in British Columbia (BC), Canada. Participants completed an online survey delivered in English using the Qualtrics platform, which assessed participants' history of COVID-19, the extent to which they were following provincial mandates and guidelines for COVID-19 prevention, and how many COVID-19 vaccine doses they had received. Participants also reported on the perceived COVID-19 history of their regular contacts, the perceived level of compliance to COVID-19 prevention guidelines and mandates among regular contacts, the vaccination status of their household and non-household contacts, and the number of household and non-household contacts with whom they had recent contacts. Supplemental Table S1 provides an overview of how these variables were measured by providing the question text and response options.

Additionally, the following demographic data of the participants were collected: age (numerical), gender (Male; Female; Non-binary), ethnicity (African, Caribbean, or Black; Arab or West Asia; East Asian; Indigenous; Latin American; South Asian; Southeast Asian; White; Other), education level (Some high school; High school diploma or equivalent; Some college or trades training; Some university; College or trades certificate or diploma; University degree or higher), annual household income (\$0 to \$150,000 or higher), and whether participants were born in Canada (Yes; No, moved to Canada in the last 5 years; No, moved to Canada more than 5 years ago). Postal code was also assessed and was used to assign participants to one of the 5 regional health authorities in BC.



**Fig. 1** Example meta advertisement used for participant recruitment

## Data analysis

### Aim 1. Characterizing vaccine homophily and analyzing

*Its association with COVID-19 risk factors and prevention behaviours* To characterize observable patterns of vaccine homophily, self-reported vaccination status, COVID-19 prevention behaviours, contact network size, and self-reported COVID-19 infection, descriptive analyses of survey responses were conducted in R version 4.1.3. [21]. Data were cleaned using the *Tidyverse* collection of R packages [22]. As a preliminary step, participants with missing data on demographic-variables (i.e., age, gender, ethnicity, income, education level, immigration status, and health authority) or poor-quality responses (i.e., those in which incongruent responses were provided across questions, indicating imprecise answering) were removed from the analysis. Removal of data with missing demographic variables was done because our sample

weighting procedure was not tolerant of missing data [23]. The remaining observations were weighted by weighting variables using iterative proportional fitting raking estimation, which is a well-established approach for multi-variable weighting when only the marginal proportions for each variable are known [24–26]. Raking estimation was implemented using the *anesrake* package [27] with marginal proportions for each weighting variable derived from the 2016 Canadian Census Profile for BC [28]. The *survey* package was used to generate weighted descriptive statistics [29]. Table 1 provides target weights used in weighting.

Weighted descriptive data were plotted using the *questionr* and *ggplot2* packages [30, 31]. For descriptive statistics, all observations were included in the weighted sample without removing them due to non-response (e.g., early survey drop off, refusal to answer). This allowed us

**Table 1** Target weights

Demographic	Target %
<b>Age</b>	
29 or younger	0.215
30–39	0.153
40–49	0.156
50–59	0.179
60–69	0.155
70 or older	0.142
<b>Gender</b>	
Men	0.485
Women	0.505
Non-binary	0.010
<b>Ethnicity</b>	
White	0.637
East Asian	0.166
South and Southeast Asian	0.092
Indigenous	0.059
Other	0.046
<b>Educational Attainment</b>	
Some Post-Secondary Training	0.304
High School Diploma or Less	0.450
University Degree	0.246
<b>Household Income</b>	
Less than \$30,000	0.189
\$30,000 to \$59,999	0.240
\$60,000 to \$89,999	0.195
\$90,000	0.376
<b>Born in Canada</b>	
No	0.305
Yes	0.695
<b>Health Authority</b>	
Fraser	0.368
Interior	0.157
Island	0.164
Northern	0.062
Vancouver Coastal	0.249

to make best use of the data available. All variables had less than 5% missingness.

To understand clustering between risk factors for COVID-19 and participant's self-reported vaccination status, the chi-square and Kruskal–Wallis tests were used to compare participants with 0, 1, 2, or 3 or more doses with regard to select COVID-19 related personal and network characteristics. Post-hoc pairwise comparisons were also conducted to examine the differences in network size across different vaccine dose categories (i.e., Zero doses, One dose, Two doses, Three doses) using the Wilcoxon rank sum test with a continuity correction. In these

additional analyses, the Benjamini-Hochberg procedure was employed for adjusting  $p$ -values to control for the false discovery rate in multiple comparisons.

The following variables are determined based on the outcomes of the survey:

- **Average Number of Doses Among Contacts.** The average numbers of doses among household, non-household, and overall contacts were calculated using self-reported estimates of the number of doses that participants believed each of their contacts had received.
- **Proportion of Contacts with  $\geq 1$  Dose.** The proportion of contacts with  $\geq 1$  dose was calculated using self-reported data on the number of doses that participants believed each of their contacts had received.
- **Prevalence adjusted Vaccine Homophily.** As homophily estimates depend on the prevalence of each vaccination group, we calculated prevalence adjusted homophily (PAH) using a method proposed in [32]. We define prevalence-adjusted vaccine homophily  $h_i$  as follows:

$$h_i = \begin{cases} \frac{\delta_i - q_i}{1 - q_i} & \text{if } \delta_i \geq q_i, \\ \frac{\delta_i - q_i}{q_i} & \text{if } \delta_i < q_i, \end{cases} \quad (1)$$

where  $\delta_i = \frac{w_i + d_i}{\sum_{j=0}^3 (w_j + d_j)}$ , which represents the proportion of contacts in an individual's network that are of the same vaccination status. Here,  $q_i$  represents the prevalence of vaccination status  $i$  which is calculated from vaccination uptake data in BC during the study period [33]. The term  $w_i$  denotes the number of people in the contact network of an individual with  $i$  doses who have also received  $i$  doses, while  $d_i$  represents the number of people in the household of an individual with  $i$  doses who have received  $i$  doses themselves. Each participant, therefore, has their own  $h_i$  value. To validate the mathematical expression for  $h_i$ , we consider three special cases:

Firstly,  $\delta_k = 1$  for the  $k^{th}$  individual, possible only if that individual only contacts those in their group. From Eq. (1), it is easy to show that in this case  $h_k = 1$ . This implies perfect homophily.

Secondly, if  $\delta_k = 0$  for the  $k^{th}$  individual, then none of the contacts are with people of the same vaccination status. In this case,  $h_k = -1$ , which implies perfect heterophily.



Thirdly, if the  $k^{th}$  individual does not have mixing preferences and prefers to mix randomly between vaccination groups, we can expect  $\delta_k$  to be approximately the prevalence of the  $k^{th}$  group, implying  $\delta_k = q_k$ , which yields  $h_k = 0$ . We have  $h_i \in [-1, 1]$  for all  $i$ , with -1 and 1 at the appropriate extremes and 0 reflecting no contact preference (by vaccination status). In summary, the PAH score is calculated by subtracting the prevalence of the participant's vaccination status from the fraction of their contact network with the same vaccination status, and then dividing by the prevalence of all other vaccination statuses. This is done if the fraction of the participant's contact network with the same vaccination status is greater than or equal to the prevalence of their group. Conversely, if this fraction is less than the prevalence of the individual's group, then the division is done by the prevalence of the individual's group.

- **Blau's Heterogeneity Index Score.** For each participant, we also calculated the diversity of vaccination statuses in their social network using Blau's heterogeneity index [34], calculated as 1 minus the sum (over the numbers of doses,  $k$ ) of the squared fraction of the participant's contact networks with  $k$  doses ( $p_k^2$ ):

$$H = 1 - \sum_{k=0}^3 p_k^2.$$

Blau's heterogeneity index scores were calculated for the number of doses ( $k=0, 1, 2, 3$ ) for each participant's overall contact network and dichotomized vaccination status (i.e.,  $k \geq 1$  dose vs.  $< 1$  dose) for each participant's household, non-household, and overall contact networks.

Homophily estimates and diversity estimates were calculated across each level of vaccination, and associations with select personal and network characteristics were tested using the Kruskal–Wallis H test (for associations with categorical variables) and Spearman's rank correlation tests (for associations with continuous variables).

## Aim 2. Demonstration of the impact of vaccine homophily on COVID-19 transmission

To demonstrate the impact of vaccine homophily on COVID-19 transmission within and between vaccination groups under conditions of high and low vaccine efficacy, we developed a deterministic model that accounts for heterogeneity in contact patterns to assess the dynamic impact of vaccine homophily on COVID-19 transmission in BC. We analyzed the effects of vaccine homophily under two broad scenarios with low and high vaccine efficacy against infection.

**Model assumptions** The current model was developed to demonstrate the impact of vaccine homophily in scenarios of both low and high vaccine efficacy. We stratified the model population based on the number of COVID-19 vaccine doses received. Interactions within and between groups occur at varying contact rates and preferences, reflecting how individuals interact with others based on their vaccination status. Since vaccine effectiveness against infection is not 100%, breakthrough infections can occur in all groups, with the rate depending on exposure, number of doses received, and vaccine efficacy. We assumed that newly recovered individuals have temporary protection against reinfection (denoted as  $\tau_j$ ). If  $\tau_j = 0$ , it implies no protection against reinfection from a previous infection; if 1, it implies perfect protection after recovery, reducing the model to the classical SIR model with sterilizing immunity after recovery. Immunity wanes at a constant rate  $\sigma_j$ , depending on the number of doses received. Vaccine efficacy  $v_j$  impacts the probability of infection given contact, and we incorporated this into the force of infection  $\lambda_{ij}$  as described below. For the unvaccinated,  $v_{(j=0)} = 0$ , reflecting no protection from vaccination. This vaccine efficacy specifically refers to immunity against infection, as disease was not explicitly modelled, focusing instead on transmission dynamics. The recovery rate  $\gamma$  is assumed to be independent of vaccination status. Without explicitly modelling importations, we assume that a constant number of susceptibles become infected due to importations at a constant rate  $f_j$ , which varies with vaccination status. Finally, our analysis focuses on a short period during which vaccination levels in the population remained stable.

**Model equation** The model equations are:

$$\begin{aligned} \frac{dS_j(t)}{dt} &= - \sum_{i=0}^3 \lambda_{ij}(t) S_j(t) + \sigma_j R_j(t) - f_j \\ \frac{dI_j(t)}{dt} &= \sum_{i=0}^3 \lambda_{ij}(t) (S_j(t) + (1 - \tau_j) R_j(t)) + f_j - \gamma I_j(t) \\ \frac{dR_j(t)}{dt} &= \gamma I_j(t) - \left( (1 - \tau_j) \sum_{i=0}^3 \lambda_{ij}(t) + \sigma_j \right) R_j(t), \end{aligned}$$

where  $i, j = 0, 1, 2, 3$  represent the number of COVID-19 doses an individual has received. Note that  $f_j$  (imported infections per day) represents a constant daily rate of importation of infected vaccinated individuals into the population. Since a minimum of two doses was required for admission into BC during the study period, we assumed no importations from unvaccinated individuals or those with only a single dose into the population ( $f_j = 0$  for  $j < 2$ ). The non-zero values of  $f_j$  are informed by a previous study [35]. The total population size is the sum of each vaccination dose group:  $N = \sum_{j=0}^3 N_j$ , where

$N_j = S_j + I_j + R_j$ . The total population is set at 5.07 million, i.e. the population of BC.

Further details, sources, and descriptions of the variables and parameters used in the model are presented in Table 2.

**Force of infection** The force of infection  $\lambda(t)$  was defined as: (number of contacts per unit time)  $\times$  (probability of disease transmission per contact)  $\times$  (proportion of contacts that are infected). We used the following expression to model the force of infection:

$$\lambda_{ij}(t) = p_{ji}c_j\beta(1 - v_j)(1 - l_j)\frac{I_i}{N_i},$$

where  $\lambda_{ij}(t)$  is the force of infection for transmitting infection from individuals with vaccination status  $i$  to those with vaccination status  $j$ ,  $p_{ji}$  is the proportion of contacts of those with vaccination status  $j$  that are of vaccination status  $i$ ,  $c_j$  is the total number of contacts per day made by individuals with  $j$  doses,  $\beta$  is the probability of infection given contact,  $v_j$  is vaccine efficacy against infection for individuals with  $j$  doses,  $l_j$  represents the extent to which individuals with  $j$  doses comply with physical distancing measures. This is used to capture the heterogeneity in adherence to population-wide physical distancing measures during the study period. During this time, mask mandates and restrictions on indoor gatherings were still in place in BC. Therefore, it's challenging to determine how respondents interpreted the survey question regarding their adherence to public health measures. One possibility is interpreting it purely as wearing a face mask or face covering during indoor contact, while another is viewing it as overall measures, including reducing contacts or avoiding large gatherings. In the former scenario, including  $l_j$  in the force of infection is necessary. In the

latter, the contact matrix inherently accounts for their level of adherence to public health guidelines. As a simplifying model assumption, we presumed that  $l_j$  represents adherence to indoor mask-wearing during contact. In another simulation we assumed that adherence to all public health measures was already accounted for and included a supplementary figure assuming  $l_j=0$  (Fig. S5 in the Supplementary Information). Under these assumptions, infections grow more rapidly, reaching higher levels. However, the direction of the force of infection essentially remains the same as when  $l_j>0$ .  $I_i$  is the number of infected individuals who have had  $i$  doses, and  $N_i$  is the total number of people with  $i$  doses. Parameter values were extracted from the Meta survey data.

**Model validation** We matched model output to reported cases of COVID-19 during the survey period from February 16 to March 3, 2022. We accounted for the underreporting of cases by assuming a constant underascertainment probability ( $r=90\%$ ), i.e. the fraction of cases that were not detected. This is similar to the 92% reported for BC by [37], during the study period. We varied the underascertainment probability around our assumed value and found that the model's predictions are not highly sensitive to underreporting. The model yielded a good fit to the data and provided reasonable initial conditions for subsequent model prediction. The model fit to data is shown in Fig. S2 (See Supplementary Information).

**Model scenarios** We analyzed the impact of vaccine homophily on COVID-19 transmission dynamics under two broad scenarios. First, we assumed that vaccine efficacy in preventing infection is relatively high, representing conditions where a reasonable proportion of the population has recently received a booster vaccination. This

**Table 2** Descriptions of variables and parameters

Variables and parameters	Description and sources
$S_j(t)$	Number of susceptible individuals
$I_j(t)$	Number of infectious individuals
$R_j(t)$	Number of recovered individuals
$\sigma_j$	Waning rate per day for immunity against infection. Set at $1/(183 \text{ days})$
$f_j$	Importation rate (e.g., due to travel). Set at 150 infections per day for those with $\geq 2$ doses and 0 for those with $< 2$ doses [35]. Assuming that travel restrictions are effective
$\tau_j$	Strength of short-term protection from reinfection. At baseline: $\tau_0=0.35, \tau_1=0.65, \tau_2=0.68, \tau_3=0.83$ (Assumed)
$\gamma$	Recovery rate per day. Set at $1/(4 \text{ days})$ [36]
$\beta$	Probability of infection given contact: 0.23 (Fitted)
$p_{ji}$	Proportion of contacts of those with vaccination status $j$ that are of vaccination status $i$ (Estimated from survey data)
$c_j$	The total number of contacts per day made by individuals with $j$ doses (Estimated from survey data)
$l_j$	The level to which those with $j$ doses comply with physical distancing measures (Estimated from survey data)

corresponds to the situation prior to the emergence of the Omicron variant, which showed substantial escape from immunity against infection, or future scenarios where more effective vaccines are available and have been widely used. Second, we modelled a scenario with low vaccine efficacy, representing time periods where immunity has waned significantly or when the dominant variant shows low sensitivity to vaccine protection. We further considered each of the above scenarios with and without homophily. For the former, we used contact-related parameter ( $p_{ji}, c_j l_j$ ) values estimated from the survey data, combining both household and non-household contacts, while in the latter, we calculated a weighted average for each of the parameters to eliminate the impact of vaccine homophily. That is, for the without homophily scenario, the total number of contacts ( $c_j$ ) for each vaccination group and the proportion of contacts individuals make with those in their group and everyone else, as well as the level of adherence to physical distancing measures, are the same for each group regardless of vaccination status.

Below, we describe the parameter values we used for model simulations under various scenarios.

The values of the invariant parameters are presented in Table 2. With vaccine homophily, we used the following parameters as estimated from the survey data: Proportion of contacts of those with vaccination status  $j$  that are of vaccination status  $i$  ( $p_{ji}$ ).

$p_{00} = 0.45, p_{01} = 0.02, p_{02} = 0.39, p_{03} = 0.14, p_{10} = 0.17, p_{11} = 0.08, p_{12} = 0.61, p_{13} = 0.14, p_{20} = 0.11, p_{21} = 0.02, p_{22} = 0.69, p_{23} = 0.18, p_{30} = 0.03, p_{31} = 0.01, p_{32} = 0.32, p_{33} = 0.64$ . For the total number of contacts per day made by individuals with  $j$  doses ( $c_j$ ), we estimated values from the survey data as follows.  $c_0 = 2.11, c_1 = 2.68, c_2 = 2.19, c_3 = 1.73$ , per day. The level of compliance to physical distancing measures ( $l_j$ ) is calculated from the survey data and obtained as follows:  $l_0 = 0.134, l_1 = 0.174, l_2 = 0.349, l_3 = 0.817$ . Furthermore, to generate a scenario of conditions without vaccine homophily, we eliminate the impact of vaccine status homophily in the number of contact per day by finding the average weighted contact  $c'_j$  per day for the population as follows:  $c' = c_1 p_1 + c_2 p_2 + c_3 p_3 = 2.43$  per day where  $p_j$  is the prevalence of each vaccination status group which was calculated from vaccination coverage data as 0.07, 0.03, 0.42, and 0.48 for  $p_0, p_1, p_2$ , and  $p_3$  respectively [33]. A similar procedure was used to calculate the weighted average rate of adherence to public health measures,  $l' = 0.514$ .

For the no-homophily scenario, we set  $p'_{ji}$  (Proportion of contacts of those with vaccination status  $j$  that are

of vaccination status  $i$ ) as the proportion of individuals with  $i$  doses in the population, for all  $j$ . Under conditions of low vaccine efficacy, we used:  $v_0 = 0, v_1 = 0.001, v_2 = 0.02, v_3 = 0.07, \tau_0 = 0.20, \tau_1 = 0.40, \tau_2 = 0.65, \tau_3 = 0.80$ .

Under conditions of high vaccine efficacy and higher temporary protection after recovery, we used:  $v_0 = 0, v_1 = 0.60, v_2 = 0.89, v_3 = 0.93, \tau_0 = 0.80, \tau_1 = 0.85, \tau_2 = 0.90, \tau_3 = 0.97$ .

For these four scenarios, we assumed  $\beta = 0.6$ , which is higher than the estimated value in Table 2, to generate scenarios that would allow for an increase in the number of infections. The values of  $\tau$  and  $v$  are also assumed to represent high and low vaccine efficacy scenarios.

### Ethics review

The study protocol was approved by the Research Ethics Board of Simon Fraser University (Protocol #30000753). All participants provided informed consent before completing the survey,

## Results

### Aim 1. Characterization of vaccine homophily and its relationship to COVID-19 transmission dynamics

Facebook and Instagram advertisements were displayed to 266,894 users. A total of 3659 participants initiated the survey and provided informed consent to participate in the study. After exclusion of responses that were of poor quality or had missing data, the final analytical sample size was 1185.

The unweighted sample was disproportionately White (86.9%), female (58.1%), had higher income ( $\geq \$90,000$ , 58.7%), and had been born in Canada (82.8%) (Table 3). Statistical weights were used to align these factors with the population distribution based on the 2016 Canadian Census Profile for British Columbia, Canada.

Table 4 presents additional descriptive statistics about self-reported COVID-19 diagnosis history and self-reported compliance with provincial COVID-19 prevention guidelines, stratified by self-reported vaccination status. Statistical comparisons across these variables indicate that participants who had received more doses of the COVID-19 vaccine were less likely to report a previous COVID-19 diagnosis ( $p < 0.0001$ ) and were more likely to report higher compliance with provincial COVID-19 prevention guidelines ( $p < 0.0001$ ).

Table 5 presents descriptive statistics for participant-reported descriptions of their household and non-household contacts stratified according to self-reported vaccination status. Statistical comparisons across these variables indicate that participants who had received more doses of the COVID-19 vaccine had networks with higher average numbers of doses ( $p < 0.0001$ ) and had a

**Table 3** Characteristics of the study population, weighted for BC population characteristics based on 2016 census profile

Variable	Weighted
<b>Age</b> , mean (SD)	47.49 (17.63)
<b>Gender</b> , <i>n</i> (%)	
Male	563.9 (47.6)
Female	590.3 (49.8)
Non-binary	30.5 (2.6)
<b>Ethnicity</b> , <i>n</i> (%) <sup>a</sup>	
White	761.2 (64.2)
Asian	175.7 (14.8)
Indigenous	173.9 (14.7)
Other	73.9 (6.2)
<b>Education level</b> , <i>n</i> (%)	
Some high school	112.1 (9.5)
High school diploma or equivalent	418.9 (35.4)
College or trades certificate or diploma	257.8 (21.8)
Some university	107.8 (9.1)
University degree or higher (e.g., Bachelors, Masters, PhD, JD, MD)	288.1 (24.3)
<b>Household income</b> , <i>n</i> (%)	
< \$30,000	227.3 (19.2)
\$30,000–\$59,999	287.5 (24.3)
\$60,000–\$89,999	231.9 (19.6)
≥ \$90,000	438.0 (37.0)
<b>Born in Canada</b> , <i>n</i> (%)	813.3 (69.8)
<b>Health authority</b> , <i>n</i> (%)	
Vancouver Coastal	310.2 (26.2)
Fraser	419.3 (35.4)
Interior	192.8 (16.3)
Island	189.2 (16.0)
Northern	73.2 (6.2)

<sup>a</sup> Due to the small sample sizes in most ethnicity categories, the statistical weight for the ethnicity variable was generated based on a binary variable measuring whether participants were either White or another ethnicity. While suboptimal, this was necessary to achieve convergence of the raking estimation algorithm. Weighted estimates may not round to whole numbers or sum to 100%

greater proportion of network contacts with at least one vaccine dose ( $p < 0.0001$ ). With regard to network vaccine heterogeneity, participants with fewer vaccine doses had more heterogeneous networks according to Blau's heterogeneity index ( $p < 0.0001$ ).

Figure 2 plots the distribution of the vaccine homophily scores, ranging from -1.0 to 1.0, for individuals with 0, 1, 2, or 3 vaccine doses. Kruskal-Wallis test indicates significant differences in homophily between individuals with differing numbers of vaccine doses ( $p < 0.0001$ ). Refer to Fig. S1 in the Supplementary Information for an alternative version of Fig. 2, which

displays homophily scores with a customized vertical axis.

Figure 3 shows boxplots of the participants' contact network sizes stratified according to vaccination status. Participants with more vaccine doses—particularly those with three or more doses—tended to have smaller average network sizes (Spearman's  $r = -0.217$ ,  $p < 0.0001$ ). Pairwise comparisons indicated that those who received zero doses exhibited a significantly different network size compared to those who received three doses,  $p < 0.0001$ . However, no significant difference was found between individuals who received zero doses and those who received either one dose ( $p = 0.958$ ) or two doses ( $p = 0.958$ ). Furthermore, individuals who received three doses showed a significantly different network size compared to those who received one dose ( $p = 0.031$ ) and two doses ( $p < 0.0001$ ). There was no significant difference between individuals who received one dose and those who received two doses ( $p = 0.958$ ).

Finally, we also tested the association between vaccine homophily and network size, finding that higher network size was associated with lower vaccine homophily (Spearman's  $r = -0.114$ ,  $p < 0.0001$ ).

## Aim 2. Demonstration of the impact of vaccine homophily on COVID-19 transmission

Our deterministic mathematical model tested the impact of vaccine homophily on COVID-19 transmission dynamics under conditions of high and low vaccine efficacy. To illustrate these effects, Fig. 4 presents four scenarios describing the intersection of vaccine homophily and vaccine efficacy. Each panel in the figure shows the number of infections from 0 to 60 days and two heat maps characterizing the force of infection at 15 (P1) and 45 (P2) days. Overall, in both low and high vaccine efficacy scenarios, the presence of vaccine homophily contributed to higher levels of epidemic growth. We describe each of the four scenarios in the following section to highlight the interaction between homophily and vaccine efficacy. The model's initial conditions were established to reflect the vaccination uptake levels in British Columbia (BC) as of February 16, 2023. It was then fitted to case report data accumulated during the study period. After calibration, the final values of the state variables in the SIR model (Susceptible, Infected, Recovered) were used to initialize the model for each simulated scenario. Subsequently, the model was simulated for two scenarios, each spanning a 60-day period.



**Table 4** Personal indicators of COVID-19 risk, weighted

	0 Doses n = 234.4	1 Dose n = 20.6	2 Doses n = 243.9	≥ 3 Doses n = 685.7
<b>COVID-19 Diagnosis/Infection, n (%)</b>				
No, and I do not think I have had COVID-19	76.3 (32.0)	10.9 (59.6)	73.2 (30.4)	486.2 (74.0)
No, but I think I have had COVID-19. I just never received a test and/or diagnosis.	86.6 (37.4)	2.8 (15.1)	92.0 (38.1)	112.9 (17.2)
Yes, I have been diagnosed with COVID-19	69.0 (29.8)	4.7 (25.3)	75.9 (31.5)	57.9 (8.8)
<b>Compliance with COVID-19 Guidelines, n (%)</b>				
Not At All	28.8 (12.3)	1.0 (5.0)	14.6 (6.0)	0.4 (0.1)
Not Very Closely	77.6 (33.1)	4.5 (22.0)	59.0 (24.3)	13.3 (2.1)
Somewhat Closely	97.1 (41.4)	9.9 (48.3)	94.6 (39.0)	116.6 (18.0)
Very Closely	30.9 (13.2)	5.1 (24.8)	74.4 (30.7)	515.9 (79.8)

Values may not round to whole numbers or sum to 100% due to missing observations on some variables and statistical weighting

#### **With vaccine homophily and high vaccine efficacy (Fig. 4A)**

In this scenario, the epidemic is primarily driven and sustained by unvaccinated individuals, as indicated by the darker heat maps. While infections increase slowly in the vaccinated groups, epidemic growth is rapid among the unvaccinated. The force of infection in this scenario reveals that transmission is predominantly sustained within the unvaccinated group. There is only a slight impact on the one-dose group due to suboptimal immunity. However, there is minimal effect on the two-dose and three-dose groups, owing to the high vaccine efficacy.

#### **Without vaccine homophily and high vaccine efficacy (Fig. 4B)**

In this scenario, the unvaccinated are mixing randomly with the vaccinated, and the epidemic quickly stabilizes as the unvaccinated obtain secondary benefits from the predominantly fully vaccinated population. Meanwhile, disease importation sustains transmission at a steady state. Furthermore, a substantial number of infections in the unvaccinated group are caused by the vaccinated groups, because contacts are primarily driven by group sizes (see Figure S3 in the Supplementary Information, which describes contact between groups). On the other hand, high vaccine efficacy against infection limits the force of infection from the unvaccinated group to the optimally vaccinated group.

#### **With vaccine homophily and low vaccine efficacy (Fig. 4C)**

In this scenario, due to the low vaccine efficacy, major outbreaks occur among the unvaccinated and suboptimally vaccinated, while infections quickly stabilize among the fully vaccinated, despite the low vaccine efficacy. The epidemic is primarily driven by the unvaccinated and those with two doses. This could be due to

the large size of the two-dose group combined with the relatively low vaccine efficacy. Although each individual has partial protection, the overall population size and low vaccine efficacy result in a substantial force of infection within the two-dose group. Unvaccinated individuals significantly impact those within their group and have some impact on those with one or two doses, but minimal impact on the group with three or more doses, as the mixing pattern limits intergroup contact. A similar pattern is observed in the group with two doses.

#### **Without vaccine homophily and with low vaccine efficacy (Fig. 4D)**

In this scenario, each group affects both itself and other groups equally, although the strength of the impact depends on the vaccination status and the size of the group. Moreover, the fully vaccinated group has a substantial impact on both the unvaccinated and partially vaccinated, as contact is driven by group sizes, coupled with the low vaccine efficacy.

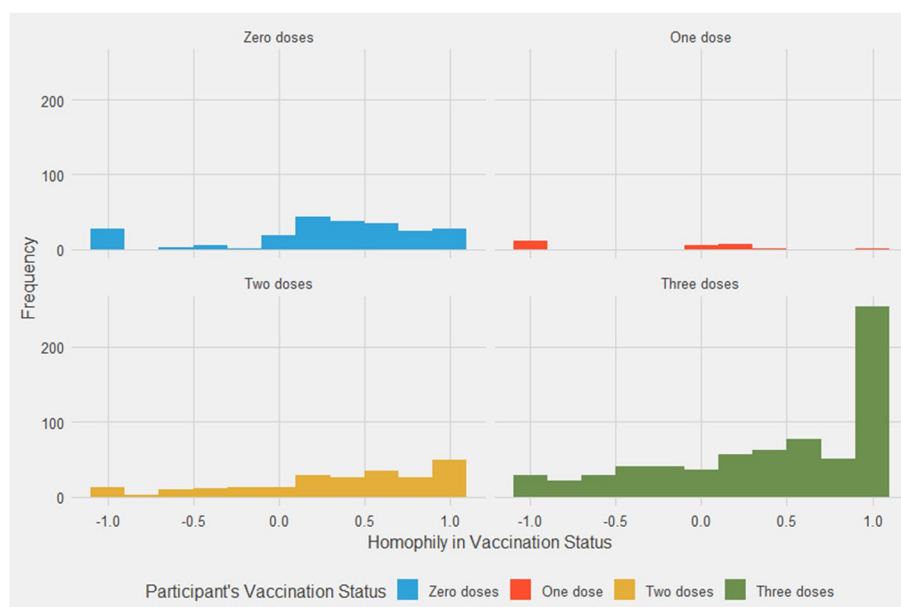
#### **Without vaccine homophily and with vaccine efficacy for Omicron**

In this scenario (See Supplementary Information Figure S4), published vaccine efficacy values against infection with the Omicron variant were used [38]. The two-dose group drives infections in both the one-dose and unvaccinated groups, albeit with a somewhat reduced impact on those with three or more doses. Furthermore, the influence of the unvaccinated group on the two- and three-dose groups is less pronounced than the converse. These dynamics likely stem from the relatively low vaccine efficacy against the Omicron variants and the absence of vaccine status homophily, which allows the prevalence of each vaccination group to drive contact patterns. This scenario was designed to reflect the

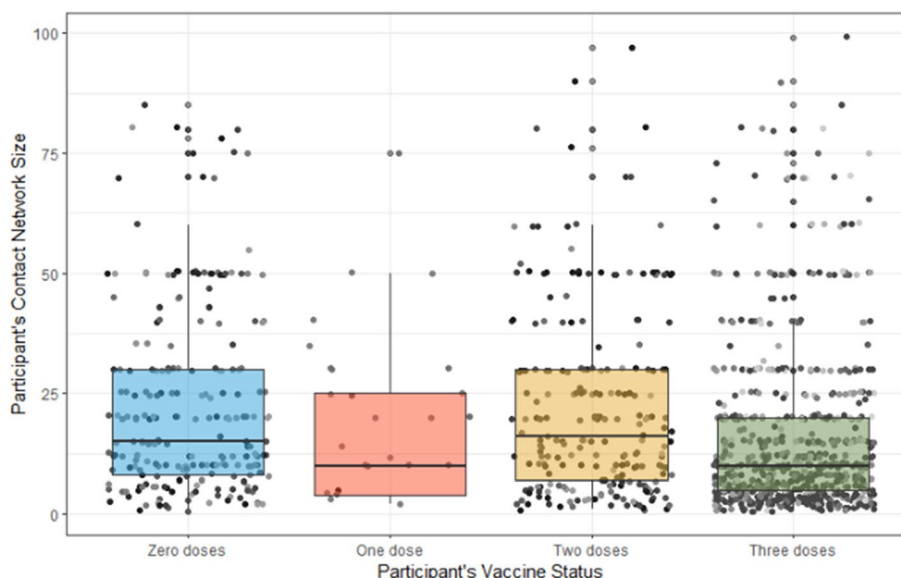
**Table 5** Social network indicators of COVID-19 risk

	0 Doses <i>n</i> = 234.4	1 Dose <i>n</i> = 20.6	2 Doses <i>n</i> = 243.9	≥ 3 Doses <i>n</i> = 685.7
Proportion of overall contacts with prior COVID-19, <i>n</i> (%)				
A few of them (i.e., 0–20%)	94.5 (40.6)	16.6 (80.6)	100.6 (41.2)	493.5 (72.0)
Some of them (i.e., 21–40%)	34.7 (14.9)	0.1 (0.4)	39.6 (16.2)	96.7 (14.1)
Around half of them (i.e., 41–60%)	25.0 (10.7)	0.0 (0.1)	49.1 (20.1)	57.2 (8.3)
Most of them (i.e., 61–80%)	54.2 (23.3)	2.8 (13.7)	34.6 (14.2)	33.9 (4.9)
Nearly all of them (i.e., 80–100%)	24.5 (10.5)	1.1 (5.2)	20.1 (8.2)	4.4 (0.6)
Proportion of overall contacts adhering closely to guidelines, <i>n</i> (%)				
A few of them (i.e., 0–20%)	31.4 (13.4)	3.9 (18.8)	29.9 (12.3)	30.7 (4.5)
Some of them (i.e., 21–40%)	41.8 (17.8)	3.9 (18.8)	46.9 (19.2)	35.8 (5.2)
Around half of them (i.e., 41–60%)	45.2 (19.3)	5.4 (26.4)	39.9 (16.4)	79.2 (11.6)
Most of them (i.e., 61–80%)	93.4 (39.9)	7.4 (35.9)	78.5 (32.2)	271.3 (39.6)
Nearly all of them (i.e., 80–100%)	22.6 (9.7)	0.0 (0.1)	48.5 (19.9)	268.6 (39.2)
Proportion of overall contacts vaccinated, <i>n</i> (%)				
A few of them (i.e., 0–20%)	20.0 (8.5)	2.3 (10.9)	12.8 (5.3)	10.3 (1.5)
Some of them (i.e., 21–40%)	40.8 (17.4)	0.3 (1.7)	14.2 (5.8)	6.0 (0.9)
Around half of them (i.e., 41–60%)	68.4 (29.2)	1.7 (8.4)	42.5 (17.4)	9.3 (1.4)
Most of them (i.e., 61–80%)	68.1 (29.0)	3.8 (18.5)	88.2 (36.2)	128.6 (18.9)
Nearly all of them (i.e., 80–100%)	37.2 (15.9)	12.5 (60.5)	86.2 (35.4)	526.2 (77.3)
Number of non-household contacts, mean (SD)	21.76 (19.24)	13.30 (12.86)	20.29 (16.03)	15.52 (16.38)
Number of non-household contacts with known vaccine status, mean (SD)	10.21 (9.75)	10.40 (10.50)	10.67 (9.82)	10.97 (12.17)
Vaccination status of non-household contacts, mean (SD)				
0 doses	3.73 (6.49)	1.23 (3.23)	1.13 (3.53)	0.35 (1.21)
1 dose	0.17 (0.55)	1.03 (1.91)	0.19 (0.66)	0.18 (0.97)
2 doses	4.41 (6.54)	4.65 (6.37)	7.19 (9.05)	4.34 (9.42)
3 doses	1.90 (4.66)	3.49 (5.07)	2.15 (3.46)	6.10 (8.01)
Household size, mean (SD)	1.85 (1.42)	1.90 (0.79)	2.14 (1.28)	1.76 (1.02)
Vaccination status of household contacts, mean (SD)				
Unknown	0.11 (0.46)	0.06 (0.35)	0.06 (0.31)	0.00 (0.00)
0 doses	1.09 (1.38)	0.56 (1.07)	0.29 (0.92)	0.03 (0.17)
1 dose	0.05 (0.22)	0.40 (0.81)	0.03 (0.18)	0.01 (0.12)
2 doses	0.40 (0.83)	0.32 (0.48)	1.38 (1.11)	0.26 (0.59)
3 doses	0.21 (0.52)	0.56 (0.91)	0.38 (0.60)	1.46 (0.97)
Calculated measures				
Average number of doses among overall contacts, mean (SD)	1.31 (0.79)	1.81 (0.70)	2.00 (0.49)	2.61 (0.45)
Average number of doses among household contacts, mean (SD)	0.83 (1.06)	1.59 (1.11)	1.93 (0.75)	2.81 (0.43)
Average number of doses among non-household contacts, mean (SD)	1.47 (0.89)	1.88 (0.57)	2.02 (0.56)	2.55 (0.55)
% of overall contacts with at least 1 dose, mean (SD)	0.58 (0.31)	0.85 (0.22)	0.89 (0.18)	0.97 (0.12)
% of household contacts with at least 1 dose, mean (SD)	0.36 (0.45)	0.74 (0.41)	0.86 (0.30)	0.99 (0.10)
% of non-household contacts with at least 1 dose, mean (SD)	0.65 (0.36)	0.91 (0.16)	0.90 (0.20)	0.96 (0.15)
Vaccine Homophily, mean, (SD)	0.27 (0.60)	-0.34 (0.63)	0.35 (0.57)	0.42 (0.60)
Network heterogeneity for ≥ 1 dose, mean (SD)	0.31 (0.20)	0.17 (0.18)	0.13 (0.19)	0.04 (0.10)
Household heterogeneity for ≥ 1 dose, mean (SD)	0.11 (0.27)	0.07 (0.20)	0.06 (0.17)	0.01 (0.06)
Non-household heterogeneity for ≥ 1 dose, mean (SD)	0.20 (0.20)	0.11 (0.18)	0.10 (0.17)	0.03 (0.10)
Network heterogeneity, same number of doses in overall contacts, mean (SD)	0.41 (0.20)	0.42 (0.21)	0.33 (0.21)	0.25 (0.22)
Network heterogeneity, same number of doses in household contacts, mean (SD)	0.13 (0.28)	0.08 (0.21)	0.15 (0.24)	0.06 (0.17)
Network heterogeneity, same number of doses in non-household contacts, mean (SD)	0.30 (0.23)	0.34 (0.24)	0.24 (0.22)	0.22 (0.22)

Values may not round to whole numbers or sum to 100% due to missing observations on some variables and statistical weighting



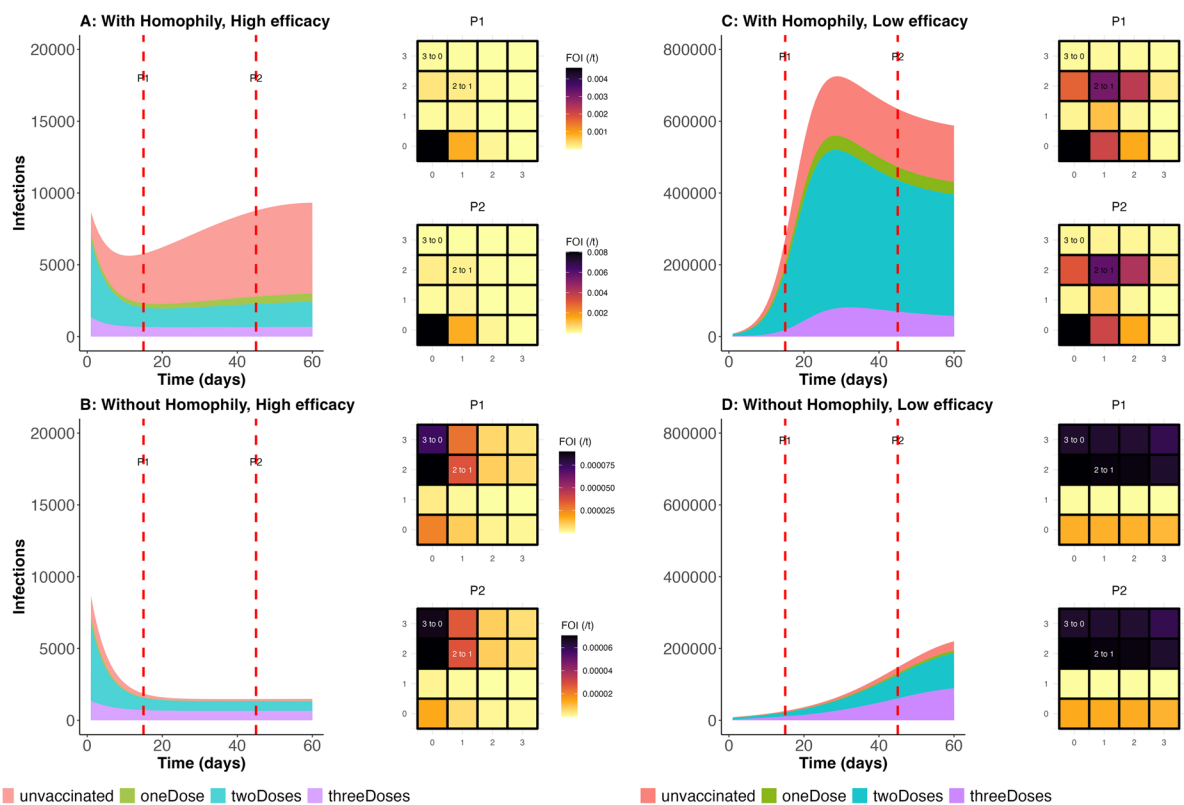
**Fig. 2** Distribution of vaccine homophily scores, by participant vaccination status



**Fig. 3** Homophily and Contact Network Size by Quantile. Figure Note: Each box in the boxplot spans from the first quartile (Q1) to the third quartile (Q3) of the data, representing the middle 50% of the data points for each vaccine status category. The black horizontal bar within each box represents the median of the data. In the context of this boxplot, it indicates the median network size for participants within each vaccine status category. The vertical lines, known as whiskers, extend from the upper and lower edges of the box to the highest and lowest values within a specific range. This range is typically defined as 1.5 times the interquartile range (IQR) above the upper quartile and below the lower quartile. Values outside this range are considered outliers and are not included in the whiskers. In this plot, the whiskers represent the spread of the network size data points, excluding outliers, for each vaccine status category

situation in British Columbia where homophily-enhancing measures, such as vaccination passports, were relaxed shortly after the study period. This may have led to a

reduction in homophily. Given that the Omicron variant was the dominant strain in BC at the time, we focused on vaccine efficacy against the Omicron variant.



**Fig. 4** Number of Infections and Force of Infection: Assessment of the Impact of Homophily Under Scenarios of Low and High Vaccine Efficacy. **A** Number of infections under a scenario with vaccine homophily and high vaccine efficacy. The trajectory is colour-coded by vaccination status. Heat maps P1 and P2 show the force of infection at 15 and 45 days, respectively. **B** Number of infections under a scenario without vaccine homophily and with high vaccine efficacy. **C** The number of infections per day for various vaccination groups under a scenario with vaccine homophily and low vaccine efficacy. **D** Daily number of infections according to vaccination status under a scenario without homophily and low vaccine efficacy. The following parameter values were used under conditions with vaccine homophily. The horizontal and vertical axes on the heat maps represent vaccination status. The numbers within the heat maps indicate which group is transmitting infection to which other group: “2 to 1” indicates that individuals with 2 doses transmit to those with only 1 dose on that grid, and “3 to 0” indicates that those with 3 doses transmit to unvaccinated individuals on that grid

## Interpretation

### Primary findings

This study was performed to characterize observable patterns of vaccine homophily and examine the impact of vaccine homophily on COVID-19 transmission both within and between vaccination status groups under conditions of high and low vaccine efficacy. The results indicated the occurrence of vaccine homophily, with a large proportion of the participants’ network contacts having the same number of vaccine doses as the participants themselves. Even adjusting for the population prevalence of each vaccine dose, those with zero, two, or three doses reported greater than expected levels of homophily. Similarly, the average number of doses received by household and non-household contacts was highest among those with  $\geq 3$  doses and lowest among those with 0 doses, demonstrating a higher prevalence of vaccination within the networks of vaccinated individuals relative

to unvaccinated individuals. Those who were unvaccinated also had more diverse social networks with regard to vaccine status, were more likely to report previous COVID-19 infection and had larger social network sizes. Mathematical models demonstrated that these dynamics contribute to elevated transmission overall under conditions of high vaccine efficacy, and transmission is driven primarily by unvaccinated individuals infecting other unvaccinated individuals. Under conditions of low vaccine efficacy, within-group transmission among unvaccinated individuals remains high, but there is also considerable impact of unvaccinated transmission on suboptimally vaccinated individuals. Those with suboptimal protection (e.g., two doses) also experience considerable within-group transmission due to high contact rates with other suboptimally protected contacts within their network.

One factor contributing to these patterns is a higher level of observed vaccine homophily among household



contacts compared to non-household contacts. Indeed, among unvaccinated participants, only 39% of household contacts had one or more doses of the COVID-19 vaccine, compared to 68% of non-household contacts. We also found that vaccine homophily appears to decrease as social network size increases, suggesting that smaller networks are more similar to one another than larger networks. This is consistent with the empirical expectation that people tend to associate with people like themselves and are more different from those who are more socially distant [11].

To our knowledge, there have been few reports of empirically measured COVID-19 vaccine homophily. However, our findings that vaccine homophily has important implications for understanding the transmission of COVID-19 were consistent with previous modeling studies [12, 13, 15]. In situating our findings within these previous studies, it is important to note that the impact of vaccine homophily differs according to the level of vaccine efficacy. Under conditions of high vaccine efficacy, transmission occurs largely among unvaccinated individuals. Meanwhile, contact patterns put suboptimally vaccinated individuals at risk of infection under conditions of low vaccine efficacy. In contrast, fully vaccinated individuals experience a lower risk of infection. Furthermore, contrary to some narratives that blame unvaccinated individuals for driving the epidemic under conditions of low vaccine efficacy, we found that the force of infection is substantially influenced by group sizes in ‘without homophily’ scenarios. Additionally, a sizeable portion of the force of infection among unvaccinated individuals originates from outside their group, a trend that becomes more pronounced when vaccine efficacy is low. With vaccine homophily, unvaccinated individuals pose significantly greater risk to other unvaccinated individuals than to other groups. The impact of unvaccinated individuals on fully vaccinated individuals is considerable only when there is low vaccine homophily, and vaccine efficacy is low. For all the scenarios we considered, the impact of homophily is amplified by increased probability of infection per contact.

The overrepresentation of the unvaccinated in the total number of infections (Fig. 4A) is similar to findings from Canada, based on case-level vaccine history data. Among individuals aged 5 years and older, the unvaccinated constituted approximately 30% of the total reported cases since the onset of the vaccination rollout, as of June 10, 2022. As the vaccination rollout progresses, the limited testing capacity has resulted in the targeted testing of the high-risk population for severe disease, which coincides with the group prioritized during the vaccination rollout. Consequently, this bias in the case report data indicates that reported case data by

vaccination status may not accurately reflect the distribution of infections by vaccination status at the population level [33]. For example, in BC, the unadjusted data indicated that the unvaccinated accounted for 14.2% of the total cases, whereas the age-adjusted cases per 100,000 population in the province showed that unvaccinated groups accounted for 58% in March 2022 [39]. This finding is consistent with the initial conditions of our model at the beginning of March 2022.

Taken together, our findings are worrisome, particularly when considering the risk for transmission within households, which are known to account for a significant proportion of COVID-19 infections [40–42]. Furthermore, the high risk of infection among unvaccinated individuals, even with an effective vaccine available, underscores the need for vaccine-status-specific COVID-19 prevention measures. These measures are crucial as unvaccinated individuals can significantly contribute to hospitalizations, even when they are a minority [43], which might raise important questions about health equity. Such measures may include mask mandates, physical distancing rules, and proof of vaccination requirements. Given the group transmission dynamics that arise due to household and non-household contact networks, it is important to engage these populations to address vaccine hesitancy [44–50]. This will likely require community-based and culturally aware public health interventions that can help reduce vaccine hesitancy. Indeed, rather than viewing unvaccinated individuals as a threat to public health, it should be taken as an opportunity to educate and work with these individuals to address their concerns, particularly given the skepticism that may be associated with the emergency use authorizations that have allowed the rapid rollout of COVID-19 vaccines [46, 47].

### Limitations

This study had some limitations that should be taken into consideration when interpreting our findings. First, we note that our findings are relevant to the promotion of vaccines across the population and emphasize the importance of continued vaccine research and efforts to provide ongoing protection as vaccine-induced immunity wanes. However, our data are from a period in which individuals were receiving third doses and facing the rising prevalence of the Omicron variant. Therefore, our results should not be read as predictive scenarios. Rather, they should be interpreted in the context of a pandemic-related mass-vaccination effort, during which there was uneven uptake of vaccines across social networks due to a variety of factors within and outside the control of individuals.

Second, we note that our survey utilized an online, opt-in convenience sampling methodology to study the effects

of interest. Online sampling is now a widespread methodology, particularly since the decline in reliability of other opt-in sampling strategies such as random digit dialing methods. Point estimates from this study are therefore likely to be non-representative and may be biased. However, we note that studies show that epidemiological and behavioural estimates from web and telephone surveys are typically comparable, and that online samples may have advantages to other survey methods (e.g., reduced favourable reporting; [51]). This is because the direction of bias may be random. While population weights may partially adjust for this issue, the direction and magnitude of potential biases are unknown. Replication in a population-based sample is warranted. Furthermore, our weighting method does not make any assumptions regarding the statistical relationship between weighting variables (i.e., each variable included in the weighting variable is iteratively fit until target weights are met, without trying to match marginal weights).

Third, it is important to acknowledge that our sample size was relatively modest. Replicating our findings in a larger sample could offer more robust evidence and enhance the accuracy of our measurements. However, we must acknowledge that replicating the study will present significant challenges, particularly given the current stage of the pandemic. Tracking the vaccination statuses of individuals within contact networks may prove to be a daunting task.

Fourth, self-reported data may be unreliable, particularly estimates regarding characteristics of participants' social networks. People may be overly confident in estimating their network's vaccination status, guideline compliance, and vaccine history of their social network contacts, which may result in a systematic bias toward the hypothesis that vaccine status homophily exists. Fifth, we do not intend to imply causality in describing any of the relationships between vaccine status and vaccine homophily. Further qualitative and quantitative studies are needed to understand the processes that give rise to vaccine homophily and how best to respond to these network characteristics.

Our modelling assumptions did not consider vaccine efficacy against infectiousness, except that if infection itself is prevented, so is infectiousness. We made this choice partly because the evidence for a reduction in infectiousness due to vaccination is still emerging [52, 53]. Additionally, during the study period, almost everyone in the population had been exposed to the infection, which suggests that the majority might exhibit some reduction in infectiousness due to vaccination or natural immunity. This generally impacts the probability of infection given a contact. A recent study suggests a 22% reduction in infectiousness post-vaccination [52]. We found that our results

remain robust with a reduction in infectiousness around that value, which slightly impacts the total infection. Detailed modelling of the interaction between acquired and natural immunities, and their efficacy against infectiousness, would require more data and details, which are beyond the scope of the current study.

## Conclusion

The present study identified evidence of homophily in COVID-19 vaccine uptake. Unvaccinated individuals are more likely to have unvaccinated network contacts, conditions that create increased risk of COVID-19 transmission among unvaccinated individuals. Nevertheless, vaccine homophily varies considerably, and further research is needed to understand the factors that shape vaccine homophily within social networks. Vaccine status-specific prevention guidelines may help to mitigate the risks to communities posed by the unique risk profiles of unvaccinated individuals.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12889-024-17957-5>.

### Additional file 1.

## Acknowledgements

PIPPS receives funding from the British Columbia Ministry of Health.

## Authors' contributions

Study design (KC, CC, EBA), Data collection (KC), Data analysis (KC, EBA, CC), Model conceptualization (EBA, CC), Simulation (EBA), Initial Draft (KC, EB), Visualization (EBA, KC), Funding acquisition (KC, CC), Supervision (CC), Resources (KC, CC) Writing, Editing and Proofreading (EBA, KC, CC).

## Funding

This study was supported with funding from the Canadian Network for Modeling Infectious Diseases (CANMOD) and the Natural Sciences and Engineering Research Council of Canada (NSERC). KGC was supported with a Michael Smith Health Research BC Scholar Award.

## Availability of data and materials

Data sets generated during the current study are available from Dr. Kiffer Card.

## Declarations

### Ethics approval and consent to participate

The study protocol was reviewed and approved by the Research Ethics Board at Simon Fraser University (Protocol #30000753). All participants and their legal guardian gave Informed consent for participation and all methods were carried out in accordance with relevant guidelines and regulations.

### Consent for publication

Not Applicable.

### Competing interests

The authors declare no competing interests.

Received: 5 June 2023 Accepted: 1 February 2024

Published online: 14 February 2024

# References

1. Mousa A, Winskill P, Watson OJ, Ratmann O, Monod M, Ajelli M, et al. Social contact patterns and implications for infectious disease transmission – a systematic review and meta-analysis of contact surveys. *eLife*. 2021;10:e70294.
2. Alimohamadi Y, Tola HH, Abbasi-Ghahramanloo A, Janani M, Sepandi M. Case fatality rate of COVID-19: a systematic review and meta-analysis. *J Prev Med Hyg*. 2021;62(2):E311.
3. Zheng C, Shao W, Chen X, Zhang B, Wang G, Zhang W. Real-world effectiveness of COVID-19 vaccines: a literature review and meta-analysis. *Int J Infect Dis*. 2022;114:252–60.
4. Feikin DR, Higdon MM, Abu-Raddad LJ, Andrews N, Araos R, Goldberg Y, et al. Duration of effectiveness of vaccines against SARS-CoV-2 infection and COVID-19 disease: results of a systematic review and meta-regression. *Lancet*. 2022;399(10328):924–44.
5. Chen X, Chen Z, Azman AS, Deng X, Sun R, Zhao Z, et al. Serological evidence of human infection with SARS-CoV-2: a systematic review and meta-analysis. *Lancet Glob Health*. 2021;9(5):e598–609.
6. Liu Q, Qin C, Liu M, Liu J. Effectiveness and safety of SARS-CoV-2 vaccine in real-world studies: a systematic review and meta-analysis. *Infect Dis Poverty*. 2021;10(1):132.
7. World Health Organization. Ten health issues WHO will tackle this year. Available from: <https://www.who.int/news-room/spotlight/ten-threats-to-global-health-in-2019>. Cited 2022 Oct 4.
8. Larson HJ, Jarrett C, Eckersberger E, Smith DMD, Paterson P. Understanding vaccine hesitancy around vaccines and vaccination from a global perspective: a systematic review of published literature, 2007–2012. *Vaccine*. 2014;32(19):2150–9.
9. Konstantinou P, Georgiou K, Kumar N, Kyprianidou M, Nicolaides C, Karekla M, et al. Transmission of vaccination attitudes and uptake based on social contagion theory: a scoping review. *Vaccines*. 2021;9(6):607.
10. Lin C, Tu P, Beitsch LM. Confidence and receptivity for COVID-19 vaccines: a rapid systematic review. *Vaccines*. 2020;9(1):16.
11. McPherson M, Smith-Lovin L, Cook JM. Birds of a feather: homophily in social networks. *Annu Rev Sociol*. 2001;27(1):415–44.
12. Kadelka C, McCombs A. Effect of homophily and correlation of beliefs on COVID-19 and general infectious disease outbreaks. *PLoS One*. 2021;16(12):e0260973.
13. Hiraoka T, Rizi AK, Kivelä M, Saramäki J. Herd immunity and epidemic size in networks with vaccination homophily. *Phys Rev E*. 2022;105(5):L052301.
14. Aral S, Muchnik L, Sundararajan A. Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proc Natl Acad Sci*. 2009;106(51):21544–9.
15. Fisman DN, Amoako A, Tuite AR. Impact of population mixing between vaccinated and unvaccinated subpopulations on infectious disease dynamics: implications for SARS-CoV-2 transmission. *Can Med Assoc J*. 2022;194(16):E573–80.
16. Ankomah AA, Moa A, Chughtai AA. The long road of pandemic vaccine development to rollout: a systematic review on the lessons learnt from the 2009 H1N1 influenza pandemic. *Am J Infect Control*. 2022;50(7):735–42.
17. Reñosa MDC, Landicho J, Wachinger J, Dalglish SL, Bärnighausen K, Bärnighausen T, et al. Nudging toward vaccination: a systematic review. *BMJ Glob Health*. 2021;6(9):e006237.
18. Wang Y, Liu Y. Multilevel determinants of COVID-19 vaccination hesitancy in the United States: a rapid systematic review. *Prev Med Rep*. 2022;25:101673.
19. Wouters OJ, Shadlen KC, Salcher-Konrad M, Pollard AJ, Larson HJ, Teerawattananon Y, et al. Challenges in ensuring global access to COVID-19 vaccines: production, affordability, allocation, and deployment. *Lancet*. 2021;397(10278):1023–34.
20. Johns Hopkins University Center for Systems Science and Engineering. COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. 2022;17(1):341. Available from: <http://trialsjournal.biomedcentral.com/articles/10.1186/s13063-016-1473-3>.
21. Team RD. R: A language and environment for statistical computing. R Foundation for Statistical Computing. 2010. <https://www.R-project.org/>.
22. Wickham H, RStudio. Tidyverse: easily install and load the “Tidyverse” (1.3.1). 2021. [Computer Software]. Available at: <https://CRAN.R-project.org/package=tidyverse>.
23. Mukaka M, White SA, Terlouw DJ, Mwapasa V, Kalilani-Phiri L, Faragher EB. Is using multiple imputation better than complete case analysis for estimating a prevalence (risk) difference in randomized controlled trials when binary outcome observations are missing? *Trials*. 2016;17(1):1–2.
24. Mercer A, Lau A, Kennedy C. 1. How different weighting methods work. *Pew Res Cent Methods*. 2018;86(415):779–84.
25. Norman P. Putting iterative proportional fitting on the researchers desk. Working paper 99/03, 1999. Leeds: University of Leeds. 1999.
26. Skinner CJ. On the efficiency of raking ratio estimation for multiple frame surveys. *J Am Stat Assoc*. 1991;86(415):779–84. <https://doi.org/10.1080/01621459.1991.10475109>.
27. Pasek J. anesrake: ANES raking implementation. 2018. Available from: <https://CRAN.R-project.org/package=anesrake>.
28. Government of Canada SC. Census of population. 2001. Available from: <https://www12.statcan.gc.ca/census-recensement/index-eng.cfm>. Cited 2022 Oct 4.
29. Lumley T. survey: analysis of complex survey samples. 2021. Available from: <https://CRAN.R-project.org/package=survey>.
30. Barnier J, Briatte F, Larmarange J. questionr: functions to make surveys processing easier. 2022. Available from: <https://CRAN.R-project.org/package=questionr>.
31. Wickham H, Chang W, Henry L, Pedersen TL, Takahashi K, Wilke C, et al. ggplot2: create elegant data visualisations using the grammar of graphics. 2021. Available from: <https://CRAN.R-project.org/package=ggplot2>.
32. Coleman J. Relational analysis: the study of social organizations with survey methods. *Hum Organ*. 1958;17(4):28–36.
33. BC Government news release. COVID-19 pandemic update: daily update of COVID-19. Ministry of Health Communications; 2022. Available: <https://news.gov.bc.ca/releases/2022HLTH0102-000434>.
34. Blau PM. Inequality and heterogeneity: a primitive theory of social structure. New York: Free Press; 1977.
35. Arino J, Portet S, Bajoux N, Ciupeanu A. Investigation of global and local COVID-19 importation risks. Report to the Public Health Risk Science division of the Public Health Agency of Canada. 2020.
36. Lima CM. Information about the new coronavirus disease (COVID-19). *Radiol Bras*. 2020;53:V-I.
37. Skowronski DM, Kaweski SE, Irvine MA, Kim S, Chuang ES, Sabaiduc S, Sekirov I, et al. Serial cross-sectional estimation of vaccine- and infection-induced SARS-CoV-2 seroprevalence in British Columbia, Canada. *Cmaj*. 2022;194(47):E1599–1609.
38. Andrews N, Stowe J, Kirsebom F, Toffa S, Rickeard T, Gallagher E, et al. Covid-19 vaccine effectiveness against the omicron (B.1.1.529) variant. *N Engl J Med*. 2022;386(16):1532–46.
39. Canada P.H.A.O. COVID-19 daily epidemiology update. 2021.
40. Madewell ZJ, Yang Y, Longini IM, Halloran ME, Dean NE. Household transmission of SARS-CoV-2: a systematic review and meta-analysis. *JAMA Netw Open*. 2020;3(12):e2031756.
41. Madewell ZJ, Yang Y, Longini IM, Halloran ME, Dean NE. Factors associated with household transmission of SARS-CoV-2: an updated systematic review and meta-analysis. *JAMA Netw Open*. 2021;4(8):e2122240.
42. Zhu Y, Bloxham CJ, Hulme KD, Sinclair JE, Tong ZWM, Steele LE, et al. A meta-analysis on the role of children in severe acute respiratory syndrome coronavirus 2 in household transmission clusters. *J Clin Infect Dis*. 2021;72(12):e1146–53.
43. Keeling MJ, Moore S, Penman BS, Hill EM. The impacts of SARS-CoV-2 vaccine dose separation and targeting on the COVID-19 epidemic in England. *Nat Commun*. 2023;14(1):740.
44. Shen SC, Dubey V. Addressing vaccine hesitancy: clinical guidance for primary care physicians working with parents. *Can Fam Physician*. 2019;65(3):175–81.
45. McClure CC, Cataldi JR, O’Leary ST. Vaccine hesitancy: where we are and where we are going. *Clin Ther*. 2017;39(8):1550–62.
46. Cataldi JR, Kerns ME, O’Leary ST. Evidence-based strategies to increase vaccination uptake: a review. *Curr Opin Pediatr*. 2020;32(1):151–9.
47. Dubé E, Gagnon D, MacDonald N, Bocquier A, Peretti-Watel P, Verger P. Underlying factors impacting vaccine hesitancy in high income countries: a review of qualitative studies. *Expert Rev Vaccines*. 2018;17(11):989–1004.
48. Dubé É, Ward JK, Verger P, MacDonald NE. Vaccine hesitancy, acceptance, and anti-vaccination: trends and future prospects for public health. *Annu Rev Public Health*. 2021;42(1):175–91.

49. Hudson A, Montelpare WJ. Predictors of vaccine hesitancy: implications for COVID-19 public health messaging. *Int J Environ Res Public Health*. 2021;18(15):8054.
50. Limaye RJ, Opel DJ, Dempsey A, Ellingson M, Spina C, Omer SB, Dudley MZ, Salmon DA, O'Leary ST. Communicating with vaccine-hesitant parents: a narrative review. *Acad Pediatr*. 2021;21(4):S24–9.
51. Jones MK, Calzavara L, Allman D, Worthington CA, Tyndall M, Iveniuk J. A comparison of web and telephone responses from a national HIV and AIDS survey. *JMIR Public Health Surveill*. 2016;2(2):e5184.
52. Tan ST, Kwan AT, Rodríguez-Barraquer I, Singer BJ, Park HJ, Lewnard JA, Lo NC, et al. Infectiousness of SARS-CoV-2 breakthrough infections and reinfections during the Omicron wave. *Nat Med*. 2023;29(2):358–65.
53. Mongin D, Bürgisser N, Laurie G, Schimmel G, Vu DL, Cullati S, Courvoisier DS, et al. Effect of SARS-CoV-2 prior infection and mRNA vaccination on contagiousness and susceptibility to infection. *Nat Commun*. 2023;14(1):5452.

# Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



## Terms and Conditions

Springer Nature journal content, brought to you courtesy of Springer Nature Customer Service Center GmbH (“Springer Nature”).

Springer Nature supports a reasonable amount of sharing of research papers by authors, subscribers and authorised users (“Users”), for small-scale personal, non-commercial use provided that all copyright, trade and service marks and other proprietary notices are maintained. By accessing, sharing, receiving or otherwise using the Springer Nature journal content you agree to these terms of use (“Terms”). For these purposes, Springer Nature considers academic use (by researchers and students) to be non-commercial.

These Terms are supplementary and will apply in addition to any applicable website terms and conditions, a relevant site licence or a personal subscription. These Terms will prevail over any conflict or ambiguity with regards to the relevant terms, a site licence or a personal subscription (to the extent of the conflict or ambiguity only). For Creative Commons-licensed articles, the terms of the Creative Commons license used will apply.

We collect and use personal data to provide access to the Springer Nature journal content. We may also use these personal data internally within ResearchGate and Springer Nature and as agreed share it, in an anonymised way, for purposes of tracking, analysis and reporting. We will not otherwise disclose your personal data outside the ResearchGate or the Springer Nature group of companies unless we have your permission as detailed in the Privacy Policy.

While Users may use the Springer Nature journal content for small scale, personal non-commercial use, it is important to note that Users may not:

1. use such content for the purpose of providing other users with access on a regular or large scale basis or as a means to circumvent access control;
2. use such content where to do so would be considered a criminal or statutory offence in any jurisdiction, or gives rise to civil liability, or is otherwise unlawful;
3. falsely or misleadingly imply or suggest endorsement, approval, sponsorship, or association unless explicitly agreed to by Springer Nature in writing;
4. use bots or other automated methods to access the content or redirect messages
5. override any security feature or exclusionary protocol; or
6. share the content in order to create substitute for Springer Nature products or services or a systematic database of Springer Nature journal content.

In line with the restriction against commercial use, Springer Nature does not permit the creation of a product or service that creates revenue, royalties, rent or income from our content or its inclusion as part of a paid for service or for other commercial gain. Springer Nature journal content cannot be used for inter-library loans and librarians may not upload Springer Nature journal content on a large scale into their, or any other, institutional repository.

These terms of use are reviewed regularly and may be amended at any time. Springer Nature is not obligated to publish any information or content on this website and may remove it or features or functionality at our sole discretion, at any time with or without notice. Springer Nature may revoke this licence to you at any time and remove access to any copies of the Springer Nature journal content which have been saved.

To the fullest extent permitted by law, Springer Nature makes no warranties, representations or guarantees to Users, either express or implied with respect to the Springer nature journal content and all parties disclaim and waive any implied warranties or warranties imposed by law, including merchantability or fitness for any particular purpose.

Please note that these rights do not automatically extend to content, data or other material published by Springer Nature that may be licensed from third parties.

If you would like to use or distribute our Springer Nature journal content to a wider audience or on a regular basis or in any other manner not expressly permitted by these Terms, please contact Springer Nature at

[onlineservice@springernature.com](mailto:onlineservice@springernature.com)