🔓 OPEN ACCESS

Vaccination, immunity, and the changing impact of COVID-19 on infant health

Florencia Torche[a,1] and Jenna Nobles[b]

In utero exposure to COVID-19 infection may lead to large intergenerational health effects. The impact of infection exposure has likely evolved since the onset of the pandemic as new variants emerge, immunity from prior infection increases, vaccines become available, and vaccine hesitancy persists, such that *when* infection is experienced is as important as *whether* it is experienced. We examine the changing impact of COVID-19 infection on preterm birth and the moderating role of vaccination. We offer the first plausibly causal estimate of the impact of maternal COVID-19 infection by using population data with no selectivity, universal information on maternal COVID-19 infection, and linked sibling data. We then assess change in this impact from 2020 to 2023 and evaluate the protective role of COVID-19 vaccination on infant health. We find a substantial adverse effect of prenatal COVID-19 infection on the probability of preterm birth. The impact was large during the first 2 y of the pandemic but had fully disappeared by 2022. The harmful impact of COVID-19 infection disappeared almost a year earlier in zip codes with high vaccination rates, suggesting that vaccines might have prevented thousands of preterm births. The findings highlight the need to monitor the changing consequences of emerging infectious diseases over time and the importance of mitigation strategies to reduce the burden of infection on vulnerable populations.

infant health | COVID-19 infection | vaccination | effect heterogeneity | administrative linkages

The COVID-19 pandemic has imposed a large burden on population health and well-being, resulting in enormous loss of life and the onset of long-term illness and disability. Though COVID-19 response efforts were initially focused on older adults with high mortality risk, evidence suggested that infection was potentially costly for younger individuals as well, with documented associations for cardiovascular health, cognition, physical ability, and reproductive health (1–3). By fall 2020, early evidence on COVID-19 risk in pregnancy had emerged, suggesting that infection may adversely affect maternal and infant health (4). Studies now demonstrate a correlation between maternal COVID-19 infection and higher risk of preterm birth and neonatal intensive care unit (NICU) hospitalization in multiple US populations (5–13). These findings raise particular concern because they indicate a health effect that crosses generations and may be among the most enduring legacies of the pandemic.

Understanding the broader health effects of COVID-19, like those occurring during pregnancy, is hindered by three prevailing challenges of infectious disease surveillance. One, infection risk is not randomly distributed in populations. During the first 18 mo of the COVID-19 pandemic, infection during pregnancy differentially occurred among the least resourced individuals (14). As a result, research must grapple with the bias introduced by selective infection risk—including patterns of selection that shift over time. Two, infection measurement varies across people, regions, and time; some COVID-19 infections are asymptomatic and rarely documented, and testing requirements change over time (15, 16). Three, viruses continually evolve, as do the tools to combat them. Shifting variants, alongside preventive and palliative innovations, can alter the impact of infection, such that *when* infection is experienced might be as important as *whether* it is experienced.

This temporal dimension is particularly relevant in the COVID-19 case. The first documented SARS-CoV-2 variant spread in the US between February 2020 and June 2020, followed by the Alpha and Epsilon variants in early 2021, the Delta strain which became universal in August 2021, and the Omicron variant, the predominant strain in the United States since December 2021 (17, 18). Variants appear to have had distinct morbidity and fatality rates (19, 20). The Delta strain was more dangerous for the health of pregnant people than Alpha or Epsilon (21) and led to more severe illness than Omicron (22). Omicron's association with lower probability of death and hospitalization resulted from multiple factors, including increased population immunity from prior infection and vaccination, lower intrinsic virulence of this variant, and Omicron's higher immune-evasion

Significance

The effects of COVID-19 on infant health may be among the most enduring legacies of the pandemic. Using linked population-level data on siblings born between 2014 and 2023 in birthing facilities with confirmed universal testing, we establish that maternal COVID-19 infection during pregnancy causally, and substantially, increased the risk of preterm birth—an infant outcome with lifelong consequences for health and socioeconomic well-being. We show that this effect disappeared by 2022 and demonstrate that the disappearance of this effect happened almost a year earlier in places that were early adopters of COVID-19 vaccination. The availability of vaccines and the decision to use them may have reduced a serious health burden for the next generation of US children.

Author affiliations: ªDepartment of Sociology, Stanford University, Stanford, CA 94305; and ᵇDepartment of Sociology, Center for Demography and Ecology, University of Wisconsin-Madison, Madison, WI 53706

capabilities. The latter likely resulted in a positively selected infected population—that is, individuals with stronger immune responses were more likely to be infected with Omicron than prior variants (23–27). The management of COVID-19 has also changed markedly; the development of vaccines and antiviral therapies was a key turning point in COVID-19 management. Vaccination is efficacious in averting symptomatic infection, severe illness, and death in pregnant populations (28, 29) and may confer protection in utero (30, 31). Vaccines became available to all people aged 16 or older in late spring of 2021, with significant variation in take-up across communities and social groups. By March 2023, about 70% of pregnant people had received two doses of COVID-19 vaccination (32). Intravenous antiviral therapy emerged in fall 2021, followed by oral antivarial medication, which became widely available in 2022.

We analyze linked administrative natality and vaccination data with econometric tools to address the three challenges to infectious disease surveillance: patterned infection risk, incomplete testing, and effect variation alongside evolving immunity, strains, and mitigation measures. With complete population data, we evaluate the changing effects of COVID-19 and accompanying protective measures on infant health. We focus on preterm birth as a marker of infant health, given existing evidence of a correlation between maternal COVID-19 and preterm birth (5–9), and the importance of preterm birth for population health and intergenerational health and socioeconomic disparities. Preterm birth is the main predictor of infant morbidity and mortality in the United States (33) and shapes long-term markers of well-being, including educational attainment and earnings. Preterm birth is also deeply unequally distributed, with African American and low-income infants experiencing particularly high risks (34, 35).

We use continuous-release restricted-access natality data for California, a large and diverse state of 40 million people that contributes 12% of all US births. Restricted natality data offer nearly universal information on live births with no sample selectivity, provide detailed location information of both the residence of the birthing person and birth facility, document maternal infection at the time of delivery hospitalization for all births, and allow the identification of siblings (births of the same mother) born between 2014 and 2023. Matched-sibling data support a fixed-effects approach comparing "treated" infants exposed to COVID-19 infection in utero with their "untreated" siblings. This approach accounts for all time-invariant unobserved characteristics of mothers—the primary threats to causal identification in all prior research—substantially advancing the causal interpretation of the association between COVID-19 infection and infant health. We then evaluate how the effects of maternal infection have evolved over time and consider adjudicative evidence on the uptake of COVID-19 vaccination, using small-area-level data on vaccination rates. We demonstrate that COVID-19 infection in pregnancy had a large, causal effect on preterm birth in the first year of the pandemic. We then show that effect disappeared entirely by fall 2022. While the shift from Delta to Omicron strands and rising infection-based immunity appears to have played a role, vaccination uptake appears to have had a substantial protective impact.

At present, a long-lasting intergenerational health impact of the pandemic has been at least temporarily curtailed. However, the emergence of new COVID-19 subvariants with higher capability to evade immunity and neutralize antibody therapy (36–38) as well as the cyclical change in virulence in other viral infections such as influenza (39) suggest that the threat of COVID-19 infection on infant health is likely not over. In fact, the evidence presented here underscores the importance of implementing protective
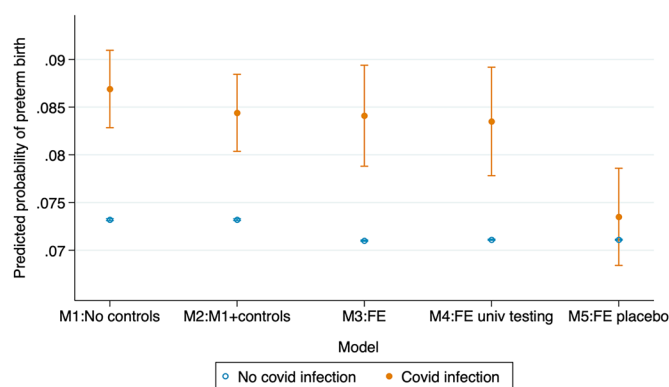

Fig. 1. Statistical models for the association between COVID-19 infection during pregnancy and preterm birth. California births July 2020 to February 2023. Solid and hollow circles are parameter estimates, vertical lines are 95% CIs. Models 1 to 2 based on linear probability models with birth facility and birth month fixed effects. Model 2 includes controls for maternal sociodemographic characteristics [educational attainment (less than high school diploma, high school graduate, some college, bachelor degree (BA), graduate degree), race/ethnicity (NH Black, NH White, Hispanic, Asian, Other), foreign-born status, age (less than 20 y of age, 20 to 24, 25 to 29, 30 to 34, 35 to 39, 40, and older), quartile of zip code of residence socioeconomic (SES) disadvantage, and parity (1st, 2nd, 3rd, or higher order)] and risk factors for preterm birth (prepregnancy smoking, prepregnancy hypertension, prepregnancy diabetes, asthma, large fibroid tumors). Model 3 uses a within-mother fixed-effects estimator comparing sibling(s) exposed to maternal COVID-19 infection during pregnancy with unexposed sibling(s). Controls for mother's educational attainment, age, quartile of the zip code of residence SES disadvantage, birth order, and month of birth included. Model 4 uses a mother fixed-effects estimator and restricts the sample to birth facilities with confirmed COVID-19 universal testing at the time of delivery. Model 5 is a falsification fixed-effects test based on a randomly selected sample of births with identical birth date distribution as the treated siblings and compared them to their (mostly older) siblings. Robust SEs clustered by birthing person. Models 1 to 2 use data from January 2018 to February 2023, models 3 to 5 use data from January 2014 to February 2023 to allow for a wider interpregnancy interval. Births occurring March 2020 to June 2020 excluded from the analysis due to unobserved maternal COVID-19 infections. Sample restricted to singleton births. Source: California Natality Data.

measures, especially vaccination, as these changes in the virus continue to occur. We conclude by discussing ways in which administrative data can be leveraged to assess the changing health effects of COVID-19 and other emerging infectious diseases to support a nimble response strategy as the risks associated with infection change over time.

Results

Fig. 1 shows regression estimates of the average impact of maternal COVID-19 infection on preterm birth over the period from July 2020 to February 2023 (parameter estimates and statistical significance testing in *SI Appendix,* Table A1). Model 1 uses the complete natality sample and includes only birth facility and month fixed effects—conceptually similar to dichotomous indicators for each birth facility in the sample and for each month between January 2014 and February 2023 (see *Materials and Methods* section for description of data and model specification). Estimates capture the within-facility and within-month difference in the probability of preterm birth among infected and uninfected populations, accounting for birth facility confounders (e.g., differences in COVID-19 testing protocols, labor and delivery protocols, socioeconomic composition of patients), and time-of-birth confounders (e.g., trends in COVID-19 infection and temporal changes in testing shared across facilities.)

Based on model 1, maternal COVID-19 infection at the time of delivery is associated with a large increase in the probability of preterm birth by 1.4 percentage points from 7.3 to 8.7%, equivalent to a 29% increase. Model 2 adjusts the estimate for a large

set of observed maternal sociodemographic and risk factors. The estimated effect of COVID-19 infection declines marginally compared to model 1 to 1.1 percentage points (from 7.3 to 8.4%), a 15% increase in the probability of preterm birth.

The introduction of within-sibling comparisons through mother fixed effects (Model 3) yields a comparable estimate: COVID-19 infection during pregnancy increases the probability of preterm birth by 1.3 percentage point—from 7.1 to 8.4%. Restricting the sample to facilities with documented universal COVID-19 testing at time of birth (Model 4) yields estimates nearly identical to that of the full sample of facilities.

Estimation based on sibling comparisons provide the strongest evidence currently available that COVID-19 infection during pregnancy negatively affects infant health. Restricting the sample to birth facilities with documented universal testing at the time of delivery avoids biases emerging from absent or selective testing. Relying on the preferred estimate—a within-mother comparison restricted to facilities with documented universal COVID-19 testing—maternal COVID-19 infection increases the probability of preterm birth by 1.2 percentage points, from 7.1 to 8.3%. This effect is roughly equivalent to in utero exposure to a 9 percentage point increase in the area-level unemployment rate (40) or to high-intensity wildfire smoke for 20 d (41)—an enormous impact. Because this estimate captures infection at the time of delivery hospitalization, it potentially underestimates the effects of infection any time in pregnancy (see *Discussion* section). This estimated effect of COVID-19 infection is not driven by unobserved characteristics of the birthing persons, nor it is an artifact of a correlation between COVID-19 infection and maternal age or birth order: a falsification test implemented in model 5 that uses a randomly-selected group of children with an identical birth date distribution to the treated siblings in model 4 and compares them to their siblings yields an insignificant estimate close to zero.

Critically, the negative impact of COVID-19 infection is not restricted to late preterm births around the 37 wk of gestation cutoff. We observe a similar negative effect for very preterm birth (births occurring <32 wk of gestation), a more severe condition identifying infants at highest risk of mortality, morbidity, and developmental difficulties (42, 43). (*SI Appendix*, Fig. A1).

Change in the Impact of Maternal COVID-19 Infection Over Time. Rising population immunity, change in variant virulence, and the emergence of vaccines and therapeutic responses may have altered the impact of COVID-19 infection on infant health as the pandemic unfolded between 2020 and 2023, the period captured in Fig. 2. Mother fixed-effects models stratified by time of birth examine variation over the course of the pandemic (Fig. 2, parameter estimates and statistical significance testing in *SI Appendix*, Table A2).

The harmful effect of COVID-19 infection was large from July 2020, the earliest stage of the pandemic for which data are available, to the end of 2020, with an estimated increase in preterm birth of 5.4 percentage points (from 6.9 to 12.3%, a 78% increase). After this initial period, the detrimental impact of maternal COVID-19 infection fluctuates between 2 and 4 percentage points in 2021, with a peak of 4.1 percentage points during the fall of 2021 when the virulent Delta variant had become universal. The impact of COVID-19 infection disappears in 2022. No indication of change is observed thereafter, even during the surge in Omicron infections in the summer of 2022.

The large impact of COVID-19 infection on infant health in the earliest stage of the pandemic is likely due to a combination of factors, including the detrimental impact of viral infection in
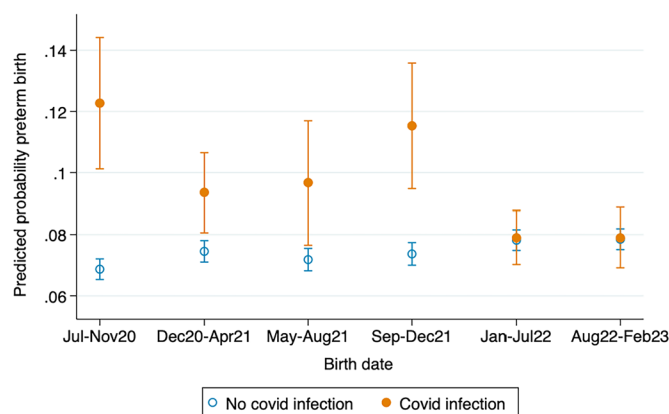

Fig. 2. Sibling fixed-effects models for the effect of COVID-19 infection during pregnancy on preterm birth (<37 wk of gestation) over the course of the pandemic. California births July 2020 to February 2023. Solid and hollow circles are parameter estimates, vertical lines are 95% CIs. Fixed-effects model comparing sibling(s) exposed to maternal COVID-19 infection during pregnancy with unexposed sibling(s). Controls for mother's educational attainment, age, quartile of zip code of residence SES disadvantage, birth order, and year of birth included. Robust SEs clustered by birthing person. Births occurring March 2020 to June 2020 excluded from the analysis due to unobserved maternal COVID-19 infections. Sample restricted to singleton births. Source: California Natality Data.

a context where vaccines and therapeutic alternatives were not yet available and immunity via prior infection was limited. It is also possible that the large impact was shaped by pandemic-related changes to obstetric practices such as labor induction or cesarean delivery intended to manage risk in overwhelmed healthcare settings. However, the fact that a large impact of COVID-19 infection early in the pandemic is also observed for very preterm birth <32 wk of gestation (*SI Appendix*, Fig. A2)—a serious condition with substantial risks for surviving newborns—suggests that the COVID-19-induced shortening in gestational age is not due to nonessential interventions designed to reduce COVID-19 spread—e.g., inductions to reduce the number of people in the labor and delivery unit. Indeed, medically unwarranted induction is highly unlikely to be used to deliver prior to 32-wk gestation.

The decline and then disappearance of an adverse impact of COVID-19 on preterm birth in 2022 occur in the context of growing infection-based immunity, the potentially reduced virulence of the Omicron variant, increased uptake of COVID-19 vaccination, and growing access to therapeutic management of COVID-19 infection. Our ability to adjudicate across these mechanisms is limited due to the absence—to date—of individual-level vaccination information among the pregnant population. However, we exploit trends in vaccine uptake at the most granular level currently available—the zip code area of residence of the birthing person—to gauge the potential impact of vaccination as a moderator of the impact of COVID-19 infection on infant health. We merge natality data with zip code–specific daily data on vaccination uptake compiled by the California Department of Public Health and available since January of 2021.

Fig. 3 shows trends in the proportion of zip code residents fully vaccinated with the initial series by quintiles of vaccination rates. Vaccination rates increased sharply in the spring and early fall of 2021, plateauing at around 70% for the entire population by March 2022. Quintiles divide population-weighted zip code areas into five equal-sized groups ranked according to the level of COVID-19 vaccination achieved by March 2022. Substantial variation exists across locations. Zip code areas with the highest vaccination rates reached a vaccination rate of 86% in March 2022. In contrast, the vaccination rate reached a maximum of
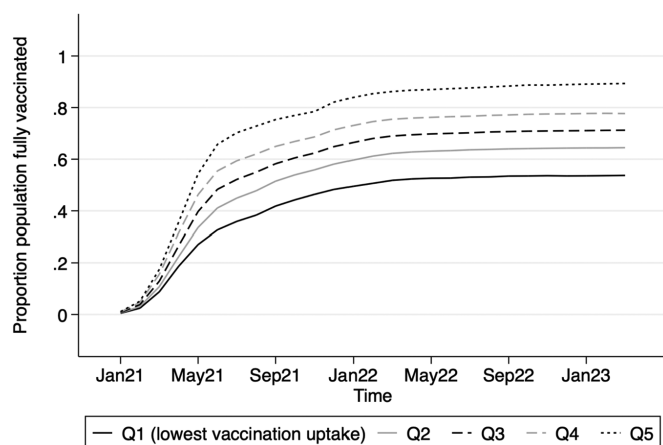
Fig. 3. COVID-19 vaccine uptake across zip code areas in California January 2021 to February 2023. Vaccine uptake defined as proportion of the population fully vaccinated (total number of people fully vaccinated divided by total population). Zip codes were sorted by vaccine uptake rate achieved by March 2020 and divided into four equal groups weighted by zip code population size. Fully vaccinated individuals defined as those with 2 Pfizer doses>=17 d apart, 2 Moderna doses>=24 d apart, 1 dose of J&J, a combination of Pfizer and Moderna doses>=17 d apart, three or more vaccination records, or only one dose in IRIS labeled as dose number 2. Source: COVID-19 Vaccine Progress Dashboard Data by zip Code, California Department of Public Health.

only 51% in the lowest-vaccination quintile. During the first 6 mo after vaccines became widely available—summer and fall 2021—population vaccination rates in higher-uptake zip codes (~70% vaccinated) were nearly double the rates in lower-uptake regions (~35% vaccinated).

We compare the impact of maternal COVID-19 infection on preterm birth for zip codes in quintiles with the lowest and highest vaccination rates. Fig. 4 shows marked differences across vaccination-rate quintiles. The impact of COVID-19 is similar across high- and low-vaccination areas until May 2021. Given

that vaccines were not available until the spring of 2021, this result is consistent with the hypothesis that vaccines reduced the harmful infant of COVID-19 infection on infant health. After the spring of 2021, the impact of COVID-19 infection fully disappeared in the highest-vaccination quintile but remained large and statistically significant in the lowest-vaccination quintile, to disappear only in early 2022 (parameter estimates and statistical significance testing in *SI Appendix,* Fig. A3 and Table A3 shows trends in middle quintiles).

The similarity across quintiles in the impact of COVID-19 infection before vaccines were available suggests that the reduction in the harmful effect of COVID-19 infection is due to vaccination uptake rather than alternative health-protective behaviors undertaken by pregnant persons residing in high-vaccination locations. Any other explanation for this impact would have to be consistent with similar associations between infection and preterm birth across zip codes prior to May 2021 and a discontinuous gap between high- and low-vaccination zip codes that widened substantially beginning in May 2021. In the low-vaccination quintile, the association between maternal COVID-19 infection and preterm birth disappeared only in January 2022, close to a year later than the high-vaccination locations. It is likely that the 51% population vaccination rate achieved in these locations was complemented by the less virulent Omicron variant, growing immunity based on prior infection, and availability of therapeutic antiviral treatment to reduce the impact of COVID-19 infection on infant health.

Discussion

Emerging infectious diseases such as COVID-19 are likely to have long-term intergenerational health effects if exposure occurs before birth. These effects are likely to change over time alongside the evolution of viral strains and the emergence of infection-mitigation strategies. Studies to capture these impacts are usually based on
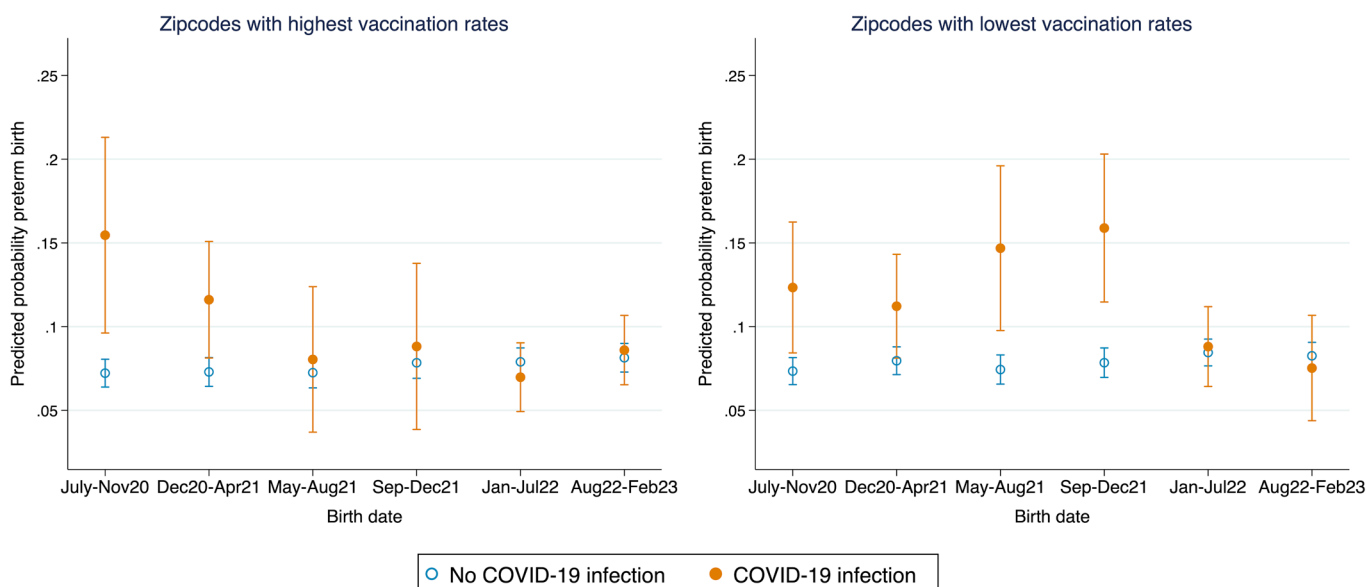

Fig. 4. Sibling fixed-effects models for the effect of COVID-19 infection during pregnancy on preterm birth over the course of the pandemic by quintile of vaccination rates at the zip code level (top and bottom quintiles). California births July 2020 to February 2023. Solid and hollow circles are parameter estimates, vertical lines are 95% CIs. Fixed-effects model comparing sibling(s) exposed to maternal COVID-19 infection during pregnancy with unexposed sibling(s), Controls for mother's educational attainment, age, quartile of zip code of residence SES disadvantage, birth order and year of birth are added. Separate models estimated for *Top* (highest vaccination rates) and *Bottom* (lowest vaccination rates) quintiles. Vaccination rate quintiles based on the proportion of population fully vaccinated by zip code area of residence. Births occurring March 2020 to June 2020 excluded from the analysis due to unobserved maternal COVID-19 infections. Robust SEs clustered by birthing person. Sample restricted to singleton births. Source: California Natality Data.

small and selected samples subject to untractable sources of bias or are only able to detect consequences of infection with lags of months, or even years, due to restricted data availability. We show how pairing administrative data and a within-person longitudinal study design can be deployed to provide near real-time evidence on the causal impact of infection—and how this impact evolves as contextual factors change.

These findings provide the first evidence on the changing impact of COVID-19 infection over time and offer a plausibly causal approach, building on prior static, associational evidence, including our own earlier research (14). Here, we have demonstrated that COVID-19 infection in pregnancy was a large risk factor for preterm birth during 2020 and 2021 and that this risk has largely disappeared. This is a profound change and represents thousands of averted preterm births and accompanying complications for those children and their families. It also represents substantial averted health care expenditures; each preterm birth incurs costs that exceed eighty thousand 2023 dollars per infant (44). We advance prior research on COVID-19 and infant health by using a within-person design—here, a within-mother design comparing siblings. The approach reduces the potential confounding bias introduced by unmeasured characteristics of mothers and their communities that plague earlier research. The strengths and limitations of such designs are well established by the methodological literature (45, 46). The main barrier to their use in understanding the impact of COVID-19 or other infectious diseases is the timely availability of administrative data with identifying individual information. These data are, however, regularly produced by state-level public health agencies around the country.

Our estimates from a mother fixed effect specification are similar in magnitude and direction to estimates from controlled regression specifications that exclude the within-person design element. This similarity could be interpreted as suggesting that more advanced identification strategies are unnecessary to capture the true health burden of infection. We strongly caution against of that interpretation. The observed similarity is not a necessary or enduring pattern and instead depends upon how the correlates of infection risk and adverse infant health change over time and how well they can be measured in standard sociodemographic regression controls. The fact that in this particular historical moment two estimators with distinct ability to account for confounding yield similar estimates does not generalize to other outcomes or contexts.

As described, the COVID-19 infection data reflect hospital-administered tests at the time of delivery, potentially rendering the effects detected here underestimates of the true effects of COVID-19 infection at any time in pregnancy. To date, research on the timing of infection as a prognostic factor is preliminary—largely because no pregnant cohort has yet to be tested continuously over pregnancy. Recent research suggests that while COVID-19 infection during the third trimester of gestation predicts an increased risk of preterm birth, first and second trimester infection is not associated with adverse pregnancy outcomes (47). However, some evidence and theoretical models about viral mechanisms suggest early-pregnancy infection might also increase preterm birth (48, 49). In that case, given that a repeated infection over the course of the pregnancy is extremely rare, early-pregnancy infection would be a confounder in our analysis: it both reduces the probability of infection at the time of birth and increases the probability of preterm birth. Only a cohort of pregnant persons with repeated testing over the pregnancy will provide adjudicative evidence.

A related issue arises if COVID-19 infection increases the risk of miscarriage or stillbirth. This study examines live births because

population data on miscarriage does not exist in the United States and data on stillbirths are released with a much longer, multiyear lag. Evidence is mixed on whether COVID-19 infection increases miscarriage early in pregnancy (50); evidence on increased risk of stillbirth (>20 wk gestation) is more consistent (51). For the present study, if pregnancies that succumb to COVID-19 also had increased risk for preterm birth, we would underestimate the effect of COVID-19 infection during pregnancy on preterm birth. Furthermore, we would underestimate the difference in the effects of infection between higher- and lower-vaccinated populations. Evidence to date suggests that the mechanisms by which COVID-19 can increase the risk of stillbirth—such as placentitis and placental insufficiency—predominately occur in unvaccinated populations (52).

Our findings similarly support a beneficial effect of vaccination in severing the association between infection exposure and infant health outcomes by using local area-level vaccination rates as a proxy for individual-level vaccination status. The main vulnerability of this approach is that zip codes with high vaccine uptake may have experienced other changes that confound the effect of vaccination—for example, selective migration patterns or early adoption of health-protective behaviors by pregnant persons residing in high-vaccination areas. For these changes to explain the observed departures between high- and low-vaccination locations, however, they must have implausibly commenced at exactly the same time COVID-19 vaccine roll-out occurred. Ancillary evaluation of trends in prenatal care initiation across locations shows no indication of such discontinuous change (*SI Appendix*, Fig. A5). Additionally, our estimates capture both the impact of individual-level vaccination and any benefits emerging from community vaccination-induced immunity at the local level. To distinguish these pathways, individual-level vaccination data among pregnant populations should be used to identify individual and aggregate protective effects of vaccination.

A further consideration is whether findings obtained in California can be extended to the rest of the country. COVID-19 infection rates in California match the rest of the country, suggesting similar trends in infection-based immunity. While vaccination rates in California are somewhat higher than for the entire United States—by March 2022, when vaccination rates plateaued, the proportion of fully vaccinated residents was 70.8% compared to 65.3% at the national level (53)—there is no reason to expect variation in the *impact* of vaccination on infant health. Still, important demographic, social, and institutional sources of heterogeneity are likely to exist across and within states. A robust data infrastructure based on administrative sources would support analysis of such variation.

In spite of these limitations, the findings indicate that vaccination is a critical point of intervention for public health policy. Vaccine avoidance is higher in pregnancy than in the general public, in part because of the absence of initial COVID-19 trials among pregnant persons, early mixed messaging on the part of public health officials, and hesitation on the part of some healthcare providers. Though evidence to date indicates no major health risks to pregnancies of vaccination, surveys of patients indicate the main barrier to vaccination is concerns about adverse consequences for the fetus and, in case of people intending pregnancy, concerns about a negative impact on fertility (54–57). The evidence here suggests substantial health costs of vaccine avoidance in pregnancy —a finding that is distinct from the observation that vaccines are safe in pregnancy—and provides relevant information for reproductive-age populations and their health care providers.

We have established that an important health legacy of COVID-19 disappeared by 2022. We caution that risk of infection during pregnancy for preterm birth will likely change again

in the future. Even if the adult population in California had reached near-universal prevalence of COVID-19 antibodies by the second trimester of 2022, which should confer some protection (58), emerging strains have shown enhanced ability to evade immunity and neutralize antibody therapy (36–38). In addition, the share of pregnant people who are protected by vaccination may also change. Although about 70% of pregnant people in the CDC's Vaccine Safety Datalink surveillance system had achieved the two-dose sequence of COVID-19 by December 2022, the true rate of vaccination is likely lower because this surveillance system disproportionately includes people with health insurance, underrepresenting marginalized populations (59). Uptake has stagnated at least since early 2022 (32), and rates of boosting have lagged, with only 21% taking the bivalent booster, including only 12% of Black pregnant people and 13% of Hispanic pregnant people (32). Though the initial vaccine series provided protection against newer variants in pregnant populations, evidence suggests that the booster dose significantly improves protection against serious adverse health effects (perinatal complications, severe maternal complications, ICU admission, death) of exposure to Omicron variants during pregnancy (60, 61), raising protection from 48% with the initial series alone to 76% with an additional initial booster (31). The reduction in vaccination uptake may contribute to an increasingly vulnerable population of pregnant people in the future.

More broadly, we have demonstrated an approach to study evolving infectious disease that could be used in the future, in the context of COVID-19 and other emerging infections, and for a wide array of health outcomes that are well-measured in vital statistics, health records, and other administrative data. Leveraging linked administrative data not only reduces the potential bias associated with nonrandom inclusion and infection risk that is present in many studies, but it also typically provides a timely way to evaluate population welfare. While compilation of data at the federal level is a lengthy process in the United States, all US states produce vital records plausibly available for use in near-real time. Many forms of electronic health records are similarly accessible with minimal delay. In light of the speed at which infectious diseases and other threats to population health evolve, timely released scientific evidence may support better targeted and prioritized resources for response to emergent health crises.

Materials and Methods

Data. Analysis relies on restricted-access natality microdata based on birth certificates including all births in California between January 2014 and February 2023. The study was approved by the Institutional Review Boards of Stanford University (Protocol # 56652), the University of Wisconsin, Madison (Study # 2020-0950-CP003), and the California Department of Public Health (Project # 2020-172). Natality data contain information about exact date of birth, infant characteristics (gestational age, birth weight, etc.), and characteristics of the person giving birth (age, race/ethnicity, educational attainment, zip code of residence, etc.). The analytic sample is restricted to singletons to limit the influence of other determinants of preterm birth that accompany multiple birth pregnancies.

Measuring maternal COVID-19 infection. All analyses of the population health effects of COVID-19 infection are thwarted by missing information on infected people who are not tested. Starting in late June 2020, the California birth certificate recorded confirmed and presumed cases of COVID-19 infection among all persons giving birth, providing an exceptional source of universal surveillance. Though testing could plausibly have occurred at any point of the pregnancy, recording protocols at the hospital level indicate that COVID-19 diagnoses are virtually always based on universal screening at admission to hospital. Analysis of trends comparing COVID-19 cases for persons giving birth with trends for the general population shows a strong temporal overlap supporting this assumption (Fig. 5). As a result, the birth records largely capture variation in infection at the
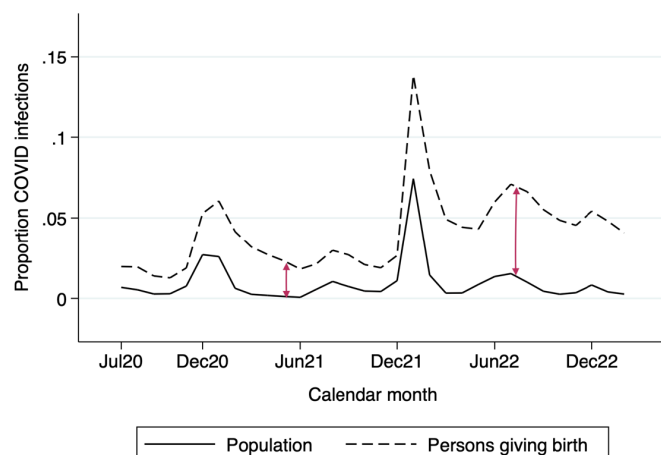

Fig. 5. COVID-19 infection among persons giving birth and general population, California July 2020 to February 2023. Trends in COVID-19 infection among persons giving birth (dashed line) and the general population (solid line), July 2020 to February 2023. Sources: Data for persons giving birth are from the California Department of Public Health; data for the general population are from The New York Times COVID-19 data repository (https://github.com/nytimes/COVID-19-data).

time of delivery hospitalization. From the perspective of infant health, this is an important point in the pregnancy in which maternal COVID-19 infection operates, likely through placental mechanisms (62, 63). It is also the only moment in pregnancy in which complete population information about COVID-19 infection is possible in *any* setting. Fig. 5 also reveals a growing divergence in infection rates between the natality data and population-level data based on selective testing and voluntary reporting compiled by the CDC (signaled by vertical arrows), which supports the importance of universal testing for infection surveillance.

We implement two strategies to alleviate bias from undetected COVID-19 infections. First, all analyses exclude births occurring between March 2020 and June 2020 in which maternal COVID-19 infection was not yet included in the birth certificate, which would lead to false negatives among infected people. Second, we successfully contacted 93% of the approximately 400 hospitals and birthing units in California and were able to ascertain whether universal testing on admission was administered and if so when it started (and ended) for 92% of births occurring since July 2020, when birth certificates started including this information. It was ascertained that 85% of births occurred in a facility with confirmed universal testing at the time of birth. We use this dataset to test the robustness of results by limiting the sample to births that occurred in a facility with confirmed universal testing at time of delivery.

Analytical Strategy. We evaluate the impact of maternal COVID-19 infection on the probability of preterm birth by estimating a series of regression models that account for a growing portion of selection bias. Model 1 captures the association between maternal COVID-19 infection and preterm birth using a linear probability model adjusting only for birth facility fixed effects and month fixed effects. Facility fixed effects account for differences in COVID-19 testing protocols, labor and delivery protocols, and other institutional characteristics that might shape preterm birth, as well as for sociodemographic features of catchment communities for each birth facility. Month fixed effects account for trends in COVID-19 infection and temporal changes in testing shared across facilities. Model 2 adds controls for a large set of potential confounders including mother's sociodemographic characteristics (age, educational attainment, race/ethnicity, SES disadvantage of zip code of residence, and parity) and the leading risk factors for preterm birth (maternal hypertension, diabetes, prior preterm birth, large fibroid tumors, asthma, and smoking) (see *SI Appendix* for detailed variable description).

Models 1 and 2 mimic current research assessing the relationship between COVID-19 pregnancy infection and infant health. The next set of models use a siblings fixed-effect approach to identify a plausibly causal effect of COVID-19 infection. Mothers with more than one birth during the 2014 to 2023 period were identified by their first name and surname, year and month of birth, race/ethnicity, and foreign-born status. Because records of maternal first and last names

might include slight spelling differences across births, we used an automatized algorithm to group similar names, allowing differences by one character (a less restrictive linking algorithm yields similar results, *SI Appendix*, Fig. A4). Model 3 relies on a within-mother estimator; identification derives from differences in outcomes between siblings. This strategy accounts for all maternal characteristics that are enduring across pregnancies such as health endowments, personality, and ability. Additionally, to account for changes in the mother's circumstances across births, fixed-effects models adjust for changes in mother's age, birth order, educational attainment, and a measure of the socioeconomic status of the zip code of residence. Models use robust SEs clustered at the mother's level. (see *SI Appendix* for statistical formulation of model).

To address bias in effect estimation resulting from missing infections among mothers who were not tested, Model 4 restricts the fixed-effects analysis to the 85% of births occurring in facilities with confirmed universal testing at the time of delivery. An additional threat to validity of the fixed-effects estimator is that siblings exposed to COVID-19 infection are overwhelmingly younger and higher birth order than control siblings. As a result, the estimated impact of COVID-19 infection might be an artifact of the effect of birth order. Model 5 implements a falsification exercise to address this potential source of bias. We randomly selected a group of siblings with identical birth date distribution as the treated siblings and compared them to their (mostly older) siblings.

We then evaluate changes in the impact of COVID-19 infection over the course of the pandemic by distinguishing six time periods: July to November 2020, December 2020 to April 2021, May to August 2021, September to December 2021, January to July 2022, and August 2022 to February 2023 and run separate fixed-effects models stratifying the sample by these periods. The first two periods precede the emergence of the Delta variant and the availability of vaccines. Period 3 between May and August 2021 marks the widespread rollout of vaccines to reproductive-age adults, and period 4 (September to December 2021) captures the dominance of Delta. Periods 5 and 6 (January 2022 to February 2023) comprise the recent time of Omicron dominance.

The data structure used for this analysis includes births nested within individuals, analogous to a more common panel data structure in which time observations are nested within individuals. Our dataset is an unbalanced short panel in which a small number of births are observed for large cross-section of individuals (births per birthing person range from 1 to 8). Individuals experience the treatment, COVID-19 infection, at different times and individuals can switch in and out of treatment (e.g., birthing people can have an early birth while infected with COVID-19 and a later, noninfected birth). The two-way fixed effect (TWFE) model estimated here is a generalization of the simple 2x2 difference-in-differences (DID) model with two treatment assignment groups (treated and control individuals) and two treatment timing groups (pretreatment and posttreatment).

While the simple DID formulation relies on the "parallel trend assumption," unbiasedness of the TWFE estimator requires an additional assumption, namely that the treatment effect is constant over time and across groups. If the effect of the treatment varies over time, bias will ensue due to what the literature has called "forbidden comparisons" i.e., using already-treated observations as counterfactual for later-treated observations (64, 65). The scope for this source of bias should be minuscule in the setting under analysis, where treatment is maternal COVID-19 infection during pregnancy given that very few birthing people experienced an infection in more than one birth or experienced a birth in the noninfected condition after a prior, infected, birth. To empirically evaluate the potential impact of bias, we restricted the analytical sample to individuals for whom treatment status is absorbent across births–i.e., where unexposed births precede but do not follow exposed births. This achieves a staggered treatment allocation and removes the possibility that already-treated observations are used as counterfactuals for later-treated observations. This resulted in a reduction in the number of individuals by 0.09% (0.18% of births). Analyses using this restricted sample yield identical (at the .001 decimal point) estimates and tests of statistical significance, suggesting that this source of bias is irrelevant in this case.

Data, Materials, and Software Availability. This information is from the records of the California Department of Public Health. The analyses, interpretations, or conclusions expressed herein represent those of the authors and do not necessarily represent the position of the California Department of Public Health or the State of California. Restricted-access natality data are produced by the California Department of Public Health. To protect user privacy and confidentiality, restrictions apply to the availability of these data, and so they are not publicly released. Information about the data application process can be obtained from the California Department of Public Health in the following website: https://www.cdph.ca.gov/Programs/CHSI/Pages/Data-Applications.aspx. Vaccination data are publicly available and can be downloaded from the California Department of Public Health website.

1. S. Lopez-Leon *et al.*, More than 50 long-term effects of COVID-19: A systematic review and meta-analysis. *Sci. Rep.* **11**, 16144 (2021).
2. J. Artico *et al.*, Myocardial involvement after hospitalization for COVID-19 complicated by troponin elevation: A prospective, multicenter, observational study. *Circulation* **147**, 364–374 (2023).
3. A. M. Digby, M. H. Dahan, Obstetrical and gynecologic implications of COVID-19: What have we learned over the first two years of the pandemic. *Arch. Gynecol. Obstet.* **308**, 813–819 (2023), 10.1007/s00404-022-06847-z (20 June 2023).
4. R. A. M. Pierce-Williams *et al.*, Clinical course of severe and critical coronavirus disease 2019 in hospitalized pregnancies: A United States cohort study. *Am. J. Obstet. Gynecol. MFM* **2**, 100134 (2020).
5. J. Allotey *et al.*, Clinical manifestations, risk factors, and maternal and perinatal outcomes of coronavirus disease 2019 in pregnancy: Living systematic review and meta-analysis. *BMJ* **370**, m3320 (2020).
6. R. S. Chamseddine *et al.*, Pregnancy and neonatal outcomes in SARS-CoV-2 infection: A systematic review. *J. Pregnancy* **2020**, 1–7 (2020).
7. M. Jafari *et al.*, Clinical characteristics and outcomes of pregnant women with COVID-19 and comparison with control patients: A systematic review and meta-analysis. *Rev. Med. Virol.* **31**, 1–16 (2021).
8. D. Karasek *et al.*, The association of COVID-19 infection in pregnancy with preterm birth: A retrospective cohort study in California. *Lancet Reg. Health Am.* **2**, 100027 (2021).
9. E. McClymont *et al.*, Association of SARS-CoV-2 infection during pregnancy with maternal and perinatal outcomes. *JAMA* **327**, 1983 (2022).
10. Z. S. Lassi *et al.*, A systematic review and meta-analysis of data on pregnant women with confirmed COVID-19: Clinical presentation, and pregnancy and perinatal outcomes based on COVID-19 severity. *J. Glob. Health* **11**, 05018 (2021).
11. S. M. Newton *et al.*, Preterm birth among pregnant persons with severe acute respiratory syndrome Coronavirus 2 infection. *J. Perinatol.* **42**, 1328–1337 (2022), 10.1038/s41372-022-01467-6 (29 September 2022).
12. M. Vouga *et al.*, Maternal outcomes and risk factors for COVID-19 severity among pregnant women. *Sci. Rep.* **11**, 13898 (2021).
13. S. Q. Wei, M. Bilodeau-Bertrand, S. Liu, N. Auger, The impact of COVID-19 on pregnancy outcomes: A systematic review and meta-analysis. *CMAJ* **193**, E540–E548 (2021).
14. F. Torche, J. Nobles, The unequal impact of the COVID-19 pandemic on infant health. *Demography* **59**, 2025–2051 (2022).
15. N. J. Irons, A. E. Raftery, Estimating SARS-CoV-2 infections from deaths, confirmed cases, tests, and random surveys. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2103272118 (2021).
16. K. E. N. Clarke *et al.*, Seroprevalence of infection-induced SARS-CoV-2 antibodies – United States, September 2021–February 2022. *MMWR Morb. Mortal. Wkly. Rep.* **71**, 606–608 (2022).
17. Centers for Disease Control and Prevention, Unpacking Variants. *COVID Data Tracker Weekly Review* (2022) (29 September 2022).
18. California Department of Public Health, COVID-19 variant data (2023). https://data.chhs.ca.gov/dataset/covid-19-variant-data. Accessed 11 June 2023.
19. S. Adjei *et al.*, Mortality risk among patients hospitalized primarily for COVID-19 during the Omicron and Delta variant pandemic periods – United States, April 2020–June 2022. *MMWR Morb. Mortal. Wkly. Rep.* **71**, 1182–1189 (2022).
20. L. Wang, M. Møhlenberg, P. Wang, H. Zhou, Immune evasion of neutralizing antibodies by SARS-CoV-2 Omicron. *Cytokine Growth Factor Rev.* **70**, 13–25 (2023).
21. E. H. Adhikari, J. A. SoRelle, D. D. McIntire, C. Y. Spong, Increasing severity of COVID-19 in pregnancy with Delta (B.1.617.2) variant surge. *Am. J. Obstet. Gynecol.* **226**, 149–151 (2022).
22. E. H. Adhikari *et al.*, COVID-19 cases and disease severity in pregnancy and neonatal positivity associated with Delta (B.1.617.2) and Omicron (B.1.1.529) variant predominance. *JAMA* **327**, 1500 (2022).
23. R. P. Bhattacharyya, W. P. Hanage, Challenges in inferring intrinsic severity of the SARS-CoV-2 Omicron variant. *N. Engl. J. Med.* **386**, e14 (2022).
24. Z. H. Strasser, N. Greifer, A. Hadavand, S. N. Murphy, H. Estiri, Estimates of SARS-CoV-2 Omicron BA.2 subvariant severity in New England. *JAMA Netw. Open* **5**, e2238354 (2022).
25. T. Nyberg *et al.*, Comparative analysis of the risks of hospitalisation and death associated with SARS-CoV-2 omicron (B.1.1.529) and delta (B.1.617.2) variants in England: A cohort study. *Lancet* **399**, 1303–1312 (2022).
26. D. J. Pascall *et al.*, Directions of change in intrinsic case severity across successive SARS-CoV-2 variant waves have been inconsistent. *J. Infection* **87**, 128–135 (2023).

27. C. Hyams *et al.*, Severity of Omicron (B.1.1.529) and Delta (B.1.617.2) SARS-CoV-2 infection among hospitalised adults: A prospective cohort study in Bristol, United Kingdom. *Lancet Reg. Health Eur.* **25**, 100556 (2023).

28. N. Dagan *et al.*, Effectiveness of the BNT162b2 mRNA COVID-19 vaccine in pregnancy. *Nat. Med.* **27**, 1693–1695 (2021).

29. H. Engjom *et al.*, Severe COVID-19 in pregnancy is almost exclusively limited to unvaccinated women–time for policies to change. *Lancet Reg. Health Eur.* **13**, 100313 (2022).

30. N. B. Halasa *et al.*, Effectiveness of maternal vaccination with mRNA COVID-19 vaccine during pregnancy against COVID-19-Associated hospitalization in infants aged <6 Months –17 States, July 2021–January 2022. *MMWR Morb. Mortal. Wkly. Rep.* **71**, 264–270 (2022).

31. J. Villar *et al.*, Pregnancy outcomes and vaccine effectiveness during the period of omicron as the variant of concern, INTERCOVID-2022: A multinational, observational study. *Lancet* **401**, 447–457 (2023).

32. CDC, COVID-19 Vaccination among pregnant people (2023). https://covid.cdc.gov/covid-data-tracker/#vaccinations-pregnant-women. Accessed 5 July 2023.

33. R. Behrman, A. Stith-Butler, *Preterm Birth* (National Academies Press, 2007), 10.17226/11622.

34. J. M. Bronstein, M. S. Wingate, A. E. Brisendine, Why is the U.S. preterm birth rate so much higher than the rates in Canada, Great Britain, and Western Europe? *Int. J. Health Serv.* **48**, 622–640 (2018).

35. A. Chen, E. Oster, H. Williams, Why is infant mortality higher in the United States than in Europe? *Am. Econ. J. Econ. Policy* **8**, 89–124 (2016).

36. A. Akif, M. A. Bhuiyan, M. R. Islam, SARS-COV-2 Omicron subvariant BF.7 is again triggering the Covid fear: What we need to know and what we should do? *J. Med. Virol.* **95**, e28551 (2023).

37. D. Ao, X. He, W. Hong, X. Wei, The rapid rise of SARS-CoV-2 Omicron subvariants with immune evasion properties: XBB.1.5 and BQ.1.1 subvariants. *MedComm* **4**, e239 (2023).

38. M. Imai *et al.*, Efficacy of antiviral agents against Omicron subvariants BQ.1.1 and XBB. *N. Engl. J. Med.* **388**, 89–91 (2023).

39. A. Dorélien, The effects of in utero exposure to influenza on birth and infant outcomes in the US. *Popul. Dev. Rev.* **45**, 489–523 (2019).

40. L. Dettling, M. S. Kearney, The cyclicality of births and babies' health, revisited: Evidence from unemployment insurance. NBER Working Paper 30937 (2023). Accessed 5 November 2023.

41. S. Heft-Neal, A. Driscoll, W. Yang, G. Shaw, M. Burke, Associations between wildfire smoke exposure during pregnancy and risk of preterm birth in California. *Environ. Res.* **203**, 111872 (2022).

42. S. Eryigit Madzwamuse, N. Baumann, J. Jaekel, P. Bartmann, D. Wolke, Neuro-cognitive performance of very preterm or very low birth weight adults at 26 years. *J. Child Psychol. Psychiatr.* **56**, 857–864 (2015).

43. A. L. Murray *et al.*, Neonatal brain pathology predicts adverse attention and processing speed outcomes in very preterm and/or very low birth weight children. *Neuropsychology* **28**, 552–562 (2014).

44. N. J. Waitzman, A. Jalali, S. D. Grosse, Preterm birth lifetime costs in the United States in 2016: An update. *Semin. Perinatol.* **45**, 151390 (2021).

45. A. Sjölander, T. Frisell, S. Öberg, Sibling comparison studies. *Annu. Rev. Stat. Appl.* **9**, 71–94 (2022).

46. D. Conley, *The Pecking Order: Which Siblings Succeed and Why* (Pantheon Books, 2004).

47. N. Fallach *et al.*, Pregnancy outcomes after SARS-CoV-2 infection by trimester: A large, population-based cohort study. *PLoS One* **17**, e0270893 (2022).

48. D. A. Badr *et al.*, Severe acute respiratory syndrome Coronavirus 2 and pregnancy outcomes according to gestational age at time of infection. *Emerg. Infect. Dis.* **27**, 2535–2543 (2021).

49. M. Y. Valdespino-Vázquez *et al.*, Fetal and placental infection with SARS-CoV-2 in early pregnancy. *J. Med. Virol.* **93**, 4480–4487 (2021).

50. C. Calvert *et al.*, A population-based matched cohort study of early pregnancy outcomes following COVID-19 vaccination and SARS-CoV-2 infection. *Nat. Commun.* **13**, 6124 (2022).

51. C. L. DeSisto *et al.*, Risk for stillbirth among women with and without COVID-19 at delivery hospitalization – United States, March 2020–September 2021. *MMWR Morb. Mortal. Wkly. Rep.* **70**, 1640–1645 (2021).

52. D. A. Schwartz, S. B. Mulkey, D. J. Roberts, SARS-CoV-2 placentitis, stillbirth, and maternal COVID-19 vaccination: Clinical–pathologic correlations. *Am. J. Obstet. Gynecol.* **228**, 261–269 (2023).

53. CDC, COVID-19 vaccinations in the United States, Jurisdiction (Centers for Disease Control and Prevention, 2023).

54. A. Battarbee, Attitudes toward COVID-19 illness and COVID-19 vaccination among pregnant women: A cross-sectional multicenter study during August-December 2020. *Am. J. Perinatol.* **39**, 75–83 (2022).

55. A. L. Hsu, T. Johnson, L. Phillips, T. B. Nelson, Sources of vaccine hesitancy: Pregnancy, infertility, minority concerns, and general skepticism. *Open Forum Infect. Dis.* **9**, ofab433 (2021).

56. C. Townsel *et al.*, COVID-19 vaccine hesitancy among reproductive-aged female tier 1A healthcare workers in a United States Medical Center. *J. Perinatol.* **41**, 2549–2551 (2021), 10.1038/s41372-021-01173-9 (25 October 2021).

57. M. J. Perez *et al.*, Characterizing initial COVID-19 vaccine attitudes among pregnancy-capable healthcare workers. *Am. J. Obstet. Gynecol. MFM* **4**, 100557 (2021).

58. California Department of Public Health, Serostatus of participants based on COVID-19 antibody results and self-reported vaccination history Waves 1, 2, and 3 (2022).

59. P. L. Moro, M. M. McNeil, Successes of the CDC monitoring systems in evaluating post-authorization safety of COVID-19 vaccines. *Expert Rev. Vaccines* **21**, 281–284 (2022).

60. J. Guedalia *et al.*, Effectiveness of a third BNT162b2 mRNA COVID-19 vaccination during pregnancy: A national observational study in Israel. *Nat. Commun.* **13**, 6961 (2022).

61. S. J. Schrag *et al.*, Estimation of COVID-19 mRNA vaccine effectiveness against medically attended COVID-19 in pregnancy during periods of Delta and Omicron Variant predominance in the United States. *JAMA Netw. Open* **5**, e2233273 (2022).

62. S. M. Glynn *et al.*, Timing of COVID-19 during pregnancy and impact on placental pathology. *Am. J. Obstet. Gynecol.* **224**, S357–S358 (2021).

63. K. Narang *et al.*, SARS-CoV-2 infection and COVID-19 during pregnancy: A multidisciplinary review. *Mayo Clin. Proc.* **95**, 1750–1765 (2020).

64. C. de Chaisemartin, X. D'Haultfoeuille, Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: A survey. *Econom. J.*, 10.1093/ectj/utac017 (2022).

65. A. Goodman-Bacon, Difference-in-differences with variation in treatment timing. *J. Econom.* **225**, 254–277 (2021).