Supplementary Online Content

Whaley CM, Cantor J, Pera M, Jena AB. Assessing the association between social gatherings and COVID-19 risk using birthdays. *JAMA Intern Med.* Published online June 21, 2021. doi:10.1001/jamainternmed.2021.2915

This supplementary material has been provided by the authors to give readers additional information about their work.

eMethods.

Representativeness of the Castlight Health insurance claims data

Our primary source of data were medical claims collected by Castlight Health, which provides health navigation services for approximately 200 US employers that provide employer-sponsored insurance. To assess the representativeness of the Castlight data for the commercially insured population nationally, we examined differences between our study population and the broader commercially insured population. To make this comparison we used data from the American Community Survey (ACS) and applied sampling weights to make the ACS nationally representative. The most recent year available in the ACS was 2018, and so we limited our comparison to 2018 claims data. We limited the ACS sample to individuals who received insurance through an employer or union and were under the age of 65. As shown in **eTable 1**, the Castlight population was similar in gender, age, and geographic distribution to the ACS population.

Additional Methods

We conducted several subgroup analyses and sensitivity tests. As described in detail below, we estimated separate regression models for each subgroup analysis and tested the statistical difference in birthday effects between relevant subgroups using interaction tests. As in the baseline model, each subgroup regression interacted weekly Covid-19 cases per 10,000 in each county (decile indicators) with household birthday.

We first estimated whether the association between household birthday and Covid-19 diagnoses varied according to whether a birthday was for a child, defined as a household member

aged 18 years or younger. Households may be more likely to socially gather to celebrate children's birthdays than adult birthdays. We estimated models in which the household birthday indicator was jointly interacted with an indicator for whether the birthday was for a child (vs. adult) and weekly Covid-19 cases per 10,000 in each county (decile indicators).

Second, we assessed whether the relationship between household birthdays and Covid-19 risk varied according to whether the birthday was a milestone birthday, defined as 16th, 18th, 20th, 30th, 40th, 50th, and 60th birthdays. We hypothesized that gatherings may be more likely for milestone birthdays than for non-milestone birthdays. We estimated models in which the household birthday indicator was jointly interacted with an indicator for whether the birthday was for milestone age and weekly Covid-19 cases per 10,000 in each county (decile indicators).

Third, we examined whether the relationship between household birthdays and Covid-19 risk varied according to whether there was particularly heavy precipitation during the period of the birthday, which could drive gatherings indoors. Specifically, we focused on rainfall occurring the Saturday of the birthday week, under the assumption that birthday-related gatherings might be most likely to occur on Saturday. To conduct this analysis, we obtained precipitation data from the National Centers for Environmental Information (https://www.ncdc.noaa.gov/data-access/quick-links) and identified days with any precipitation. We estimated models in which the household birthday indicator was jointly interacted with an indicator for whether the birthday occurred in a week with any Saturday precipitation and weekly Covid-19 cases per 10,000 in each county (decile indicators). As a sensitivity test, we also categorized rainfall at the weekly level and identified weeks as having heavy precipitation if precipitation in a given week was above the weekly national mean in 2020 of 3.2 millimeters. Results were similar when using any precipitation or heavy precipitation.

Fourth, due to political differences in responses to the Covid-19 pandemic (e.g., differences in preferences for social distancing or masking across states or differences in state policies), we also analyzed whether the relationship between household birthdays and Covid-19 risk varied according to voting patterns in the 2016 presidential election. Specifically, we tested for differences in the birthday / Covid-19 relationship among counties won by Hillary Clinton versus Donald Trump in the 2016 election. We estimated models in which the household birthday indicator was jointly interacted with an indicator for whether the county was won by Trump in the 2016 election and weekly Covid-19 cases per 10,000 in each county (decile indicators). We also estimated models that used the county-level share of votes won by Donald Trump in the 2016 election, rather than a binary indicator for counties in which Donald Trump won a majority of votes.

Fifth, we examined differences based on the presence of county-level shelter-in-place policies and social distancing guidelines. We used publicly available data collected by the New York Times on what counties implemented shelter-in-place policies and the timing of implementation.[1] These policies were designed to increase social distancing and limit social mobility. Existing studies have found that these policies limited social mobility and were associated with reduced Covid-19 transmission.[2,3] If observed behavior changes following the implementation of these policies were similar for birthday-related social gatherings as other forms of social mobility, then the estimated association between birthdays and Covid-19 transmission may be smaller in counties with these policies. On the other hand, finding similar rates of transmission in counties with and without these policies would suggest that the social mobility responses to these policies demonstrated in other studies may not translate to smaller and less formal forms of social gatherings, such as birthday parties. We estimated models in

which the household birthday indicator was jointly interacted with an indicator for whether a shelter-in-place order was in effect in the household's county during the birthday week (i.e., the indicator was allowed to turn off during weeks when the shelter-in-place order was not in effect) and weekly Covid-19 cases per 10,000 in each county (decile indicators).

Finally, we tested the sensitivity of the regression specification for our main results by including household-level fixed effects and clustering standard errors at the household-level. This additional specification provided a test of the assumption implicit in our main regression model that within-household infections were independent. The household fixed effects accounted for all time-invariant differences between households.

1. Additional Results

eFigure 1 presents the weekly distribution in the share of households with a birthday (**Panel A**) and the number of birthdays (**Panel B**). On average, 4.1% of households had a birthday in a given week, with a median of 4.1%, as well. The number of households in the sample with a birthday in a given week averaged 120,317, with a median of 120,433. Birthdays were evenly distributed throughout the year, suggesting that systematic variations in birthday timing were unlikely to occur throughout the Covid-19 pandemic. The lower birthday rates in the first week of 2020 is due to January 1, 2020 falling on a Wednesday, and thus the first week of 2020 is a partial week.

Sensitivity Analyses

eFigure 2 presents results that use quintiles, rather than deciles, of COVID-19 prevalence. In the top quintile of COVID-19 prevalence, households with a birthday in the two

weeks prior experienced a 5.9 per 10,000 household increase in COVID-19 infections, which is qualitatively similar to our main results.

eFigure 3 presents results from a falsification test that replaced household birthdays with randomly assigned birthdays. In each decile of Covid-19 prevalence, the association between the randomly assigned household birthdays and Covid-19 diagnoses was not statistically different from 0. The lack of an association between randomly assigned birthdays and Covid-19 diagnoses suggests that household or market characteristics unrelated to household birthdays did not contribute to the observed association between household birthdays and Covid-19 diagnoses.

eFigure 4 presents results from a sensitivity analysis that estimated whether there was a relationship between Covid-19 diagnosis and the presence of a birthday in the household in the following 4 or 8 weeks, as opposed to the preceding 2 weeks (our primary analysis). We found no evidence that Covid-19 risk was associated with the presence of a birthday in the household in the 4 or 8 weeks following, suggesting that the increased risk of household Covid-19 infection in the two weeks following a birthday in a household was unlikely to be due to unobserved household-specific factors.

eFigure 5 presents results from a sensitivity analysis that used COVID-19-related hospitalizations as the dependent variable, an analysis conducted to assess whether observed increases in COVID-19 diagnoses following a birthday could simply be due to ascertainment bias arising from families being more likely to seek testing for COVID-19 if they recently had a birthday party. We hypothesized that COVID-19 hospitalizations, which by definition reflect severe infections, would not be susceptible to ascertainment bias. COVID-19-related hospitalizations were identified as hospital visits with a diagnosis code for COVID-19. To account for lags between COVID-19 infection and hospitalization, we used a four-week lag

period between household birthdays and hospitalizations. We found similar patterns for COVID-19 hospitalizations as for diagnoses. Among county-week combinations in the bottom five deciles of COVID-19 prevalence, we did not find that household birthdays were associated with COVID-19-related hospitalizations. However, in the sixth, eight, and ninth deciles of COVID-19 prevalence, we found positive and statistically significant associations between household birthdays and COVID-19-related hospitalizations.

eTable 2 presents results that use the continuous county-level share of Trump 2016 voter share, rather than an indicator for the majority of voter share going to Trump. When using the continuous Trump 2016 voter share, the COVID-19 infection rate was statistically different at the $p<0.05$ level in the sixth through tenth deciles of COVID-19 prevalence. However, the extent of effect modification was small. For example, the results suggest that a 10-percentage point increase in Trump voter share (e.g., equivalent to a 45 percent to 55 percent voter share spread, which is almost 40% larger than the 2008 Obama voter share margin) in counties in the tenth decile of COVID-19 prevalence only increased the association between household birthdays and COVID-19 infection by 1.4 infections per 10,000 households.

eFigure 6 presents results from a regression model that included household-level fixed effects, rather than county-level fixed effects. We also clustered standard errors at the household level, instead of the county-level used in the main results. Overall, these results were similar to our main regression specification. Among counties in the top decile of COVID-19 prevalence, household birthdays in the two weeks prior were associated with an increase of 7.6 (95% CI: 5.8 to 9.3) per 10,000 household rate of COVID-19 infections.

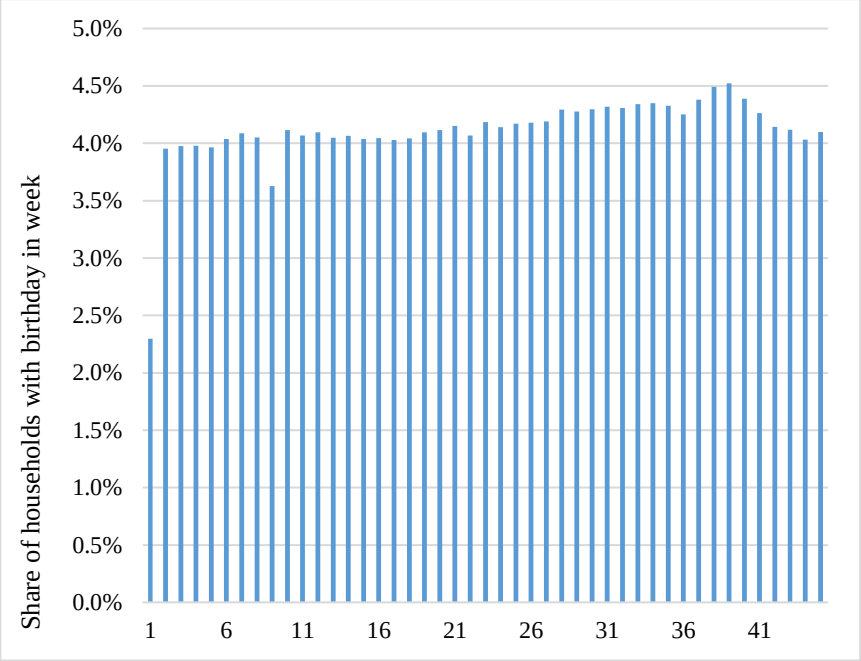
eTable 1. Comparison Between Castlight Study Population and Population With Employer-Sponsored Insurance in American Community Survey (ACS)

Characteristics	Castlight (2018)	ACS (2018)*
Number of enrolled persons	5,608,888	162,136,077
Gender, percent female (%)	50.0%	50.0%
Average age	34.3	33.5
Census region		
South	40.6%	35.70%
Midwest	23.0%	22.80%
Northeast	10.4%	18.20%
West	26.0%	23.40%

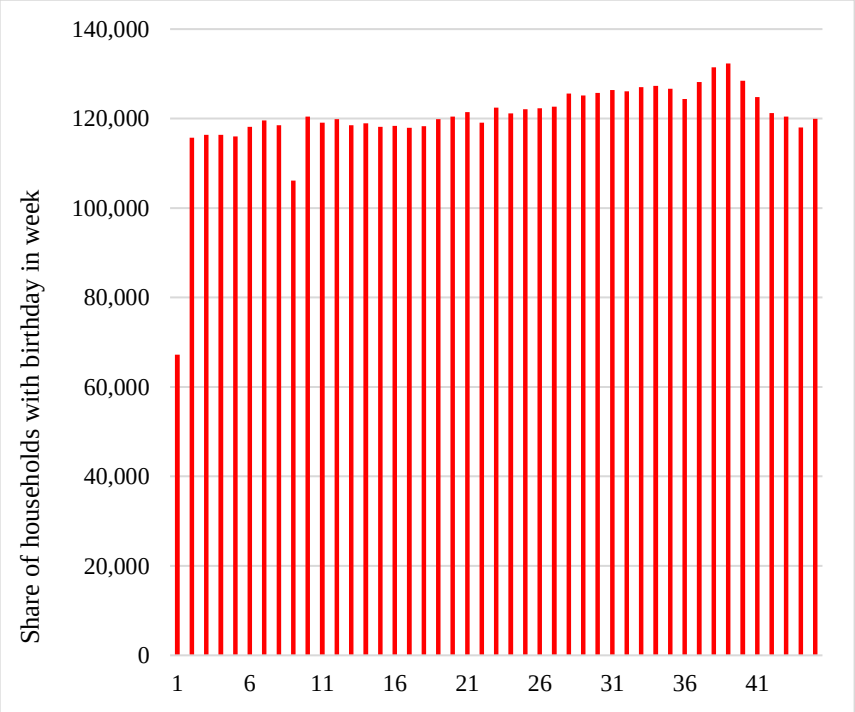
Notes: *Nationally representative survey weights applied.

eFigure 1. Weekly Distribution of Household Birthdays

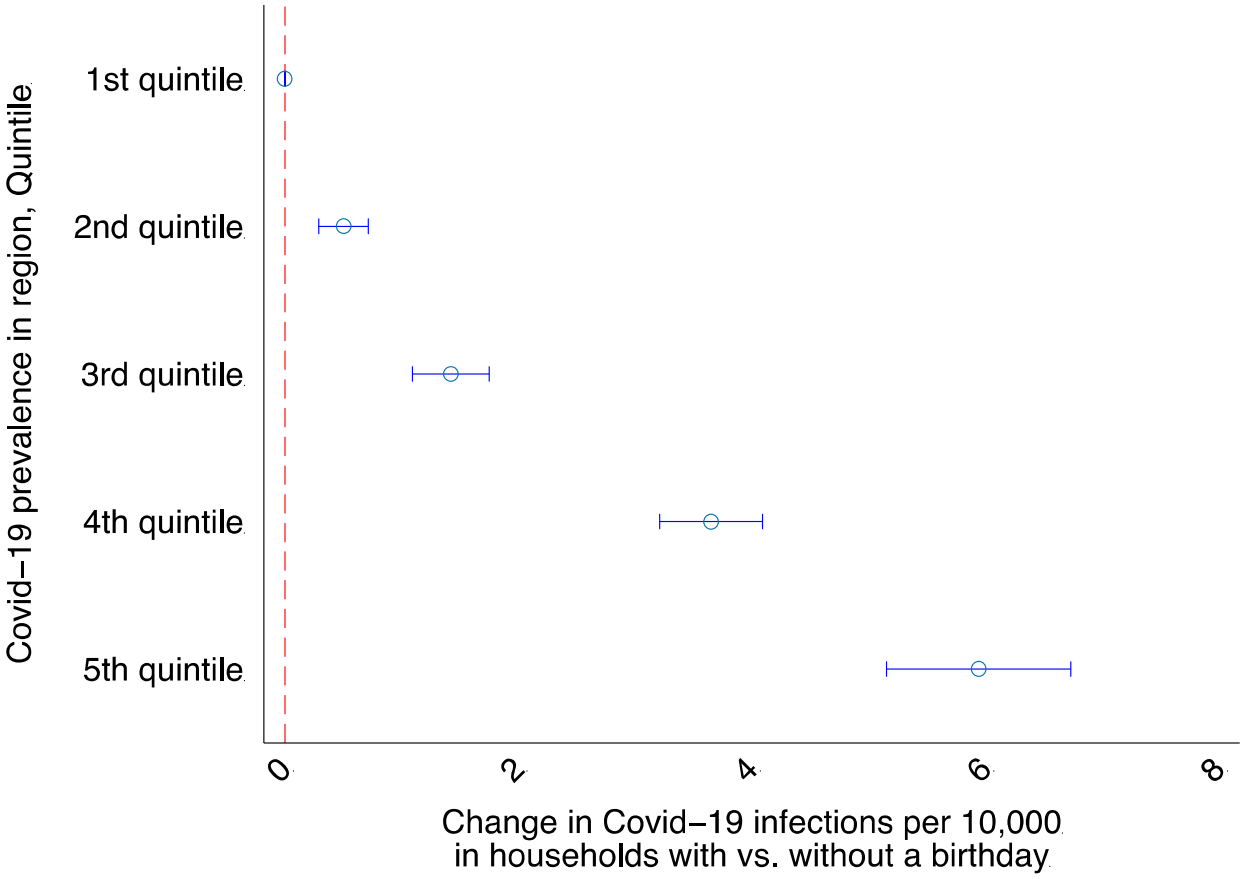
(a) Share of Households with Birthday in Each Week



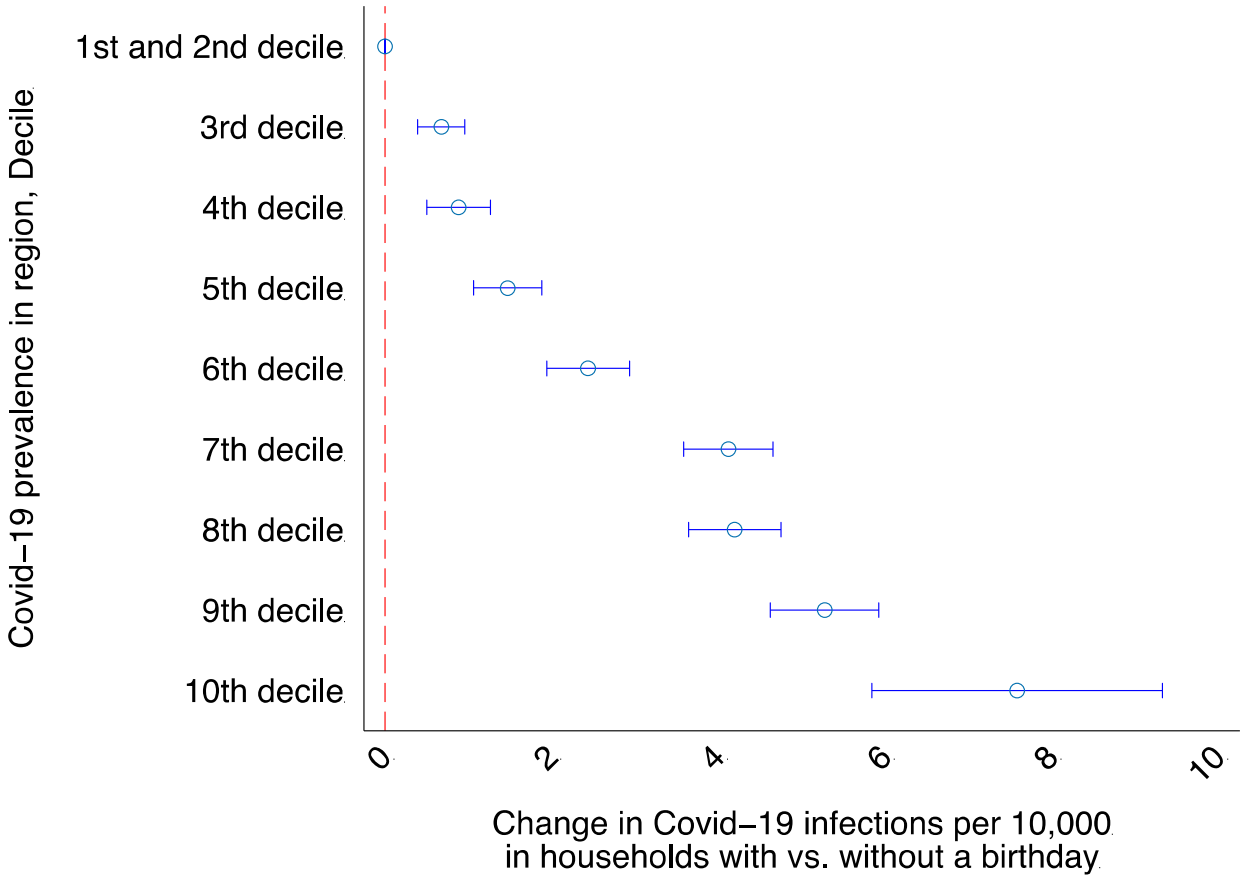
(b) Number of Households with Birthday in Each Week



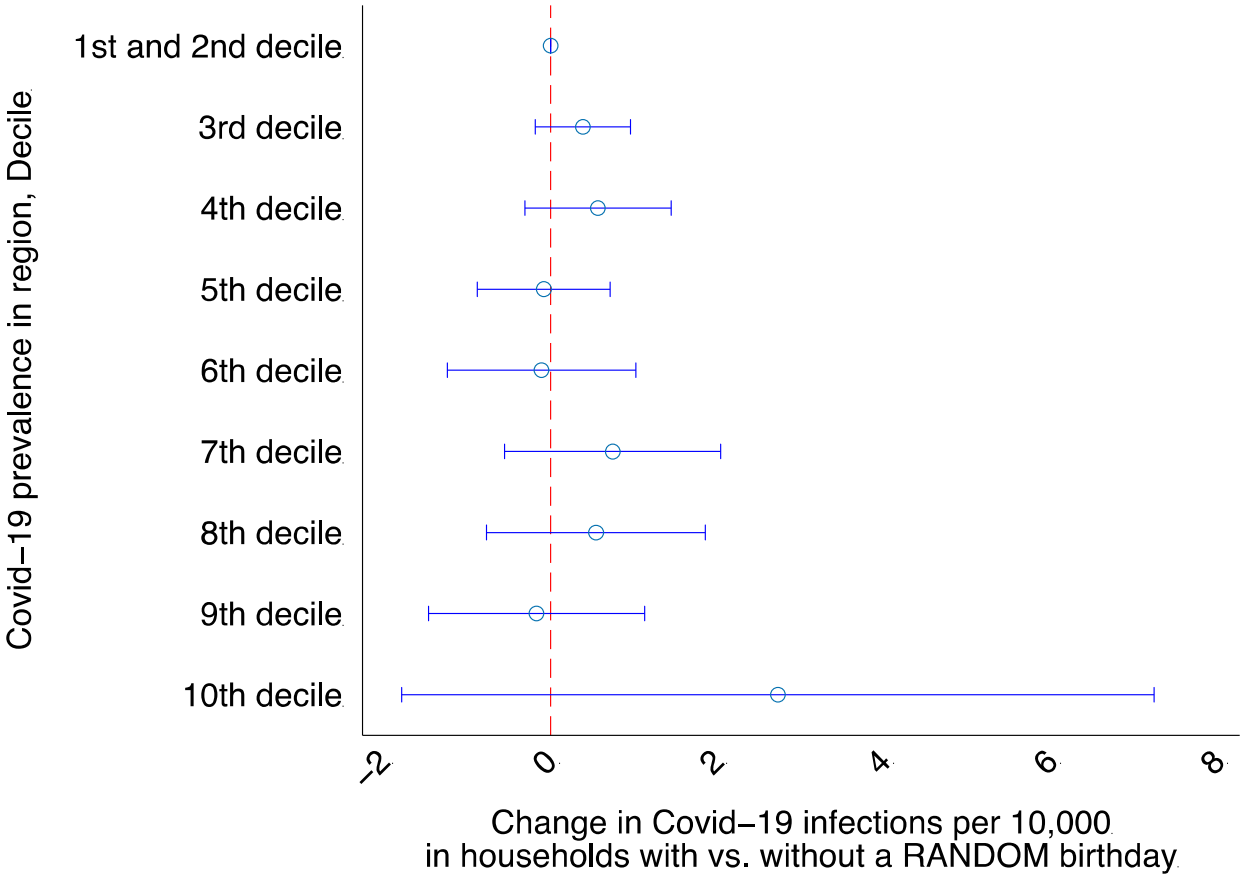
eFigure 2. Differences in Rates of COVID-19 According to Presence of a Birthday in an Individual's Household, Sensitivity Test That Uses COVID-19 Prevalence Quintiles



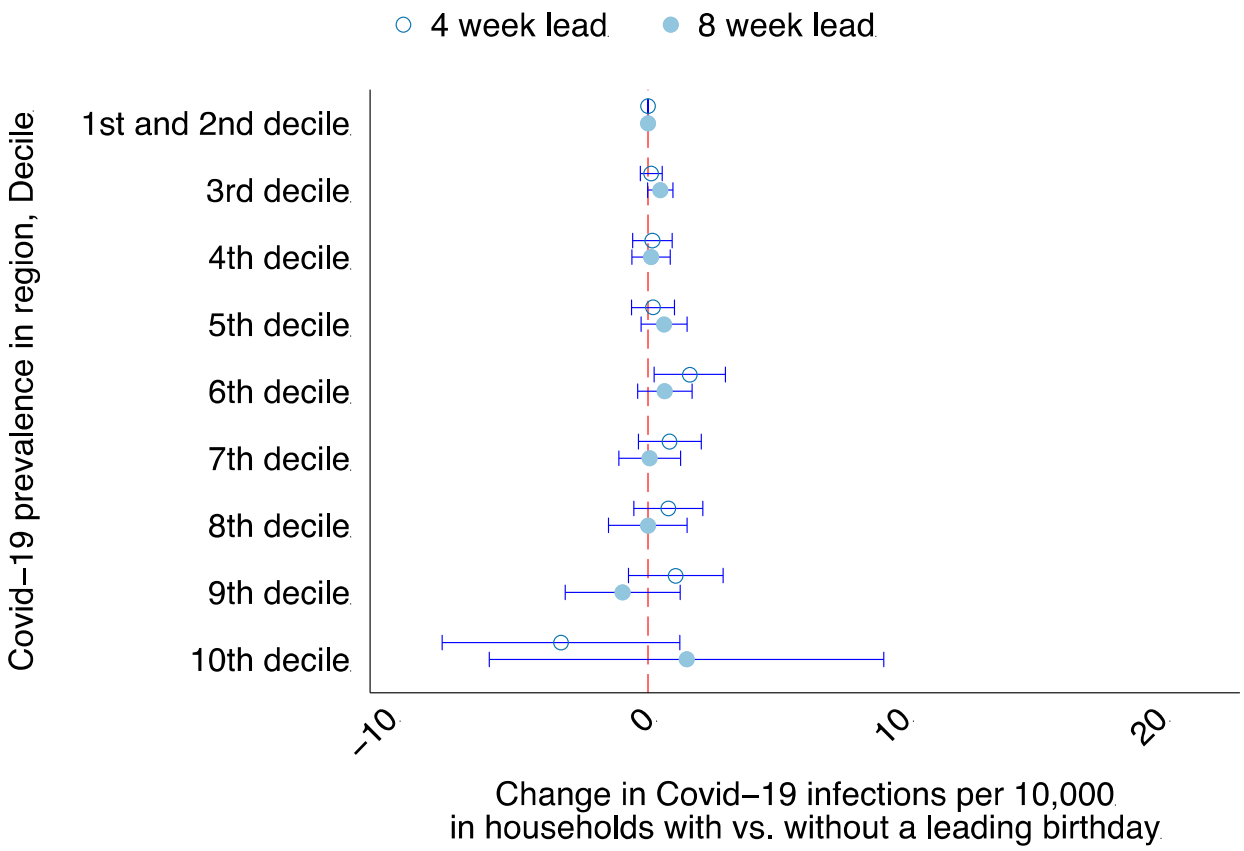
eFigure 3. Differences in Rates of COVID-19 According to Presence of a Birthday in an Individual's Household, Sensitivity Test That Includes Household Fixed Effects and Clusters Standard Errors at Household Level



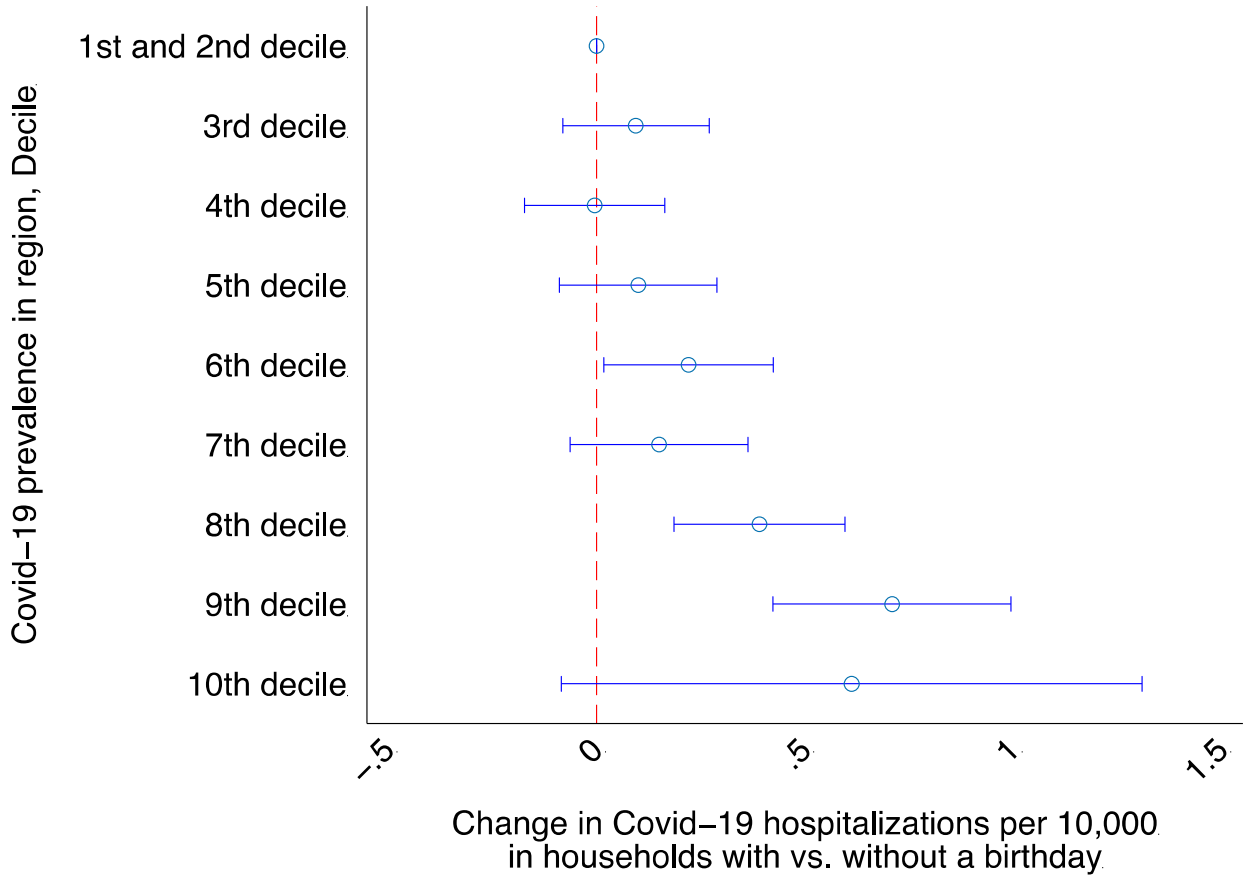
eFigure 4. Falsification Analysis Based on Randomly Assigned Birthdays



eFigure 5. Differences in Rates of COVID-19 According to Presence of a Birthday in an Individual's Household, Falsification Analysis Using Diagnoses in the 4 or 8 Weeks Before Birthday



eFigure 6. Differences in Rates of COVID-19–Related Hospitalization According to Presence of a Birthday in an Individual's Household



eTable 2. Differences in Rates of COVID-19 Diagnosis, According to Presence of a Household Birthday, County-Level COVID-19 Prevalence, and County-Level Share of Votes Won by Donald Trump in the 2016 US Presidential Election

	(1)
Household birthday X Trump 2016 vote share X 3rd decile	2.01e-05
	(7.38e-05)
Household birthday X Trump 2016 vote share+A18 X 4th decile	3.15e-08
	(9.88e-05)
Household birthday X Trump 2016 vote share X 5th decile	-4.08e-05
	(0.000114)
Household birthday X Trump 2016 vote share X 6th decile	0.000408***
	(0.000152)
Household birthday X Trump 2016 vote share X 7th decile	0.000615***
	(0.000166)
Household birthday X Trump 2016 vote share X 8th decile	0.000385**
	(0.000173)
Household birthday X Trump 2016 vote share X 9th decile	0.000837***
	(0.000223)
Household birthday X Trump 2016 vote share X 10th decile	0.00140**
	(0.000596)
Observations	120,267,042
Robust standard errors in parentheses	
*** p<0.01, ** p<0.05, * p<0.1	

eReferences.

1. Cook J, Newberger N, Smalling S. The spread of social distancing. Economics Letters 2020;196:109511.
2. Courtemanche C, Garuccio J, Le A, Pinkston J, Yelowitz A. Strong Social Distancing Measures In The United States Reduced The COVID-19 Growth Rate. Health Aff (Millwood) 2020;39:1237-46.
3. Gao S, Rao J, Kang Y, et al. Association of Mobile Phone Location Data Indications of Travel and Stay-at-Home Mandates With COVID-19 Infection Rates in the US. JAMA Network Open 2020;3:e2020485-e.