





# Substantial transmission of SARS-CoV-2 through casual contact in retail stores: Evidence from matched administrative microdata on card payments and testing

Niels Johannesen<sup>a,b,c,1</sup>, Alessandro Tang-Andersen Martinello<sup>d</sup>, Bjørn Bjørnsson Meyer<sup>d</sup> , Emil Toft Vestergaard<sup>d</sup>, Asger Lau Andersen<sup>b,c</sup> , and Thais Lærkholm Jensen<sup>d</sup>

Edited by Jose Scheinkman, Columbia University, New York, NY; received October 10, 2023; accepted March 21, 2024

This paper presents quasiexperimental evidence of Covid-19 transmission through casual contact between customers in retail stores. For a large sample of individuals in Denmark, we match card payment data, indicating exactly where and when each individual made purchases, with Covid-19 test data, indicating when each individual was tested and whether the test was positive. The resulting dataset identifies more than 100,000 instances where an infected individual made a purchase in a store and, in each instance, allows us to track the infection dynamics of other individuals who made purchases in the same store around the same time. We estimate transmissions by comparing the infection rate of exposed customers, who made a purchase within 5 min of an infected individual, and nonexposed customers, who made a purchase in the same store 16 to 30 min before. We find that exposure to an infected individual in a store increases the infection rate by around 0.12 percentage points ( $P < 0.001$ ) between day 3 and day 7 after exposure. The estimates imply that transmissions in stores contributed around 0.04 to the reproduction number for the average infected individual and significantly more in the period where Omicron was the dominant variant.

COVID-19 | reproduction number | disease transmission | transaction data

Understanding the transmission of infectious diseases is of fundamental importance for public health. It allows individuals to take adequate measures to protect themselves against infection and enables health authorities to design effective policies to mitigate spreading of the disease.

Most existing studies of disease transmission in real-world environments focus on social networks and report secondary attack rates or reproduction numbers within households (1–6), school classes (7–9), or work places (10–12). These studies have delivered important messages for public policy. For instance, low estimated secondary attack rates in classrooms constituted a strong argument for open schools in the recent Covid-19 pandemic (13).

By contrast, there is barely any evidence on transmissions outside of social networks, i.e., through casual contact in the public space. This is unfortunate as such transmissions may be particularly important for aggregate infection dynamics, by allowing the virus to jump from one social network to another. Credible estimates of the individual and aggregate risks associated with casual contact in the public space are valuable in settings like the recent Covid-19 pandemic as they can guide policies that restrict personal mobility and impose social distancing in public places.

Here, we report the findings from a study of Covid-19 transmission through casual contact between unconnected customers in supermarkets and grocery stores. We develop a research design that harnesses large, naturally occurring datasets on card payments and Covid-19 testing. Specifically, we match payment data, indicating exactly where and when each individual in our sample made purchases, with Covid-19 test data, indicating when each individual was tested and whether the test was positive. This allows us to identify instances where an infected individual made a purchase in a store and track the infection dynamics of other individuals who made purchases in the same store at around the same time.

Our quasiexperimental empirical design provides evidence on transmission in stores by comparing the infection rates for customers in the same store whose exposure to the infected individual differed due to the precise timing of their store visits. For example, if an infected individual made a purchase at 14.13, we compare individuals like Ms. Jones who made a purchase in the same store at 14.15 (“exposed”) to individuals like Ms. Jennings who made a purchase at 13.51 (“nonexposed”). Assuming that the exact timing of transactions within a short time span is quasi-random—i.e., that Ms. Jones and Ms. Jennings are not systematically different with respect to exposures outside the store—the probability of transmission in the store can be inferred from the difference in the infection rates of the two groups after the exposure.

## Significance

The recent Covid-19 pandemic highlighted that understanding the channels of disease transmission is crucially important for public health policies. However, measuring transmissions occurring through casual contact in the public space is highly challenging as researchers generally do not observe when infected individuals intersect casually with noninfected individuals. We overcome this methodological challenge in the context of the Covid-19 pandemic by combining card payment data, indicating exactly where and when individuals visited stores, with test data indicating when they were infected. We document that exposure to an infected individual in a store is associated with a significantly higher infection rate in the following week. Our estimates imply that transmissions between retail shoppers made a substantial contribution to the Covid-19 pandemic.

Author contributions: N.J., A.T.-A.M., B.B.M., E.T.V., A.L.A., and T.L.J. designed research; N.J., A.T.-A.M., B.B.M., E.T.V., A.L.A., and T.L.J. performed research; N.J. analyzed data; N.J., A.T.-A.M., A.L.A., and T.L.J. performed project administration; A.T.-A.M., B.B.M., and E.T.V. prepared data; and N.J. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Copyright © 2024 the Author(s). Published by PNAS. This article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

<sup>1</sup>To whom correspondence may be addressed. Email: niels.johannesen@sbs.ox.ac.uk.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2317589121/-/DCSupplemental>.

Published April 17, 2024.

Our analysis documents substantial Covid-19 transmission between consumers in supermarkets and grocery stores. We find that making a purchase in a store within 5 min of an infected person raises the infection rate between day 3 and day 7 after the purchase by 0.12 %-points ( $P < 0.001$ ). This estimate of in-store transmission compares to a baseline infection rate of 1.3% for the nonexposed, implying a relative risk ratio of 1.09. The estimated transmission rate in stores is considerably higher in the period where Omicron was dominant, which is consistent with other evidence of this variant's exceptional transmissibility (14–16).

Our estimates have strikingly different implications for the individual and aggregate risks associated with casual contact in the public space. On the one hand, we find that the average store visit in the estimation period involved a very small probability of contracting Covid-19 of around 0.00025. This reflects that both the risk of exposure conditional on going to a store and the infection risk conditional on in-store exposure were relatively low. On the other hand, we find that each infected individual on average transmits the virus to around 0.04 others through casual contact in supermarkets and grocery stores, implying a substantial contribution of shopping to aggregate reproduction of the virus in society. These two seemingly contradictory implications are reconciled by the fact that grocery shopping is a frequent activity that involves casual contact with a large number of people.

It is illustrative to compare our results to existing evidence on Covid-19 transmission within households. Our estimated transmission rate of 0.12% for casual contacts in stores is two orders of magnitude smaller than the transmission rate of 20 to 30% typically found for household members (2, 3). Despite the low transmission rate in stores, the contribution to overall reproduction remains quantitatively relevant due to the high number of exposures. Specifically, for the average infected individual, the number of other customers making a transaction within 5 min is around 30, one order of magnitude more than the number of household members (2, 3).

A key contribution of our study is to provide estimates of the epidemiological risks associated with retail shopping in the Covid-19 pandemic. Most governments around the world imposed restrictions on consumer activity with the aim of reining in the pandemic at high costs for both households and businesses. One can interpret the estimated reproduction number in stores of around 0.04 as an upper bound on the reduction in infections that can be achieved through restrictions on stores. With an overall reproduction number hovering around one throughout most of the pandemic, this interpretation suggests that such restrictions can have a limited, but meaningful dampening effect on the infection dynamics.

Our study also makes an important methodological contribution: we develop a unique quasiexperimental method for measuring disease transmission through casual contact in the public space. Our method represents a major advance over existing analytical approaches, i.e., case studies of specific outbreaks (17–19) and case–control studies using surveys to investigate differences in potential community exposures across infected and noninfected individuals (20–22). Further, our method overcomes many of the limitations inherent to existing studies of transmission within social networks (cited above) by using an objective measure of physical proximity rather than self-reported close contact and by accounting rigorously for background infection risk.

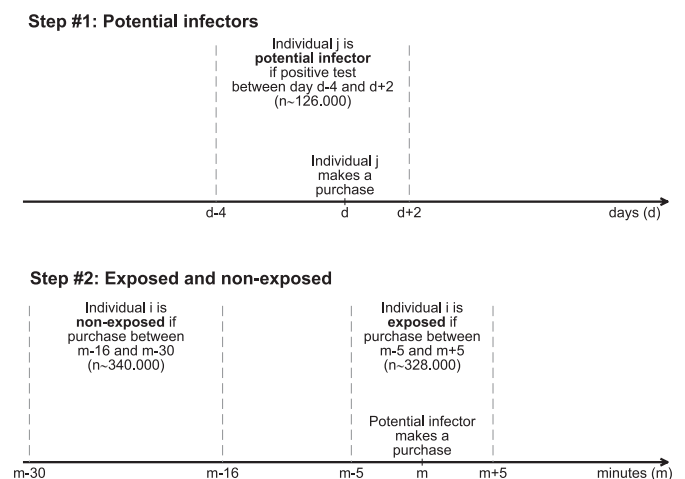
## Data and Research Design

To implement our research design, we first combine data from two sources. From Danske Bank, the largest bank in Denmark, we obtain card transaction data for around 630,000 customers (23). The transaction dataset includes unique store identifiers and

indicates the exact time of each payment. Matching on unique personal identifiers, we link the transaction data to comprehensive administrative data on Covid-19 tests from the national health authorities (24) (*SI Appendix, section 2.1*). These two data sources are attractive in the Danish setting where card payments account for almost 90% of all payments in stores (25), where test frequencies during the pandemic were among the highest in the world and up to 80% of all infections were detected (26, 27), and where the administrative test data comprises both molecular and antigen tests performed by public as well as private test providers.

In the resulting dataset, we identify around 126,000 instances where a transaction on day  $d$  was made by an individual with a positive Covid-19 test between day  $d - 4$  and day  $d + 2$ . As individuals infected with Covid-19 are generally contagious from around 2 d before the onset of symptoms and at least until 5 d after, these individuals were potential infectors when making the purchase on day  $d$ . Our main estimation sample consists of individuals who made a purchase within 5 min of a potential infector on day  $d$  (around 328,000 exposed individuals) and individuals who made a purchase between 16 and 30 min before a potential infector (around 340,000 nonexposed individuals). We do not consider individuals who made purchases after a potential infector as nonexposed due to possible contamination of air and surfaces (28–31). Fig. 1 illustrates the selection of potential infectors as well as the groups of exposed and nonexposed individuals.

Importantly, individuals socially connected to the infector may confound the analysis if they are more likely to make purchases around the same time as the infector and additionally are more likely to interact with the infector outside of the store. If the reason Mrs. Jones made a purchase only 2 min after the potential infector is that they are work colleagues and went to the store together in the break, a higher infection rate for Mrs. Jones may reflect transmission during working hours rather than transmission in the store. We augment the main dataset with information from administrative sources (32–34) with the aim of excluding individuals with social ties to the potential infector. Specifically, we exclude members of the same family or household as the potential infector (around 2,200 individuals); employees at the same firm (around 800 individuals); students at the same school (around 1,300 individuals); as well as individuals with any mobile money transfers to or from the potential infector during the sample period



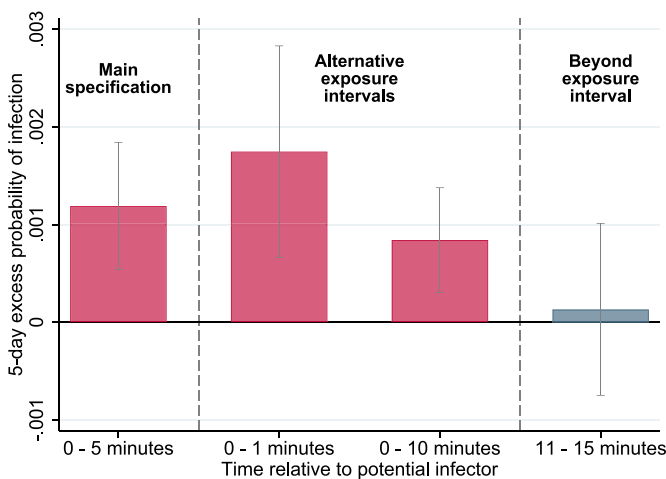
**Fig. 1.** Empirical design. The figure illustrates the empirical design. In the first step, we identify potential infectors who made a purchase on day  $d$  and tested positive between day  $d - 4$  and day  $d + 2$ . In the second step, we identify exposed individuals who made a purchase within 5 min of a potential infector in the same store and a reference group of nonexposed individuals who made a purchase between 16 and 30 min before a potential infector in the same store.

(around 7,000 individuals) (*SI Appendix, section 2.2*). The potentially confounding effects of social networks is an important motivation for restricting the analysis to grocery shopping, which is presumably less social than other consumer activities such as eating out, going to shows, or shopping for clothes.

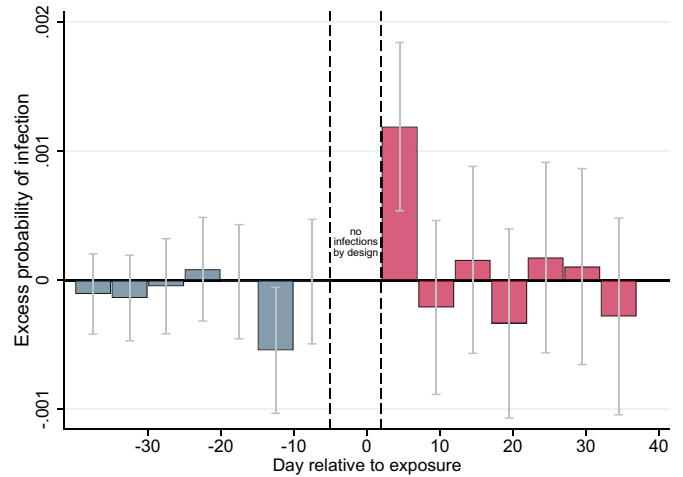
In the main analysis, the outcome is a variable indicating a positive Covid-19 test between day  $d + 3$  and day  $d + 7$ , corresponding to the typical period where symptoms would emerge following infection on day  $d$ . We regress this outcome variable on an indicator for being exposed while including a separate intercept for each set of individuals, exposed and nonexposed, linked to the same potential infector. With this specification, we capture the differential infection rate of exposed individuals relative to nonexposed individuals who made a purchase in the same store at almost the same time of the same day. This parameter reflects transmission in the store under the assumption that there are no confounding differences in exposures outside the store (*SI Appendix, section 3*). While we generally expect consumers to sort in ways that correlate strongly with other exposures—e.g., the elderly may shop in local stores during normal working hours whereas young families may prefer malls on weekends—our research design overcomes this challenge by making comparisons within stores and narrow time windows. Indeed, the differences between exposed and nonexposed individuals are generally immaterial across a range of observable characteristics such as age, gender, household size, income, and occupation (*SI Appendix, section 4.1*).

## Results

Fig. 2 illustrates the first set of results. The main specification suggests that in-store exposure increases the 5-d infection rate by around 0.12%-points ( $P < 0.001$ ). When we vary the time interval that defines exposure, we obtain results consistent with the notion that individuals who make transactions nearer to the potential infector are more likely to have close contact in the store and therefore more likely to be infected. Specifically, the estimated effect increases to around 0.18%-points ( $P = 0.002$ ) when exposure is defined more narrowly as transactions within 1 min of the potential infector and decreases to around 0.08%-points ( $P = 0.002$ ) when it is defined



**Fig. 2.** Main results. The first bar indicates the excess probability of testing positive between day  $d + 3$  and day  $d + 7$  for individuals making a purchase within 5 min of a potential infector on day  $d$  (exposed) relative to individuals making a purchase between 16 and 30 min before a potential infector on day  $d$  (nonexposed). The next two bars show analogous estimates for alternative definitions of exposure, i.e., purchases within 1 and 10 min of a potential infector. The final bar shows an analogous estimate for placebo exposure, i.e., purchases between 11 and 15 min before a potential infector. The estimated coefficients and SE are reported in *SI Appendix, Table S4*.



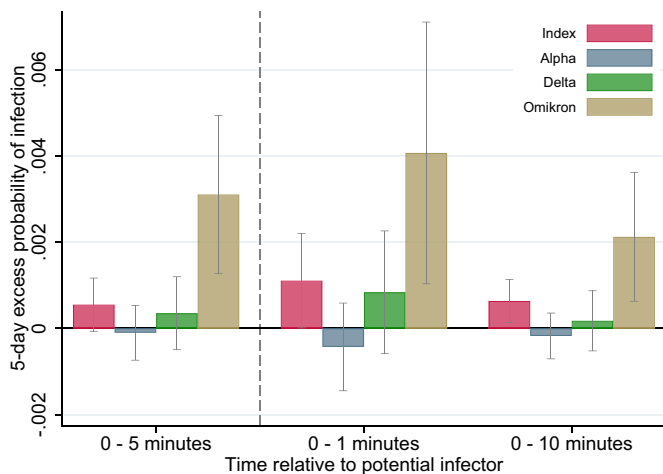
**Fig. 3.** Dynamic results. The bars indicate the excess probability of testing positive in different 5-d periods for individuals making a purchase within 5 min of a potential infector on day  $d$  (exposed) relative to individuals making a purchase between 16 and 30 min before a potential infector on day  $d$  (nonexposed). There are seven 5-d periods before the purchase (blue bars) and seven 5-d periods after the purchase (red bars).

more broadly as transactions within 10 min. Moreover, individuals with transactions further away from the potential infector than 10 min do not seem to be materially exposed. Specifically, the estimated effect drops to 0.01%-point ( $P = 0.773$ ) when we estimate the model with a placebo measure of exposure covering transactions between 11 and 15 min before the potential infector.

These results are highly robust to nonparametric controls for observable characteristics as well as sample restrictions that further address the concern about social networks (*SI Appendix, section 4.2*). Specifically, the estimates do not change materially when we absorb differences in observable characteristics that correlate strongly with infection risks (i.e., age, income, and municipality). The estimates are also robust to excluding individuals who have no observable social link to the potential infector in the data but are more likely to have an unobserved link because their age is close to the potential infector's or because they have made in-store transactions close to the potential infector on other occasions.

Fig. 3 illustrates the differential infection dynamics of exposed relative to nonexposed individuals. By design, individuals in the estimation sample have no positive tests between day  $d - 4$  and day  $d + 2$ —if they had, they would themselves be potential infectors. We therefore construct infection variables for 5-d periods postexposure, i.e.,  $[d + 3, d + 7]$ ,  $[d + 8, d + 12]$ ,  $[d + 13, d + 17]$ , etc., and preexposure, i.e.,  $[d - 9, d - 5]$ ,  $[d - 14, d - 10]$ ,  $[d - 19, d - 15]$ , etc. When we use these infection variables as outcomes in a series of separate regressions, we find that exposed and nonexposed individuals generally followed a similar infection trajectory throughout the time window, except in the first 5-d period after the in-store exposure. Thus, the main estimates do not reflect general differences in infection risk across the two groups (*SI Appendix, section 4.3*).

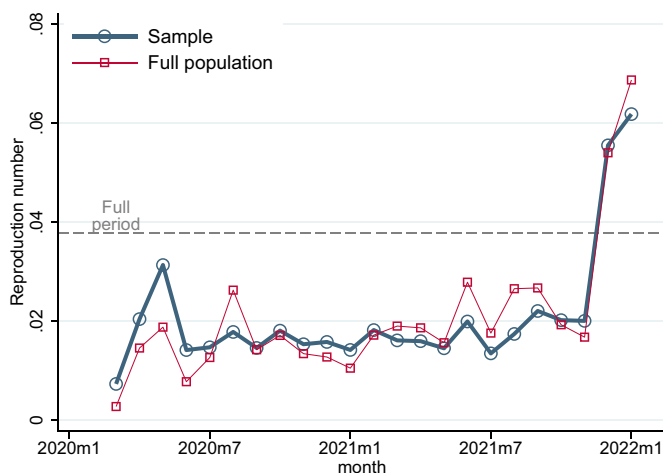
Fig. 4 shows how the baseline estimates of in-store transmission vary across four sample periods, defined by the Covid-19 variant with the highest prevalence among infected individuals in Denmark. Regardless of the specific definition of exposure, i.e., transactions within 1, 5, and 10 min of the potential infector, the qualitative pattern is the same: transmission in stores was relatively small when the Index, Alpha, and Delta variants were dominant and increased sharply when Omicron took over. The difference is robust to controlling for other factors that may affect transmission, such as seasonal variation and the age of potential infectors (*SI Appendix, section 4.4*). Analyzing heterogeneity by personal



**Fig. 4.** Heterogeneity by Covid-19 variant. The bars indicate the excess probability of testing positive between day  $d + 3$  and day  $d + 7$  associated with in-store exposure in four distinct time periods defined by the dominant variant: Index (red bars), Alpha (blue bars), Delta (green bars), and Omicron (brown bars). The first cluster of bars illustrates the estimates for the baseline definition of exposure, i.e., purchases within 5 min of a potential infector. The next two clusters illustrate the estimates for alternative definitions of exposure, i.e., purchases within 1 and 10 min of a potential infector. The estimated coefficients and SE are reported in *SI Appendix, Table S5*.

characteristics, we find that the transmission rate decreases with the age of the exposed individual, i.e., decreasing susceptibility, and increases with the age of the potential infector, i.e., increasing onward transmissibility (*SI Appendix, section 4.5*).

Finally, to gauge how much transmission in supermarkets and grocery stores contributes to the pandemic, we compute the reproduction number implied by the estimated transmission rates (*SI Appendix, section 5*). For each infected individual in our sample, we use age-dependent sampling probabilities to scale up the number of exposed individuals in the sample to the expected number in the population. We then aggregate expected exposures and multiply by the estimated transmission rates to obtain the expected number of transmissions, and divide by the number of infections in the sample to arrive at the reproduction number. Fig. 5 illustrates the results: The total reproduction number for all infected individuals in our sample is just below 0.04, with a striking increase from around 0.02 to 0.06 around the arrival of the Omicron variant. The analogous estimates for all infected individuals in the population are highly similar.



**Fig. 5.** Reproduction. The figure shows monthly estimates of the reproduction number in supermarkets and grocery stores for the infected individuals in our sample (blue line) and for the population (red line).

Our methodology for computing the reproduction number implicitly assumes that our gross sample of Danske Bank customers is representative of the overall population conditional on age. Concretely, the risk of being exposed to a potential infector in a store as well as the risk of being infected given exposure should be the same for individuals who are in our gross sample and same-aged individuals who are not. Importantly, our methodology does not assume that the estimation sample of individuals who make transactions within 30 min of a potential infector is representative of the overall population.

## Discussion

A key methodological concern is how imperfect detection of Covid-19 cases may affect the results. There are three separate mechanisms. First, there may have been potential infectors whom we do not classify as such because they did not test. This has no bearing on our estimates of in-store transmission, assuming that undetected potential infectors were equally likely to have contact with individuals classified as exposed and nonexposed. Second, there may have been in-store transmission that was not detected because the exposed individuals did not test. This implies that the true transmission rate in stores may be somewhat higher than our estimates. Third, differential testing across exposed and nonexposed, e.g., due to alerts from the health authorities' contact tracing, could potentially confound the estimates. However, we document that the two groups generally exhibit almost identical test behavior. We also show that the differential increase in infections after exposure is not accompanied by a differential increase in negative tests (*SI Appendix, section 4.6*), which seems inconsistent with a surge in testing among the exposed caused by contact tracing.

To the extent that individuals are exposed to multiple potential infectors in the same store, our estimates could, in principle, overstate the transmission risk associated with a single exposure. However, we show that, in practice, the vast majority of exposed individuals only have one potential infector within 10 min of their transaction. Moreover, the estimated transmission rate barely changes when we restrict the estimation sample to this subsample of exposed individuals (*SI Appendix, section 4.7*).

Our analysis excludes individuals who are socially connected to the potential infector as they are likely to be exposed not just inside the store but also outside. Consistent with this notion, we find very large estimates when we apply the model to exposed individuals with social links to the potential infector (*SI Appendix, section 4.8*). Clearly, these results should not be interpreted as estimates of transmission inside the store.

## Materials and Methods

**Data.** We combine microdata from three sources: Danske Bank (23), Statens Serum Institut in Denmark (24), and Statistics Denmark (32–36). We are able to match the three sources at the individual level as they use the same unique personal identifiers. For confidentiality and data protection, the data are stored in a secure environment, unique identifiers are pseudoanonymized and all identifying information removed by employees of Statistics Denmark.

From *Danske Bank*, we obtain comprehensive transaction data for each of the bank's customers for the period between 1 January 2018 and 15 January 2022.

First, we use a unique store identifier to determine the place and a time stamp to determine the time for each in-store purchase. We limit the analysis to transactions with MasterCard because transactions through other payment circuits do not provide accurate time stamps. Second, we use data on money transfers through a mobile app to identify individuals who have made transfers to one another and who are therefore likely to be socially connected. From *Statens Serum Institut*, we obtain administrative information about the Covid-19 tests performed by public and private test providers. We observe the date at which the test was performed, the type of test, and the test result. From *Statistics Denmark*, we obtain

administrative microdata from a range of government registers, which contribute further to the identification of social networks and provide detailed information about background characteristics. First, we draw on the registers for population, employment, and education to identify individuals who are socially connected through the extended family, household, workplace or educational institution. Second, we draw on the registers for population, employment, and income to obtain information about sociodemographic characteristics for each individual. We provide more details on the data in *SI Appendix, section 2*.

We create the estimation sample in three steps. First, we identify all the instances where a purchase in a supermarket or grocery store on day  $d$  was made by an individual with a positive COVID-19 test between day  $d - 4$  and day  $d + 2$ . These are potential infector transactions. Second, we identify a sample of exposed individuals, who made a purchase within 5 min of a potential infector transaction in the same store, and a group of nonexposed individuals, who made a transaction between 16 and 30 min before a potential infector transaction in the same store. Third, we stack the exposed and nonexposed samples while excluding individuals with social connections to the potential infector (i.e., individuals connected through the extended family, household, workplace, educational institution, or money transfers).

**Estimation.** Letting  $i$  denote an individual in our estimation sample and letting  $q$  denote the transaction of a potential infector that assigns individual  $i$  to the exposed or nonexposed sample, we estimate the following model:

$$infected_{i,q} = \alpha_q + \beta exposed_{i,q} + \epsilon_{i,q},$$

where  $infected_{i,q}$  is an indicator that individual  $i$  tests positive for Covid-19 between day 3 and day 7 after transaction  $q$  and  $\alpha_q$  represents a separate intercept for each potential infector transaction  $q$ . The variable of interest,  $exposed_{i,q}$ , is an indicator for individual  $i$  being exposed at transaction  $q$ . Thus, the parameter  $\beta$  captures the *differential infection rate* for the exposed, measured relative to the nonexposed individuals associated with the same infector transaction  $q$ .

We interpret  $\beta$  as the probability of transmission from the potential infector to exposed individuals. This interpretation requires two assumptions. First, we need to assume that other infection risks are uncorrelated with exposure across individuals associated with the same infector transaction (Assumption #1). This implies that individuals transacting within 5 min of a potential infector are not systematically different from individuals transacting between 16 and 30 min before the same potential infector in dimensions that correlate with background risk. Second, we need to assume that there is no transmission from the potential infector to the nonexposed individuals (Assumption #2).

We perform a range of tests that allow us to assess the validity of these assumptions. First, we estimate the model with alternative dependent variables, e.g., age, gender, and household size as well as indicators of testing and infections prior to exposure. If exposed and nonexposed individuals are highly similar in terms of characteristics, behavior, and infection outcomes that are strong correlates of background risk, it is suggestive that they are also highly similar in terms of this background risk (Assumption #1). Second, we estimate the model with a placebo definition of exposure, i.e., an indicator for making a transaction between 11 and 15 min before a potential infector. If this placebo exposure is not associated with a differential infection rate, it is suggestive that individuals transacting more than 10 min before the potential infector are effectively not exposed and, hence, that the reference group whose transactions are even longer before the potential infector's is not contaminated by exposure to the potential infector (Assumption #2). We provide more details on the empirical design in *SI Appendix, section 3*.

**Risks.** We use the estimated transmission rates to estimate the individual and aggregate risks associated with retail shopping. In either case, we start by computing a sampling probability  $\omega_i$  for each individual  $i$  in the sample of Danske Bank customers. Letting  $n_i$  and  $N_i$  denote the number of individuals at the same age as  $i$  in the bank sample and in the population respectively, we define  $\omega_i = n_i/N_i$ . Note that  $n_i$  weighs individuals by the share of MasterCard transactions in their overall transactions in grocery stores and supermarkets, e.g., an individual who uses a MasterCard for every second transaction contributes only 0.5 to  $n_i$ .

We estimate the probability of getting infected for the average store visit in our sample as the expected aggregate number of in-store transmissions  $T$  divided by the aggregate number of store visits  $V$ , i.e.,  $T/V$ , both measured in the sample of Danske Bank customers. In a first step, we estimate transmission rates while

allowing  $\beta$  to vary with the age group of the potential infector (i.e., <25 y, 25 to 45 y, and >45 y), with the age group of the exposed individual (i.e., same groups), and with the dominant variant on the day of the exposure (i.e., Omicron and other). While  $V$  is directly observable, we compute  $T$  by multiplying the inverse sampling probability of the potential infector and the relevant estimated transmission rate for each exposure and summing over all exposures. Letting  $i, j, v$  index the exposed individual, the potential infector, and the time variant, we obtain:

$$T = \sum_i \frac{1}{\omega_j} \hat{\beta}_{jiv}.$$

We estimate the number of in-store transmissions from the average infected person in our sample as the expected aggregate number of transmissions  $S$  from infected individuals divided by the aggregate number of infections  $F$ , i.e.,  $S/F$ , both measured in the sample of Danske Bank customers. While  $F$  is directly observable, we compute  $S$  by multiplying the inverse sampling probability of the exposed individual and the relevant estimated transmission rate for each exposure and summing over all potential infectors:

$$S = \sum_j \frac{1}{\omega_j} \hat{\beta}_{jiv}.$$

The assumption that our gross sample is representative in terms of the risk of being infected given exposure is reflected in the fact that we apply  $\beta$ , the transmission rate estimated in our sample, to exposed individuals outside our sample. As  $\beta$  is allowed to vary with age, we effectively assume that individuals in the gross sample are representative of same-aged individuals in the population in this dimension.

The assumption that our gross sample is representative in terms of the risk of being exposed is reflected in the fact that we scale up exposures in the sample with inverse sampling probabilities  $1/\omega$ . As  $\omega$  varies with age in the baseline estimation, we effectively assume that individuals in the gross sample are representative of same-aged individuals in the population in this dimension. In a robustness test described in *SI Appendix*, we allow  $\omega$  to vary with gender, income, region, and occupation in addition to age (*SI Appendix, section 5*).

**Ethical Approval.** As our study is based solely on administrative data processed at the servers of Statistics Denmark, this research did not require ethical approval according to the policy of the Research Ethics Committee at the Department of Economics of the University of Copenhagen. This is because Statistics Denmark ensures compliance with GDPR, which requires that the research is socially relevant and has introduced strict need-to-know principles for data analysis, as well as anonymization rules securing the anonymity of individual citizens.

**Data, Materials, and Software Availability.** The dataset combines customer data from Danske Bank and administrative data from the government registers in Denmark. Statistics Denmark merged the two data sources using social security numbers. The data is stored at the secure facility of Statistics Denmark and may not be transferred outside of this facility for security reasons. Generally, researchers interested in obtaining access to Danish administrative microdata need to submit a written application to Statistics Denmark. They can obtain access if they are themselves affiliated with a Danish research institution or if they collaborate with researchers affiliated with such an institution. Access is provided remotely through the internet. The procedure is described by Statistics Denmark at their website (37). To access the customer data from the bank, separate permissions from Danske Bank and the Ministry of Industry, Business and Financial Affairs are required. If a researcher wishes to analyze our data for replication purposes, we will assist in the process of acquiring access and provide all the programs necessary for replication. The use of the data is subject to the European Union's General Data Protection Regulation (GDPR) per Danish regulations.

**ACKNOWLEDGMENTS.** The Center for Economic Behavior and Inequality at the University of Copenhagen is supported by Danish National Research Foundation Grant DNRF134.

Author affiliations: <sup>a</sup>Saïd Business School, Oxford University, Oxford OX1 1HP, United Kingdom; <sup>b</sup>Department of Economics, University of Copenhagen, Copenhagen K 1353, Denmark; <sup>c</sup>Center for Economic Behavior and Inequality, University of Copenhagen, Copenhagen K 1353, Denmark; and <sup>d</sup>Danmarks Nationalbank, Copenhagen Ø 2100, Denmark

1. Q. Bi *et al.*, Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: A retrospective cohort study. *Lancet Infect. Dis.* **20**, 911-919 (2020).
2. S. B. Jørgensen, K. Nygård, O. Kacelnik, K. Telle, Secondary attack rates for Omicron and Delta variants of SARS-CoV-2 in Norwegian households. *JAMA* **327**, 1610-1611 (2022).
3. F. P. Lynsø *et al.*, Household transmission of the SARS-CoV-2 Omicron variant in Denmark. *Nat. Commun.* **13**, 5573 (2022).
4. J. Wu *et al.*, Household transmission of SARS-CoV-2, Zhuhai, China, 2020. *Clin. Infect. Dis.* **71**, 2099-2108 (2020).
5. Q. L. Jing *et al.*, Household secondary attack rate of COVID-19 and associated determinants in Guangzhou, China: A retrospective cohort study. *Lancet Infect. Dis.* **20**, 1141-1150 (2020).
6. H. A. Thompson *et al.*, Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) setting-specific transmission rates: A systematic review and meta-analysis. *Clin. Infect. Dis.* **73**, e754-e764 (2021).
7. K. Macartney *et al.*, Transmission of SARS-CoV-2 in Australian educational settings: A prospective cohort study. *Lancet Child Adolesc. Health* **4**, 807-816 (2020).
8. E. Larosa *et al.*, Reggio Emilia Covid-19 Working Group, Secondary transmission of COVID-19 in preschool and school settings in northern Italy after their reopening in September 2020: A population-based study. *Eurosurveillance* **25**, 2001911 (2020).
9. S. A. Ismail, V. Saliba, J. L. Bernal, M. E. Ramsay, S. N. Ladhani, SARS-CoV-2 infection and transmission in educational settings: A prospective, cross-sectional analysis of infection clusters and outbreaks in England. *Lancet Infect. Dis.* **21**, 344-353 (2021).
10. D. Hijnen *et al.*, SARS-CoV-2 transmission from presymptomatic meeting attendee, Germany. *Emerging Infect. Dis.* **26**, 1935 (2020).
11. S. Y. Park *et al.*, Coronavirus disease outbreak in call center, South Korea. *Emerging Infect. Dis.* **26**, 1666 (2020).
12. W. Zhang *et al.*, Secondary transmission of coronavirus disease from presymptomatic persons, China. *Emerging Infect. Dis.* **26**, 1924 (2020).
13. R. Lordan, G. A. FitzGerald, T. Grosse, Reopening schools during COVID-19. *Science* **369**, 1146-1146 (2020).
14. S. S. A. Karim, Q. A. Karim, Omicron SARS-CoV-2 variant: A new chapter in the COVID-19 pandemic. *Lancet* **398**, 2126-2128 (2021).
15. K. P. Hui *et al.*, SARS-CoV-2 Omicron variant replication in human bronchus and lung ex vivo. *Nature* **603**, 715-720 (2022).
16. J. R. Pulliam *et al.*, Increased risk of SARS-CoV-2 reinfection associated with emergence of Omicron in South Africa. *Science* **376**, eabn4947 (2022).
17. M. M. Böhmer *et al.*, Investigation of a COVID-19 outbreak in Germany resulting from a single travel-associated primary case: A case series. *Lancet Infect. Dis.* **20**, 920-928 (2020).
18. J. Z. Zhang *et al.*, Investigation on a cluster epidemic of COVID19 in a supermarket in Liaocheng, Shandong province. *Zhonghua Liuxingbingxue Zazhi* **41**, E055 (2020).
19. K. Danis *et al.*, Cluster of coronavirus disease 2019 (COVID-19) in the French Alps, February 2020. *Clin. Infect. Dis.* **71**, 825-832 (2020).
20. P. K. Munch, L. Espenhain, C. H. Hansen, T. G. Krause, S. Ethelberg, Societal activities associated with SARS-CoV-2 infection: A case-control study in Denmark, November 2020. *Epidemiol. Infect.* **150**, e9 (2022).
21. P. K. Munch, L. Espenhain, C. H. Hansen, T. G. Krause, S. Ethelberg, Case-control study of activities associated with SARS-CoV-2 infection in an adult unvaccinated population and overview of societal COVID-19 epidemic counter measures in Denmark. *PLoS ONE* **17**, e0268849 (2022).
22. A. E. White *et al.*, Community exposures among Colorado adults who tested positive for SARS-CoV-2—A case-control study, March–December 2021. *PLoS ONE* **18**, e0282422 (2023).
23. Danske Bank (2022). Customer transactions. Data provided by Danske Bank, Accessed on 5 April, 2022.
24. Statens Serum Institut (2022). Dataset on COVID-19 tests. <https://miba.ssi.dk/forskningsbetjening/tilgaengeligt-data>. Accessed on 9 March, 2022.
25. Danmarks Nationalbank, Kontantens anvendelse i samfundet. Analyse No. 3. NationalBank. <https://www.nationalbanken.dk/media/3ghdes5h/analyse-nr-3-kontanters-anvendelse-i-samfundet.pdf>. Accessed 4 April 2024.
26. L. W. Krogsgaard *et al.*, Seroprevalence of SARS-CoV-2 antibodies in Denmark: Results of two nationwide population-based surveys, February and May 2021. *Infect. Drug Resist.* **16**, 301-312 (2023).
27. L. Espenhain *et al.*, Prevalence of SARS-CoV-2 antibodies in Denmark: Nationwide, population-based seroepidemiological study. *Eur. J. Epidemiol.* **36**, 715-725 (2021).
28. M. Klompas, M. A. Baker, C. Rhee, Airborne transmission of SARS-CoV-2: Theoretical considerations and available evidence. *JAMA* **324**, 441-442 (2020).
29. T. Greenhalgh *et al.*, Ten scientific reasons in support of airborne transmission of SARS-CoV-2. *Lancet* **397**, 1603-1605 (2021).
30. N. Van Doremalen *et al.*, Aerosol and surface stability of SARS-CoV-2 as compared with SARS-CoV-1. *N. Engl. J. Med.* **382**, 1564-1567 (2020).
31. P. Y. Chia *et al.*, Detection of air and surface contamination by SARS-CoV-2 in hospital rooms of infected patients. *Nat. Commun.* **11**, 2800 (2020).
32. Statistics Denmark (2022a). <https://www.dst.dk/extranet/ForskningVariabellister/BEF-Befolkningen.html>. Accessed on 4 February 2022.
33. Statistics Denmark (2022b). Employment IDAN. <https://www.dst.dk/extranet/ForskningVariabellister/IDAN-IDA-ansættelser.html>. Accessed on 30 September 2022.
34. Statistics Denmark (2022c). Education UDDA. [https://www.dst.dk/extranet/ForskningVariabellister/UDDA-Uddannelser\(BUE\).html](https://www.dst.dk/extranet/ForskningVariabellister/UDDA-Uddannelser(BUE).html). Accessed on 4 February 2022.
35. Statistics Denmark (2022d). Indkomster IND. <https://www.dst.dk/extranet/ForskningVariabellister/IND-Indkomst.html>. Accessed on 4 February 2022.
36. Statistics Denmark (2022e). Job Classifications AKM. <https://www.dst.dk/extranet/ForskningVariabellister/AKM-Arbejdsklassifikationsmodulet.html>. Accessed on 4 February 2022.
37. Statistics Denmark, Access to Data. Statistics Denmark. <https://www.dst.dk/en/TilSalg/Forskningservice/Dataadgang>. Accessed 5 April 2024.