



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



ELSEVIER

Contents lists available at ScienceDirect

Annals of Epidemiology

journal homepage: [sciencedirect.com/journal/annals-of-epidemiology](https://www.sciencedirect.com/journal/annals-of-epidemiology)

Original article

## A computer modeling method to analyze rideshare data for the surveillance of novel strains of SARS-CoV-2

Conrad W. Safranek, BS<sup>a,b</sup>, David Scheinker, PhD<sup>b,c,d,\*</sup>

<sup>a</sup> Department of Biology, Computational Biology, Stanford University, CA

<sup>b</sup> Department of Management Science and Engineering, Stanford University School of Engineering, CA

<sup>c</sup> Department of Pediatrics, Stanford University School of Medicine, CA

<sup>d</sup> Clinical Excellence Research Center, Stanford University School of Medicine, CA

### ARTICLE INFO

#### Article history:

Received 5 December 2021

Revised 15 July 2022

Accepted 29 August 2022

Available online xxx

#### Keywords:

COVID-19

SARS-CoV-2

Epidemiological monitoring

Proof of concept study

Computer simulation

Epidemiology

### ABSTRACT

**Purpose:** No method is available to systematically study SARS-CoV-2 transmission dynamics using the data that rideshare companies share with government agencies. We developed a proof-of-concept method for the analysis of SARS-CoV-2 transmissions between rideshare passengers and drivers.

**Method:** To assess whether this method could enable hypothesis testing about SARS-CoV-2, we repeated ten 200-day agent-based simulations of SARS-CoV-2 propagation within the Los Angeles County rideshare network. Assuming data access for 25% of infections, we estimated an epidemiologist's ability to analyze the observable infection patterns to correctly identify a baseline *viral variant A*, as opposed to *viral variant A with mask use* (50% reduction in viral particle exchange), or a more infectious *viral variant B* (300% higher cumulative viral load).

**Results:** Simulations had an average of 190,387 potentially infectious rideshare interactions, resulting in 409 average diagnosed infections. Comparison of the number of observed and expected passenger-to-driver infections under each hypothesis demonstrated our method's ability to consistently discern large infectivity differences (*viral variant A* vs. *viral variant B*) given partial data from one large city, and to discern smaller infectivity differences (*viral variant A* vs. *viral variant A with masks*) given partial data aggregated across multiple cities.

**Conclusions:** This novel statistical method suggests that, for the present and subsequent pandemics, government-facilitated analysis of rideshare data combined with diagnosis records may augment efforts to better understand viral transmission dynamics and to measure changes in infectivity associated with nonpharmaceutical interventions and emergent viral strains.

© 2022 Elsevier Inc. All rights reserved.

### Introduction

The emergence of novel, partially vaccine-resistant strains of SARS-CoV-2 poses a serious threat to ongoing public health efforts [1,2]. Policies to contain the spread of COVID-19 rely on understanding what such strains mean for the risk of transmission and how they may impact optimal strategies to contact trace, test, quarantine, and vaccinate [3,4]. Important factors in our understanding of viral transmission include the effectiveness of nonphar-

maceutical interventions (NPIs) such as facemasks [4,5]; the prevalence of superspreaders in the population [6]; the increased infectivity of new viral strains [7,8]; and the ongoing extent to which vaccines reduce infectivity given emergent strains and waning immunity [2,9]. The emergence and rapid spread of the Delta and Omicron variants have highlighted the limitations of the current infrastructure, even after data-sharing improvements during the first year of the pandemic, to offer only retrospective identification of emergent viral strains and their transmission dynamics [1].

The study of the transmission of SARS-CoV-2 has been constrained primarily to specialized settings, such as laboratory experimentation, by examining infections in households, trains, hospitals, and military facilities, and through placebo-controlled vaccine clinical trials [5,10–14]. The usefulness of these data is limited by the difficulty of translating the results to other settings and the emergence of new viral strains. Practical and ethical limi-

**Abbreviations:** NPI, nonpharmaceutical intervention; RIDE, Rideshare Infection Detection; LA, Los Angeles County.

\* Corresponding author. Stanford University School of Engineering, Department of Management Science and Engineering, 475 Via Ortega, CA. Tel.: 94305-4121; fax: 650-683-5623.

E-mail address: [dscheink@stanford.edu](mailto:dscheink@stanford.edu) (D. Scheinker).

<https://doi.org/10.1016/j.annepidem.2022.08.051>

1047-2797/© 2022 Elsevier Inc. All rights reserved.

Please cite this article as: C.W. Safranek and D. Scheinker, A computer modeling method to analyze rideshare data for the surveillance of novel strains of SARS-CoV-2, *Annals of Epidemiology*, <https://doi.org/10.1016/j.annepidem.2022.08.051>

tations to clinical trials and the challenges of drawing causal inference from observational data limit our ongoing understanding of real-world transmission dynamics for COVID-19. For example, more than 9 months after the emergence of the Delta variant with over 50,000 diagnosed “breakthrough” infections in vaccinated individuals [15], the research community still lacked consensus on whether these infected vaccinated individuals had a reduced level of transmissibility relative to infected unvaccinated individuals [16,17], or whether they were equally contagious [18,19].

Analysis of data collected via digital disease surveillance may help elucidate evolving SARS-CoV-2 transmission dynamics [20]. Mobile phone applications for contact tracing (e.g., Google Apple Exposure Notification technology) gather extensive real-world data, but with limited information about the highly variable settings in which these interactions take place [21]. In contrast, in-vehicle transmission dynamics can be well characterized and are relatively consistent. Aerosolized viral dynamics within cars have been studied [22], and rideshare companies already share data with government health agencies for contact tracing [23], but to the best of our knowledge no statistical method is available for the use of these data for the systematic study of the transmission dynamics of COVID-19.

The interactions between passengers and drivers facilitated by rideshare platforms such as Uber and Lyft are, essentially, a series of partially controlled, standardized, pseudo-random experiments of SARS-CoV-2 transmissions. Rideshare trips are often the only connection between individuals; are governed by mask-wearing policies implemented on specific dates; occur in a relatively controlled environment with fairly consistent spatial dynamics; and are sporadic with respect to time, location, and duration. Previous studies have associated rideshare trips with up to a 10-fold increase in COVID-19 infection risk [24], suggesting that infection patterns among rideshare trips may be substantial. Furthermore, rideshare location and time data are stored in a machine-readable format that could be linked to data such as diagnoses or self-reported symptoms (anonymized in accordance with local legislation). Thus, rideshare trip data may facilitate the automated identification of the emergence of new viral strains and the study of their transmission dynamics, with and without the presence of NPIs and vaccination, and based on passenger and driver characteristics.

The unknown and potentially low “signal-to-noise” ratio of detected-to-undetected viral infections in a rideshare network presents a challenge for the usefulness of a statistical method to analyze rideshare data merged with diagnosis data [25]. The potential utility of such a method depends on its performance accurately detecting rideshare-acquired infections while accounting for undiagnosed infections as well as for “false-positive” rideshare interactions that appear to have resulted in transmission even though the potential infectee contracted the virus elsewhere.

We develop Rideshare Infection Detection (RIDE), a probabilistic method and model of rideshare transmissions designed to test hypotheses about the emergence of novel strains of SARS-CoV-2 and their transmission dynamics. Since aggregated rideshare and diagnosis data are not currently publically available, we simulate viral transmission data in a hypothetical rideshare network based on empirical data from a large US city. We use these simulated data to test hypotheses about transmission dynamics, while assuming access to only the kinds of data that may be available in practice. Primarily this study attempts to assess (1) RIDE’s ability to detect changes in infectivity associated with either mask use or a more infectious, emergent viral variant, and (2) RIDE’s ability to detect an increase in superspreading. Ultimately, we aim to demonstrate that future application of RIDE with real-world data has the potential to augment understanding of viral transmission dynamics, thereby informing policy makers in their

response to changing conditions during the current and future pandemics.

## METHODS

### Overview

We derive a mathematical model of in-vehicle viral transmission. Using large-scale, agent-based computer simulation with empirical diagnoses and rideshare data from Los Angeles County, we apply this mathematical transmission model to generate hypothetical SARS-CoV-2 rideshare infection patterns representative of a large US city during quarantine. We develop RIDE, a statistical method for the analysis of these simulated infection patterns assuming access only to the kind of data that would be available to an epidemiologist, for example, with incomplete knowledge of infections and infection origins. As a proof-of-concept, we apply a limited Monte Carlo technique to test the power of this statistical method to differentiate between the mathematical transmission parameters corresponding to a baseline strain versus other hypothetical transmission scenarios.

### Mathematical model of transmission

We estimate patient infectivity relative to symptom onset with a general mathematical model of viral transmission adapted with SARS-CoV-2-specific parameter estimates from the literature [10,26,27]. In-vehicle probability-of-infection functions incorporating ride duration and ride timing relative to the infector’s symptom onset are derived for unique passenger-to-driver and driver-to-passenger transmission dynamics. We define four hypothetical scenarios of transmission (Table 1). Scenario “viral variant A” assumes no facemask use and overall viral load homogeneity across infected individuals. This serves as the baseline. In scenario “viral variant A with masks,” we represent masking with a 50% reduction in viral particle exchange between infector and potential infectee. In scenario “viral variant A with increased superspreading,” we increase the likelihood of superspreading by introducing infectivity asymmetry in the population, with 1 in 20 infected individuals having a five-fold higher overall probability of transmitting SARS-CoV-2 [6]. In scenario “viral variant B,” based on research detailing the emergent D614G SARS-CoV-2 strain, we introduce a viral variant with 300% higher viral load relative to the baseline viral variant A (see Supplementary Section II for details) [7].

### Generation of synthetic rideshare transmission data

We apply agent-based modeling to generate synthetic data to represent the data that may become available if public health agencies partner with rideshare companies. The characteristics of urban rideshare networks, COVID-19 diagnoses, and estimates of the fraction of true infections represented by the diagnoses from March 17, 2020 to October 3rd, 2020 (200 days) in Los Angeles County (LA) were derived from the literature [28–30]. This time period represents the “first wave” of the pandemic in LA, occurring in spring and summer of 2020. LA rideshare volume, with an estimated 75% reduction during this period of quarantine, was simulated [30]. The total number of infections in LA was assumed to be twice the number of diagnosed infections reported over the period considered [25,31]. Each network was initialized by assigning a day of symptom onset to randomly selected passengers and drivers, with the number of infections proportional to the estimated historical number of infections reported 3 days later (to account for the average delay between symptom onset and testing) [32]. The time of symptom onset was chosen from a truncated normal distribution calibrated to best fit the distribution of empirical LA di-

**Table 1**

Hypothesized sets of SARS-CoV-2 transmission parameters and probability of infection for populations with differing transmission characteristics

Parameters	Estimates from literature (mean [95% confidence interval])	"Viral variant A ..."			
		"... "	"... with masks"	"... with increased super-spreading"	"Viral variant B"
		No masks, baseline super-spreading	Masks, baseline super-spreading <sup>†</sup>	No masks, 5% super-spreaders <sup>‡</sup>	No masks, baseline super-spreading
<b>Viral expulsion</b> relative to viral variant A – %	Not available	100%	-	-	300%
Viral exchange reduction due to NPIs – %	Not available	-	50%	-	-
Infectivity of "superspreader" relative to baseline viral variant A – %	Not available	-	-	500%	-
Days infectious before symptom onset – days	2.3 [0.8-3.0]	2.3	2.3	2.3	3
Peak infectivity relative to symptom onset – days*	-0.7 [-2.0-0.2]	-0.7	-0.7	-0.7	-2
probability of <b>passenger-to-driver</b> infection from 20 min ride w/ passenger on day of symptom onset – %	-	1.5%	0.7%	7.40%	2.8%
probability of <b>driver-to-passenger</b> infection from 20 min ride w/ driver on day of symptom onset – %	-	2.9%	1.5%	14.70%	5.6%

See Supplementary Section II for details of the mathematical model of in-vehicle viral transmission.

\* Adjusted for via gamma infectivity distribution shape and scale parameters.

<sup>†</sup> Passenger and driver both wearing masks, blocking 50% of viral particle exchange.

<sup>‡</sup> Increased likelihood of superspreading: Assumes 5% of infected individuals are "superspreaders" with 20x increased cumulative viral load; the remaining 95% of infected individuals are 78.9% as infectious as baseline individuals with viral variant A.

agnoses during the simulated time period (see Supplementary Section III for details).

For each rideshare interaction involving an infected individual, the probability of infection was calculated using the parameter set corresponding to the hypothetical viral scenario being tested and via applying the previously derived passenger-to-driver or driver-to-passenger transmission functions. The interaction was assumed to result in transmission based on a draw from a Bernoulli random variable with probability of success equal to the calculated probability of infection. For those infected, an incubation period was drawn from a normal distribution, and a subsequent symptom onset time was assigned.

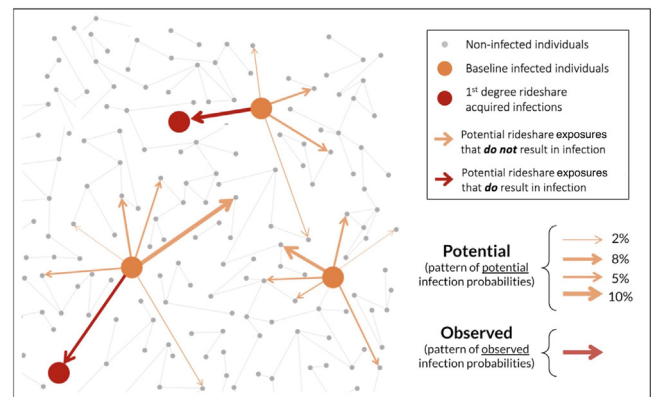
From this simulated data, we then derive data representative of the empirical data that would be available to an epidemiologist. We consider a *Partial Reporting* scenario with access to diagnosis and symptom onset data for 25% of infections and a *Full Reporting* scenario in which data are available for all infections.

Full mathematical details for the generation of the simulated data are available in Supplementary Section III.

*Hypothesis testing with RIDE*

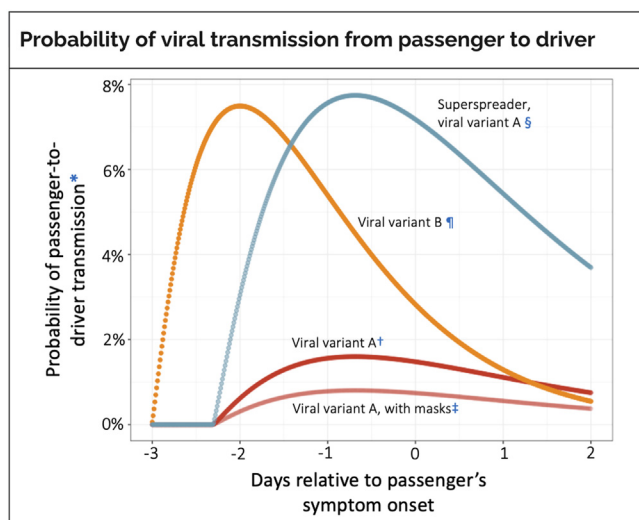
Using only the data that would be available to an epidemiologist (neither undiagnosed infections nor information on the origin of each diagnosed infection), we introduce Rideshare Infection Detection (RIDE) to analyze the simulated rideshare infection patterns in order to calculate the number of expected infections and the number of observed infections (Fig. 1).

Using synthetic data produced via simulated propagation of viral variant A, we test hypotheses about which scenario (viral variant A, viral variant A with masks, or viral variant B) best corresponds to the apparent number of rideshare infections observable in the network. For each scenario, we use the parameter set corresponding to the scenario being tested and knowledge only of diagnosed infections to calculate the expected number of rideshare infections, equal to the sum of the probabilities of infection across all potentially infectious rideshare interactions (i.e., the sum of the expected values of the Bernoulli distributions). The observed number of rideshare infections is determined by counting the num-



**Fig. 1.** Simplified schematic of RIDE analytical method to identify potential and observed infections. Observed infections are tabulated by counting individuals with symptom onset following a potentially infectious rideshare interaction. The expected number of infections is the sum of all potential infections, each weighted by its probability of infection. Arrow thickness for each potential infection corresponds to the probability-of-infection magnitude, which is calculated given a mathematical model of rideshare transmission that depends on ride duration, the timing of the ride relative to the potential infector's symptom onset, and the assumed SARS-CoV-2 transmission parameters given the hypothesized scenario and viral variant.

ber of interactions in which a diagnosed individual within their infectious window shared a rideshare vehicle with a potential infectee who had a positive diagnosis with symptom onset between 1.5 and 10 days following the rideshare trip. This observed number is then adjusted to account for the percentage of infections diagnosed (this diagnosis percentage is assumed known) and for an estimation of the average number of "false positives" (interactions that appeared to have resulted in an infection even though the infectee was infected elsewhere) given the overall infection density in the network. For each respective level of reporting (*Full* vs. *Partial Reporting*), this simulated propagation and analysis with RIDE was repeated 10 times to determine the impact of transmission stochasticity and rideshare network variability. Across the 10 simulations, the differences in the expected number of passenger-to-



**Fig. 2.** Modeled probability of transmission from passenger to driver. Probability of infection varies significantly depending upon interaction dynamics (time of ride relative to passenger's symptom onset and ride duration) and upon assumptions defining the hypothetical SARS-CoV-2 transmission parameters (representing the passenger's viral variant; whether the passenger is a superspreader; and whether face masks are used). \*Probability of driver infection given 20 min ride with infected passenger. †Baseline viral variant. ‡Passenger and driver wearing masks, blocking 50% of viral particle exchange. §Superspreaders with viral variant A are 500% more infectious than baseline individuals with viral variant A. ¶Alternative viral variant B is 300% more infectious than baseline and has different infectivity parameters within the previously estimated 95% confidence intervals.

driver infections (given each set of hypothesized parameters) and the observed number of passenger-to-driver infections are compared with a pairwise Kruskal-Wallis test (Supplementary Section IV).

Separately, two synthetic rideshare infection patterns were simulated and compared: propagation of *viral variant A* and propagation of *viral variant A* with increased superspreading. For each round of analysis, the differences were calculated between the number of observed passengers infected per infectious driver for the increased superspreading scenario minus the respective values for the baseline scenario. This was repeated 10 times each for both *Partial* and *Full Reporting*, and the resulting distributions were compared with the Kruskal-Wallis test. All P-values were adjusted for multiple testing (Supplementary Section IV).

All simulated viral propagation and analysis with RIDE were performed with R (Version 4.0.3, 2010; Vienna, Austria) and executed with Stanford's Sherlock High-Performance Computing Cluster.

## RESULTS

### Mathematical model

Differences in viral strains and the various hypothetical propagation scenarios considered resulted in notable differences in the probability of rideshare-acquired SARS-CoV-2 infection (Fig. 2). Relative to the baseline scenario probability-of-infection given a 20-minute ride without mask use with a passenger with *viral variant A* one day before symptom onset, the probability of driver infection was found to be 51% lower when both driver and passenger are masked (scenario *viral variant A* with masks), 480% higher when the passenger was one of the more infectious individuals from scenario *viral variant A* with increased superspreading, and 342% higher when the passenger was infected with *viral variant B*.

### Synthetic rideshare transmission data

The simulation was initiated with a baseline infection probability of 528,828 out of 10 million, assuming an infection prevalence double the 264,414 diagnosed, reported COVID-19 infections in LA during this period. Data from 10 simulated trials of SARS-CoV-2 propagation within the LA rideshare network resulted in an average of 190,387.1 (range 187,898–193,645) potentially infectious rideshare interactions, encompassing possible passenger-to-driver and driver-to-passenger transmissions. When these 10 simulated trials were propagated assuming SARS-CoV-2 transmission parameters corresponding to *viral variant A*, there were an average of 409.0 (range 384–424) rideshare infections resulting in a diagnosis for the *Partial Reporting* scenario (access to data on 25% of infections), and 1666.4 (range 1,614 to 1,698) diagnosed rideshare infections for the *Full Reporting* scenario.

### Hypothesis testing with RIDE

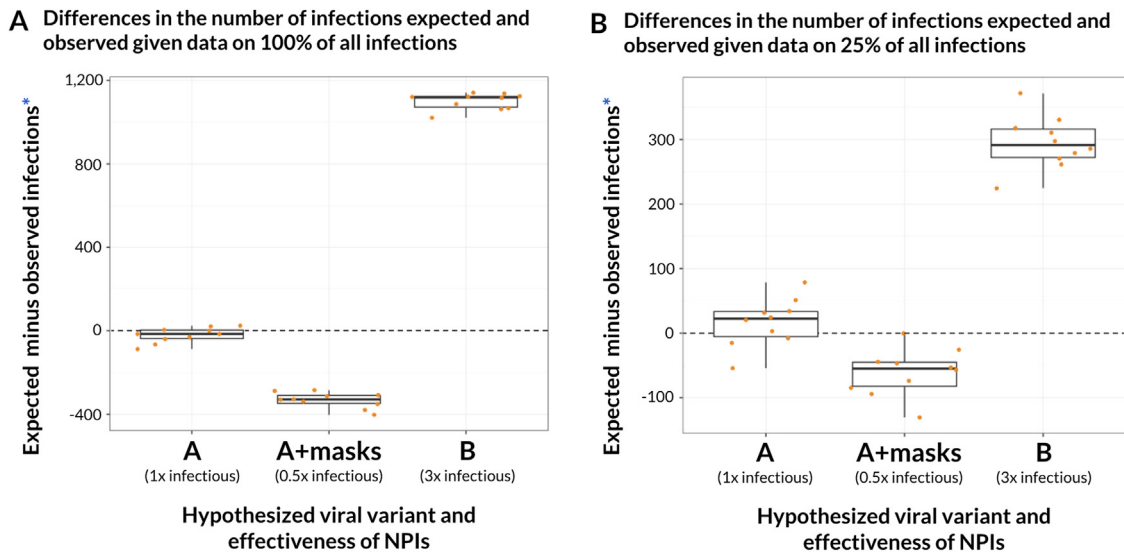
Across 10 trials of propagation with *viral variant A* followed by hypothesis testing with RIDE given *Partial Reporting*, the difference between the number of expected minus observed passenger-to-driver rideshare infections was 16.7 (range –54.4 to 78.8) when assuming *viral variant A* without masks or increased superspreading; –61.0 (range –130.3 to 0.2) when assuming *viral variant A* with masks without increased superspreading; and 294.9 (range 224.8–371.4) when assuming *viral variant B* without masks or increased superspreading (all adjusted P-values < .001). The results were qualitatively similar in the *Full Reporting* scenario, with greater differences, less variation, and a higher level of significance (Fig. 3).

For the secondary investigation comparing analytical results given simulated propagation of *viral variant A* with increased superspreading relative to simulated propagation of *viral variant A*, the mean observed number of drivers that infected exactly one passenger was 9.2 (range 29 to –6) lower in the superspreader scenario and the mean number of drivers that infected exactly two passengers was 1.8 (range –8 to 7) higher in the superspreader scenario given *Partial Reporting*. The results were qualitatively similar in the *Full Reporting* scenario, with greater differences (Fig. 4). In both the *Full* and *Partial Reporting* scenarios, the combined data from the 10 trials led to significant differences in the distributions (*Full Reporting*,  $P < .001$ ; *Partial Reporting*,  $P < .05$ ).

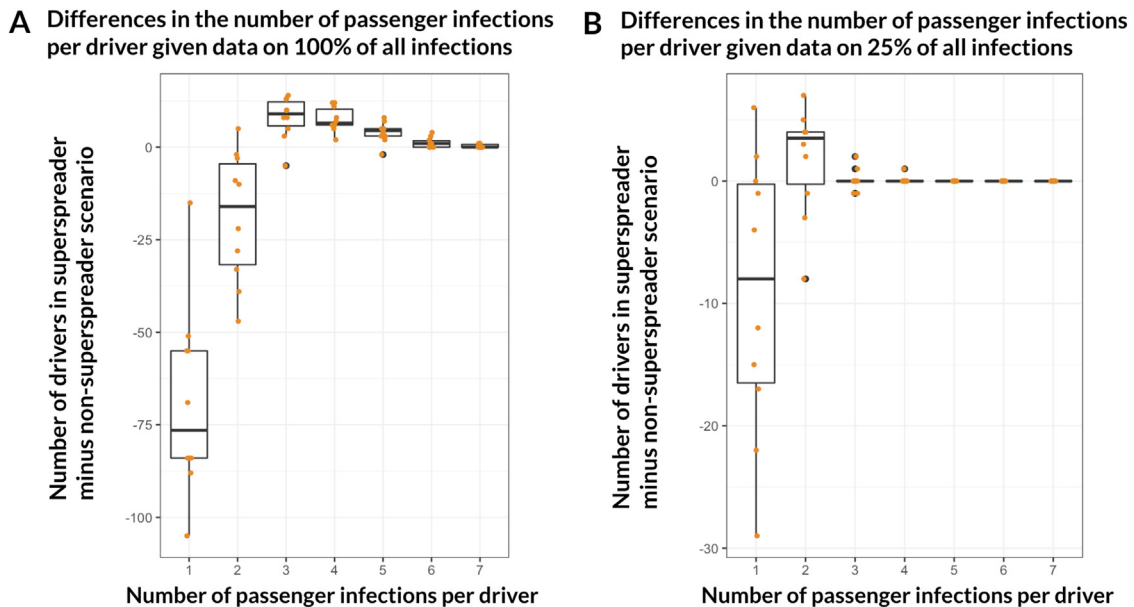
## LIMITATIONS

The primary limitations of this work stem from the use of synthetic data. Although we include sources of variability in our model by incorporating stochastic ride durations, ride timing relative to symptom onset, transmission heterogeneity, etc., real-world infection patterns would include additional sources of noise. For example, we do not consider some factors affecting in-vehicle air circulation, such as the potential for open windows [22], seating arrangements [33], whether the airflow system is operating [34], potential barriers between the front and back seat, or differences in vehicle sizes. Furthermore, we do not incorporate the possibility of systematic infectivity differences between diagnosed and undiagnosed individuals, immunity from recent SARS-CoV-2 infection, non-uniform diagnostic testing and reporting rates between drivers and passengers, nor systematic differences between neighborhoods within cities. Finally, our hypothesis testing with RIDE assumes accurate estimation of the total fraction of infections that are diagnosed (which is not precisely known by the scientific community), as well as accurate data about infected individuals' symptom onset time relative to diagnosis, which may not be consistently recorded.

The simulated propagations for this proof-of-concept testing considered COVID-19 infections in LA during the first wave of the



**Fig. 3.** Differences given 10 trials in the number of passenger-to-driver infections expected and observed in the simulation based on analysis with different hypotheses of SARS-CoV-2 transmission, according to the percent of infections reported (left, *Full Reporting*; right, *Partial Reporting*). Variability in results given analyses of 10 simulated synthetic datasets resulting from SARS-CoV-2 propagation in Los Angeles County when true propagation conditions correspond to *viral variant A* with no masks. For each simulated dataset, and with access to only diagnosed infections, the epidemiologist assesses the difference in the expected number of passenger-to-driver infections given analysis with hypothesized parameters (assuming either *viral variant A*, *viral variant A with masks*, or *viral variant B*) minus the adjusted number of observed rideshare infections in the network. Each dot within a box-plot represents results with the given hypothesis for one round of Los Angeles County simulation and analysis. Box-plot midline represents the median of analysis results across the 10 trials, box edges show interquartile range, and whisker tips show the minimum and maximum result values. \*Observed infections adjusted to account for undiagnosed infections and baseline "false-positive" transmissions (Supplementary IV).



**Fig. 4.** Differences in the number of passenger infections per driver in simulations with baseline variant versus variant with increased superspreading, according to the percent of infections reported (left, *Full Reporting*; right, *Partial Reporting*). Variability in the differences between results given analyses of 10 simulated synthetic datasets resulting from SARS-CoV-2 propagation of *viral variant A* in Los Angeles County given either a population with homogeneous cumulative infectivity (baseline superspreading scenario) or asymmetric infectivity (increased superspreading scenario, where 5% of infected individuals are "superspreaders"). Each dot within a box-plot represents analysis results with the given hypothesis for one round of Los Angeles County simulation and analysis. Box-plot midline represents the median of analysis results across the 10 trials; box edges show interquartile range; whisker tips are minimum and maximum values; and black dot shows outliers, as specified by Tukey.

pandemic, when there were 132,207 diagnosed infections over 6 months. However, in the more recent Omicron wave, in just one month (December 25, 2021 to January 25, 2022) LA's cumulative diagnosed case count rose from 1.5 million to 2.5 million [29]. Given this higher relative rate of infections, RIDE's detection power would potentially be 24-fold greater during the Omicron wave. Moreover, given improved PCR and antigen testing infrastructure, diagnosis rates have since likely increased above the conservative 25% estimate used for the *Partial Reporting* scenario. Finally, real-

world deployment of RIDE could aggregate data across multiple cities and combine analyses of passenger-to-driver and driver-to-passenger transmissions. Thus, the additional noise introduced by the aforementioned limitations would likely be mitigated by these factors given real-world deployment of RIDE (see Supplementary Section V for details).

For simplicity, we did not address all possible scenario permutations (e.g., mask use  $\times$  new viral variant). Instead, we focused on assessing RIDE's detection ability with regards specific events of

the COVID-19 pandemic, such as the start of the mask mandate and the rapid Delta variant outbreak. Other scenario permutations may represent opportunities for future research.

## DISCUSSION

Using simulated viral propagation patterns generated based on empirical ridesharing data and COVID-19 diagnoses records from Los Angeles County (LA), we demonstrate that analysis of rideshare data may allow for the identification of emergent SARS-CoV-2 strains and the study of their transmission characteristics. Our analyses of LA's first COVID-19 wave show that such an approach may be effective given data aggregated over 6 months and when as few as 25% of infectious individuals have been identified. While additional sources of noise may confound real-world deployment of RIDE, consideration of our simulations' conservative assumptions suggests that meaningful, even real-time pandemic monitoring may be feasible. Together, our findings indicate that further investigation is warranted for the development of such a system at the national scale and with access to more detailed data.

We demonstrate that current research-based estimates for the range of SARS-CoV-2 transmission parameter values leave significant uncertainty in transmission modeling, leading to substantial variability in probability-of-infection predictions for a typical rideshare interaction. We demonstrate the ability of RIDE to differentiate between values within the current estimated range, potentially enabling measured appraisal of the emergence of more infectious viral strains; the effectiveness of Uber and Lyft's nationwide mask mandates for all drivers and passengers; and the extent of superspreading in the population.

Previous epidemiological modeling research combining diagnoses records and retrospective passenger transportation data from high-speed trains in China sets precedent for RIDE [33]. The standardized format of rideshare data at an international scale presents a significant opportunity for a more extensive study of viral transmission. While our simulations have similarities with previous agent-based modeling studies predicting SARS-CoV-2 transmission patterns within other settings (e.g., universities) [35], the main contribution of our work is not estimating the number of rideshare-based infections, but rather developing a method to analyze real-world rideshare infection patterns if we were to have access to retrospective data.

Straightforward expansions of RIDE could increase real-world hypothesis-testing power via the aggregation of data across multiple cities and with the integration of additional forms of data, such as vaccination records. These extensions could be used to evaluate the extent to which vaccines reduce the infectivity of vaccinated individuals relative to unvaccinated infected individuals, and how this reduction may diminish over time given emergent viral variants or due to naturally waning immunity.

Other avenues of potential investigation include assessment of the effectiveness of other NPIs beyond masks and the prevalence and dynamics of possible passenger-to-passenger infections due to indirect contact through shared surface contact and leftover aerosols in the vehicle. Finally, this analytical method paired with the communication features of rideshare platforms could facilitate a largely automated "radar" for contact tracing within the rideshare network.

The US Center for Disease Control and Prevention may have grounds to require rideshare dataset access so that it can be merged with the list of positive COVID-19 diagnoses, strain sequencing data, vaccination records, and other relevant data. The case has been made for digital disease surveillance that maintains considerations of ethics and patient privacy [36]. Large rideshare companies such as Uber and Lyft are already sharing data with public health officials to assist with contact tracing [23], but no

standardized analytics framework is available to aggregate and derive insights from these data. While logistics—the details of which are beyond the scope of this article—for the large-scale merging of rideshare and infection data are complex, it could be accomplished in an anonymous fashion given careful data processing.

## CONCLUSION

The rideshare network of exposure is unlike any other, with tens of millions of potentially infectious connections between individuals worldwide. Unlike more general cellphone-based contact data, rideshare contacts occur in a relatively controlled environment with fairly consistent spatial dynamics, are often the only connection between individuals, and are sporadic with respect to time, location, and duration. We demonstrated, via simulations of COVID-19 propagation through a rideshare network based on Los Angeles County, that viral strains with differing SARS-CoV-2 transmission parameters lead to detectably different patterns of infections, even in the presence of limited diagnostic information. For the present and subsequent pandemics, analysis of rideshare data combined with diagnosis records may augment efforts to better understand viral transmission dynamics and to measure the changes in infectivity associated with nonpharmaceutical interventions, vaccination, or emergent viral strains.

## Author contributions

David Scheinker had a primary role in conceptualization, methodology development, supervision, validation, literature search, and writing (review and editing). Conrad Safranek had a primary role in conceptualization, methodology development, software (programming for simulation and analysis), tables, figures, literature search, and writing (original draft). Both authors (Conrad Safranek and David Scheinker) verified the underlying data.

## Ethics approval

Not applicable; No private data on human subjects was used in this research.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

We would like to thank James Safranek for his contributions to the development of the mathematical model of rideshare viral transmission. We also thank John Tamareis and the Stanford Data Studio consulting group for their feedback on the statistical methods presented in this paper. Some of the computing for this project was performed on the Stanford University Sherlock High-Performance Computing cluster.

## Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:[10.1016/j.annepidem.2022.08.051](https://doi.org/10.1016/j.annepidem.2022.08.051).

## REFERENCES

- [1] Tao K, Tzou PL, Nouhin J, Gupta RK, de Oliveira T, Kosakovsky Pond SL, et al. The biological and clinical significance of emerging SARS-CoV-2 variants. *Nat Rev Genet* 2021;17:757–73. Published online September. doi:[10.1038/s41576-021-00408-x](https://doi.org/10.1038/s41576-021-00408-x).

- [7] Wang P, Nair MS, Liu L, Iketani S, Luo Y, Guo Y, et al. Antibody resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7. *Nature* 2021;593(7857):130–5.
- [8] Anderson RM, Heesterbeek H, Klinkenberg D, Hollingsworth TD. How will country-based mitigation measures influence the course of the COVID-19 epidemic? *Lancet* 2020;395(10228):931–4.
- [9] Barnes M, Sax PE. Challenges of “Return to Work” in an Ongoing Pandemic. *N Engl J Med* 2020 Published online. doi:10.1056/NEJMSr2019953.
- [10] Wang X, Ferro EG, Zhou G, Hashimoto D, Bhatt DL. Association between universal masking in a health care system and SARS-CoV-2 positivity among health care workers. *JAMA* 2020;324(7):703–4.
- [11] Adam DC, Wu P, Wong JY, Lau EHY, Tsang TK, Cauchemez S, et al. Clustering and superspreading potential of SARS-CoV-2 infections in Hong Kong. *Nat Med* 2020;26(11):1714–19.
- [12] Korber B, Fischer W, Gnanakaran S, Yoon H, Theiler J, Abfalterer W. Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus. *Cell* 2020;182(4):812–27 Published online June 26. doi:10.1016/j.cell.2020.06.043.
- [13] European Centre for Disease Prevention and Control. *Threat Assessment Brief: Rapid Increase of a SARS-CoV-2 Variant with Multiple Spike Protein Mutations Observed in the United Kingdom.*; 2020. <https://www.ecdc.europa.eu/sites/default/files/documents/SARS-CoV-2-variant-multiple-spike-protein-mutations-United-Kingdom.pdf>.
- [14] Pegu A, O’Connell SE, Schmidt SD, O’Dell S, Talana CA, Lai L, et al. Durability of mRNA-1273 vaccine-induced antibodies against SARS-CoV-2 variants. *Science* 2021;373(6561):1372–7.
- [15] He X, Lau EHY, Wu P, Deng X, Wang J, Hao X, et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nat Med* 2020;26(5):672–5.
- [16] van Doremalen N, Bushmaker T, Morris DH, Holbrook MG, Gamble A, Williamson BN, et al. Aerosol and surface stability of SARS-CoV-2 as compared with SARS-CoV-1. *N Engl J Med* 2020;382(16):1564–7.
- [17] Harris RJ, Hall JA, Zaidi A, Andrews NJ, Dunbar JK, Dabrera G. Effect of vaccination on household transmission of SARS-CoV-2 in England. *N Engl J Med* 2021;385(8):759–60.
- [18] Letizia AG, Ramos I, Obla A, Goforth C, Weir DL, Ge Y, et al. SARS-CoV-2 Transmission among marine recruits during quarantine. *N Engl J Med* 2020;383:2407–16 Published online November 11. doi:10.1056/NEJMoa2029717.
- [19] Baden LR, El Sahly HM, Essink B, Kotloff K, Frey S, Novak R, et al. Efficacy and Safety of the mRNA-1273 SARS-CoV-2 vaccine. *N Engl J Med* 2021;384(5):403–16.
- [20] Scobie HM, Johnson AG, Suthar AB, Severson R, Alden NB, Balter S, et al. Monitoring incidence of COVID-19 cases, hospitalizations, and deaths, by vaccination status – 13 U.S. jurisdictions, April 4–July 17, 2021. *MMWR Morb Mortal Wkly Rep* 2021;70(37):1284–90.
- [21] Levine-Tiefenbrun M, Yelin I, Katz R, Herzl E, Golan Z, Schreiber L, et al. Initial report of decreased SARS-CoV-2 viral load after inoculation with the BNT162b2 vaccine. *Nat Med* 2021;27(5):790–2.
- [22] Kang M, Xin H, Yuan J, ST Ali, Liang Z, Zhang J, et al. Transmission dynamics and epidemiological characteristics of Delta variant infections in China. *Euro Surveill* 2022;27(10):2100815. doi:10.2807/1560-7917.ES.2022.27.10.2100815.
- [23] Brown CM, Vostok J, Johnson H, Burns M, Gharpure R, Sami S, et al. Outbreak of SARS-CoV-2 infections, including COVID-19 vaccine breakthrough infections, associated with large public gatherings — Barnstable County, Massachusetts, July 2021. *MMWR Morb Mortal Wkly Rep* 2021;70(31):1059–62. doi:10.15585/mmwr.mm7031e2.
- [24] Griffin JB. SARS-CoV-2 infections and hospitalizations among persons aged  $\geq 16$  years, by vaccination status — Los Angeles County, California, May 1–July 25, 2021. *MMWR Morb Mortal Wkly Rep* 2021;70(34):1170–6. doi:10.15585/mmwr.mm7034e5.
- [25] Budd J, Miller BS, Manning EM, Lampos V, Zhuang M, Edelstein M, et al. Digital technologies in the public-health response to COVID-19. *Nat Med* 2020;26(8):1183–92.
- [26] Grantz KH, Meredith HR, Cummings DAT, Metcalf CJE, Grenfell BT, Giles JR, et al. The use of mobile phone data to inform analysis of COVID-19 pandemic epidemiology. *Nat Commun* 2020;11(1):4961.
- [27] Mathai V, Das A, Bailey JA, Breuer K. Airflows inside passenger cars and implications for airborne disease transmission. *Sci Adv* 2021;7(1):1–7. doi:10.1126/sciadv.abe0166.
- [28] Ferguson NM, Cummings DAT, Cauchemez S, Fraser C, Riley S, Meeyai A, et al. Strategies for containing an emerging influenza pandemic in Southeast Asia. *Nature* 2005;437(7056):209–14.
- [29] Martinez DL, Das TK. Design of non-pharmaceutical intervention strategies for pandemic influenza outbreaks. *BMC Public Health* 2014;14(1):1328.
- [30] COVID-19 Pandemic Planning Scenarios. Centers for Disease Control and Prevention. Published July 10, 2020. <https://www.cdc.gov/coronavirus/2019-ncov/hcp/planning-scenarios.html>.
- [31] LA County COVID-19 Surveillance Dashboard. County of Los Angeles; Public Health. Accessed December 3, 2020. [http://dashboard.publichealth.lacounty.gov/covid19\\_surveillance\\_dashboard/](http://dashboard.publichealth.lacounty.gov/covid19_surveillance_dashboard/).
- [32] Lyft maintains end-2021 profitability goal even as COVID-19 guts demand. Reuters. Published August 12, 2020. <https://www.reuters.com/article/us-lyft-results/lyft-maintains-end-2021-profitability-goal-even-as-covid-19-guts-demand-idUSKCN2582PN>.
- [33] Alwan NA. Surveillance is underestimating the burden of the COVID-19 pandemic. *Lancet* 2020;396(10252):e24.
- [34] Kretzschmar ME, Rozhnova G, Bootsma MCJ, van Boven M, van de Wijgert JHHM, Bonten MJM. Impact of delays on effectiveness of contact tracing strategies for COVID-19: a modelling study. *Lancet Public Health* 2020;5(8):e452–9.
- [35] Hu M, Lin H, Wang J, Xu C, Tatem AJ, Meng B, et al. Risk of Coronavirus disease 2019 transmission in train passengers: an epidemiological and modeling study. *Clin Infect Dis* 2021;72(4):604–10.
- [36] Jones LD, Chan ER, Zabarsky TF, Cadnum JL, Navas ME, Redmond SN, et al. Transmission of severe acute respiratory syndrome Coronavirus 2 (SARS-CoV-2) in a patient transport van. *Clin Infect Dis* 2022;74(2):339–42.
- [37] Harper PR, Moore JW, Woolley TE. Covid-19 transmission modelling of students returning home from university. *Health Syst* 2021;10(1):31–40. doi:10.1080/20476965.2020.1857214.
- [38] Mello MM, Wang CJ. Ethics and governance for digital disease surveillance. *Science* 2020;368(6494):951.