The relative incidence of COVID-19 in healthcare workers versus non-healthcare workers: evidence from a web-based survey of Facebook users in the United States [version 1; peer review: 1 approved with reservations]

Abraham D. Flaxman1, Daniel J. Henning2, Herbert C. Duber1,2

1Institute for Health Metrics and Evaluation, University of Washington, Seattle, WA, 98195, USA
2Department of Emergency Medicine, University of Washington, Seattle, WA, 98195, USA

Abstract
Background: Healthcare workers are at the forefront of the COVID-19 pandemic and it is essential to monitor the relative infection rate of this group, as compared to workers in other occupations. This study aimed to produce estimates of the relative incidence ratio between healthcare workers and workers in non-healthcare occupations.

Methods: Analysis of cross-sectional data from a daily, web-based survey of 1,788,795 Facebook users from September 6, 2020 to October 18, 2020. Participants were Facebook users in the United States aged 18 and above who were tested for COVID-19 because of an employer or school requirement in the past 14 days. The exposure variable was a self-reported history of working in healthcare in the past four weeks and the main outcome was a self-reported positive test for COVID-19.

Results: On October 18, 2020, in the United States, there was a relative COVID-19 incidence ratio of 0.7 (95% UI 0.6 to 0.8) between healthcare workers and workers in non-healthcare occupations.

Conclusions: Currently in the United States, healthcare workers have a substantially and significantly lower COVID-19 incidence rate than workers in non-healthcare occupations.

Keywords
COVID-19, healthcare workers
Corresponding author: Abraham D. Flaxman (abie@uw.edu)

Author roles: Flaxman AD: Conceptualization, Methodology, Writing – Original Draft Preparation, Writing – Review & Editing; Henning DJ: Methodology, Writing – Review & Editing; Duber HC: Methodology, Writing – Review & Editing

Competing interests: ADF has consulted recently for Janssen; SwissRe; Sanofi; Merck for Mothers; and Agathos, Ltd. DJH has received research funding from Baxter and performed consulting services for Cytovale. HCD has no competing interests to disclose.

Grant information: This work was supported by the Bill and Melinda Gates Foundation [OPP1170133] and the National Science Foundation [DMS-1839116].
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2020 Flaxman AD et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Flaxman AD, Henning DJ and Duber HC. The relative incidence of COVID-19 in healthcare workers versus non-healthcare workers: evidence from a web-based survey of Facebook users in the United States [version 1; peer review: 1 approved with reservations] Gates Open Research 2020, 4:174 https://doi.org/10.12688/gatesopenres.13202.1

First published: 27 Nov 2020, 4:174 https://doi.org/10.12688/gatesopenres.13202.1
Introduction
In August, the Peterson-KFF Health System Tracker published a collection of charts showing how healthcare utilization has declined during the COVID-19 pandemic in the United States1, showing that facility discharge volume dropped by over 25% and cancer screening volumes dropped by over 85% from levels in 2019. This decrease is consistent with evidence from other sources2-5, and could be driven by a perceived risk of interacting with workers at health facilities. It is yet to be seen how much this delayed and foregone care will reduce population health. Meanwhile, a Wall Street Journal analysis of Centers for Disease Control and Prevention (CDC) data found that at least 7,400 COVID-19 infections were transmitted in US hospitals in 20206. Access to adequate resources for infection prevention among health care workers (HCWs) remains a topic of urgent importance.7

There is currently no population-based evidence quantifying the relative COVID-19 incidence rate among HCWs as compared to workers in non-healthcare occupations (non-HCWs) in the US. We hypothesized that there is not a substantially elevated rate of COVID-19 infection among HCWs and that HCWs might even have lower incidence rate than non-HCWs, and we analyzed data from a large survey of Facebook users to investigate.

Methods
Study design
We analyzed individual participant data from a large, web-based survey of Facebook users aged 18 and above in the United States (around 300,000 respondents per week). Every day Facebook offered a random sample of US-based users a Qualtrics survey run by the Delphi lab at Carnegie Mellon University who made it rapidly available to other academic researchers8. Facebook also provided survey weights to adjust for the demographics of the active Facebook user population9. This sort of survey data has been used previously to perform population based analyses related to COVID-19, though never before at such large scale10. Our analysis relied on the responses to two lines of questions: (1) questions about recent work history, worded as, “In the past 4 weeks, did you do any kind of work for pay?” and if so, “[p]lease select the occupational group that best fits the main kind of work you were doing in the last four weeks”; and (2) questions about COVID-19 testing history, worded as, “Have you ever been tested for coronavirus (COVID-19)?”, “[h]ave you been tested for coronavirus (COVID-19) in the last 14 days?”, “[d]id this test find that you had coronavirus (COVID-19)?”, and “[d]o any of the following reasons describe why you were tested for coronavirus (COVID-19) in the last 14 days? Please select all that apply.”

We analyzed the most recently available six weeks of data from September 6, 2020 to October 18, 2020, which provided more than 80% power to detect a 30% difference between COVID-19 prevalence in HCWs and non-HCWs (details below).

Variables
To quantify the relative risk of COVID-19 among healthcare workers (HCWs) versus workers in non-healthcare occupations (non-HCWs), we used the response to the occupational group question as our exposure variable (we coded respondents who selected option “Healthcare practitioners and technicians” or “Healthcare support” as HCWs, and all others, including those with a missing value, as non-HCWs). We identified individuals with COVID-19 as those who reported that they had tested positive for COVID-19 in the last 14 days.

Statistical methods
We calculated the endorsement rate of positive COVID-19 test (ER) for the HCW and non-HCW population as the survey-weighted percent of respondents in either group who reported COVID-19, and calculated the relative COVID-19 incidence ratio (RR) between the HCWs and non-HCWs who were tested in the last 14 days because their employer or school required it.

As sensitivity analyses, we considered also alternative inclusion criteria and more restrictive subsets of HCWs. The survey provided sample weights that adjust for non-response bias, which we used in our main analysis. As a sensitivity analysis, we repeated our calculation using the unweighted data. To investigate the possibility that workplace testing practices differ between HCW and non-HCW occupational settings, we also repeated our analysis with additional filtering based on the “why you were tested” question. In the main result we used the subset of individuals who responded that they were tested in the last 14 days because of employer/educational requirements, and this question has a “select all that apply” answer type, and also includes “I felt sick” as an option. As a sensitivity analysis, we used only those individuals who were tested because of a workplace requirement and did not feel sick.

Power calculation: To determine the sample size necessary to detect a difference of 30% between the COVID-19 prevalence of HCWs and non-HCWs, we developed a small simulation model where the fraction of HCWs in the general population and the COVID-19 prevalence in the general population both match that observed in the survey data.

Of respondents who were tested in the last 14 days because their employer or school required it, 33.9% were HCWs and 4.9% tested positive for COVID-19, so we simulated populations of size n with these fractions of HCWs and this positive rate among the non-HCW population. We made the positive rate among the HCW population 30% lower:
Then for populations of ranging in size from $n = 500$ to 9,500, we repeatedly synthesized a simulated population, calculated the RR of COVID-19 between the HCWs and non-HCWs as described in the main text, and checked if the upper bound of the uncertainty interval was less than 1.0. We replicated this experiment 10,000 times for each population size $n$ and found the $n$ where at least 80% of the experimental replications where the uncertainty interval upper bound was less than one.

**Ethical statement**
These research activities used no identifiable private information and were therefore exempt from institutional board review.

**Results**
The survey data contained 40,552 respondents who were tested due to workplace requirements in the time period we focused on, 13,747 HCWs and 26,805 non-HCWs (see Table 1 for demographic details). There were 1,993 respondents who reported a positive test for COVID-19 in the last 14 days (527 among HCWs and 1,466 among non-HCWs).

Among HCWs with a required test, 527 of 13,747 (3.8%) reported a positive test in the last 14 days, while among non-HCWs with a required test, 1,466 of 26,805 (5.5%) reported a positive test, for a relative COVID-19 prevalence ratio of 0.7 (95% UI 0.6 to 0.8) (Table 2).

---

**Table 1. Characteristics of survey respondents.**

<table>
<thead>
<tr>
<th></th>
<th>Non-healthcare workers</th>
<th>Healthcare workers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>1,672,980 (100.0)</td>
<td>115,814 (100.0)</td>
</tr>
<tr>
<td>Tested in last 14 days</td>
<td>123,830 (7.4)</td>
<td>21,071 (18.2)</td>
</tr>
<tr>
<td>Test required by work or school</td>
<td>26,805 (1.6)</td>
<td>13,747 (11.9)</td>
</tr>
<tr>
<td>Among those with required test</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male gender</td>
<td>8,662 (32.3)</td>
<td>1,972 (14.3)</td>
</tr>
<tr>
<td>Age in years</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18 to 24</td>
<td>3,356 (12.5)</td>
<td>761 (5.5)</td>
</tr>
<tr>
<td>25 to 34</td>
<td>4,648 (17.3)</td>
<td>2,374 (17.3)</td>
</tr>
<tr>
<td>35 to 44</td>
<td>4,784 (17.8)</td>
<td>3,058 (22.2)</td>
</tr>
<tr>
<td>45 to 54</td>
<td>4,797 (17.9)</td>
<td>3,377 (24.6)</td>
</tr>
<tr>
<td>55 to 64</td>
<td>3,983 (14.9)</td>
<td>3,141 (22.8)</td>
</tr>
<tr>
<td>65 to 74</td>
<td>1,204 (4.5)</td>
<td>920 (6.7)</td>
</tr>
<tr>
<td>75 and older</td>
<td>476 (1.8)</td>
<td>105 (0.8)</td>
</tr>
</tbody>
</table>

**Table 2. Relative COVID-19 incidence rate (RR) and counts of healthcare workers and non-healthcare workers and their crude prevalence counts and rates.**

<table>
<thead>
<tr>
<th>Healthcare workers</th>
<th>Non-healthcare workers</th>
<th></th>
<th></th>
<th>RR</th>
<th>95% UI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tested</td>
<td>Positive</td>
<td>%</td>
<td>Tested</td>
<td>Positive</td>
<td>%</td>
</tr>
<tr>
<td>13,747</td>
<td>527</td>
<td>3.8</td>
<td>26,805</td>
<td>1,466</td>
<td>5.5</td>
</tr>
</tbody>
</table>
Our power calculation simulation results showed that 7,000 simulants provide 80% power to reject a null hypothesis that HCWs and non-HCWs have the same RR if, in truth, the RR is 0.7. Since the survey currently collects a weekly volume of around 7,000 individuals who report taking a required COVID-19 test, the simulation results imply that six weeks of data will provide more than sufficient power.

**Sensitivity analyses**

When we repeated our calculation using the unweighted survey responses to calculate the COVID-19 incidence ratio, we found an even smaller relative incidence ratio of 0.4 (95% UI 0.3 to 0.5).

When we repeated our analysis restricted to only specific types of HCWs, as afforded by the questionnaire, we found a range of risks, usually less than 1.0, with substantially less certainty due to small sample sizes (Table 3).

When we used only those individuals who were tested because of a workplace requirement and did not feel sick, we obtained a relative risk closer to 1.0. Using only those tested because of a workplace requirement who also did feel sick we still obtained a relative risk substantially smaller than 1.0 (Table 4). Although this finding could suggest that differences in testing patterns between healthcare and other work settings are partially responsible for the different positivity rates among HCWs and non-HCWs, it could also be driven by greater access to COVID-19 testing for confirmation of illness among HCWs experiencing symptoms. The recall period of 14 days provides ample time for an individual to receive a workplace test without symptoms, then develop symptoms.

| Table 3. Relative COVID-19 incidence rate (RR) and counts of healthcare workers (HCWs) and non-healthcare workers stratified by worker subtype. |
|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
| Number of non-HCWs | Number of HCWs    | Relative risk     | Lower bound       | Upper bound       |
| All HCWs           | 26,805            | 13,747            | 0.7               | 0.6               | 0.8               |
| Physician or surgeon | 40,277           | 275               | 2.6               | 1.8               | 3.5               |
| Registered nurse (including nurse practitioner) | 37,573 | 2,979 | 0.6 | 0.6 | 0.8 |
| Licensed practical or licensed vocational nurse | 38,560 | 1,992 | 0.6 | 0.5 | 0.8 |
| Physician assistant | 40,405           | 147               | 0.7               | 0.4               | 1.3               |
| Dentist            | 40,518            | 34                | 0.4               | 0.0               | 0.8               |
| Any other treating practitioner | 40,189 | 363 | 0.5 | 0.3 | 0.9 |
| Pharmacist          | 40,473            | 79                | 0.3               | 0.1               | 0.8               |
| Any therapist       | 39,371            | 1,181             | 0.5               | 0.4               | 0.7               |
| Any health technologist or technician | 39,062 | 1,490 | 1.0 | 0.7 | 1.2 |
| Veterinarian        | 40,519            | 33                | 0.3               | 0.0               | 1.1               |
| Nursing assistant or psychiatric aide | 39,045 | 1,507 | 1.0 | 0.8 | 1.3 |
| Home health or personal care aide | 39,999 | 553 | 0.8 | 0.5 | 1.0 |
| Occupational or physical therapy assistant or aide | 40,477 | 75 | 1.3 | 0.5 | 1.9 |
| Massage therapist   | 40,549            | 3                 | 4.6               | 0.0               | 8.1               |
| Dental assistant    | 40,534            | 18                | 0.0               | 0.0               | 0.0               |
| Medical assistant   | 40,415            | 137               | 1.1               | 0.5               | 1.7               |
| Medical transcriptionist | 40,526     | 26                | 0.6               | 0.0               | 1.5               |
| Pharmacy aide       | 40,536            | 16                | 0.0               | 0.0               | 0.0               |
| Phlebotomist        | 40,524            | 28                | 3.4               | 0.7               | 4.8               |
| Veterinary assistant | 40,547           | 5                 | 3.4               | 0.0               | 12.0              |
| Any other healthcare support worker | 38,379 | 2,173 | 0.5 | 0.4 | 0.6 |
and then receive another test to determine if the symptoms are due to COVID-19, and HCWs might have more opportunity to access such a follow-up test, since they are visiting a healthcare setting for work already.

**Discussion**
This study utilized a population-based approach to examine the relative risk of COVID-19 infection among HCW compared with non-HCW. Finding a relative COVID-19 incidence ratio substantially and significantly less than 1.0 is an unequivocally positive finding, indicating that infection control measures being taken by HCWs in total are effective. Our findings are consistent with the limited other evidence available on the risk of COVID-19 in healthcare facility settings\(^1\)–\(^4\), and, taken together, this growing body of evidence suggests that providing and seeking healthcare at this point in the epidemic is quite safe. HCWs need not fear contracting or transmitting infections more than other workers do, and patients should not defer needed care at present over concern that they will be exposed to COVID-19 during their interactions with HCWs.

This outbreak and our understanding of it have both changed rapidly in the past, and may do so again, so we will continue to update this information.

**Limitations**
This work has at least three limitations. First, our results are based on self-reported data and therefore subject to both recall bias and social desirability bias, although the questions we relied on did not seem particularly at risk for either of these biases; the question “have you been tested for COVID-19 in the last 14 days?” likely included positive responses from individuals who received seroprevalence testing as well as PCR testing as well, which could also introduce a small amount of bias. Second, our approach required a large sample size to obtain a sufficiently precise estimate of RR, but this seems safer than including respondents who did not report receiving a required test, as that could introduce confounding. Third, it is possible that there was still uncontrolled confounding due to differential access to tests between HCWs and non-HCWs. Our sensitivity analysis found substantively similar results when restricted only to individuals who had workplace testing when they did not feel sick, but since we have only considered respondents with tests required by their employer or school, this might focus on non-HCW setting with better-than-average infection control policies (for example, they are doing asymptomatic testing) and therefore the relative risk for HCWs might be even lower than our method estimated.

**Conclusion**
As of October, 2020, in the United States the relative infection ratio of HCWs to non-HCWs is reassuringly low. Infection control remains essential and HCWs must continue to be protected as the COVID-19 pandemic continues, to ensure safety to themselves, their co-workers, and their patients.

**Data availability**
Underlying data
The underlying data used in this study are available to academic researchers for research purposes from Facebook at: https://www.facebook.com/research-operations/rfp/?title=covid19-symptom-survey-data-access. Conditions of access and instructions for applications can be found at https://dataforgood.fb.com/docs/covid-19-symptom-survey-request-for-data-access/.

**Code availability**
Reproducibility code available from: https://github.com/aflaxman/covid_hcw_r

Archived code at time of publication: http://doi.org/10.5281/zenodo.4270368\(^\text{16}\).

License: GNU General Public License v3.0

---

**Table 4. Relative COVID-19 incidence rate (RR) and counts of healthcare workers and non-healthcare workers stratified by those who reported they felt/did not feel sick as an additional reason for getting tested.**

<table>
<thead>
<tr>
<th></th>
<th>Number of non-HCWs</th>
<th>Number of HCWs</th>
<th>Relative risk</th>
<th>Lower bound</th>
<th>Upper bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test required, did not feel sick</td>
<td>23,523</td>
<td>12,789</td>
<td>1.1</td>
<td>1.0</td>
<td>1.2</td>
</tr>
<tr>
<td>Test required, felt sick</td>
<td>3,282</td>
<td>958</td>
<td>0.8</td>
<td>0.7</td>
<td>0.9</td>
</tr>
</tbody>
</table>

---

**References**


Open Peer Review

Current Peer Review Status: ?

Version 1

Reviewer Report 04 December 2020

https://doi.org/10.21956/gatesopenres.14411.r30079

© 2020 Reinhart A. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Alex Reinhart
Department of Statistics & Data Science, Carnegie Mellon University, Pittsburgh, PA, USA

This presents a timely and useful analysis of large-scale survey data. For an analysis like this, it's very important to clearly present the meaning of the data and the caveats in the survey design; the authors do a good job here, and my comments here focus on making the paper even clearer.

The analysis seems reasonable overall, and, subject to the limitations of the survey design, a useful contribution to the area.

I've separated my comments into "Main comments", which I think should be addressed to make the article more sound, and "Minor comments" that just make minor improvements to the paper.

Main comments:

- The "Sensitivity analyses" section (page 5) explains that "When we repeated our calculation using the unweighted survey responses to calculate the COVID-19 incidence ratio, we found an even smaller relative incidence ratio of 0.4 (95% UI 0.3 to 0.5)." This seems surprising. Do you have any hypotheses that could explain why this is? It suggests that either the age and gender distributions for HCWs and non-HCWs are quite different (since the survey weights correct for age and gender) or that the estimated non-response for the groups are quite different.

- The last paragraph of the Discussion suggests the possibility that "since we have only considered respondents with tests required by their employer or school, this might focus on non-HCW setting with better-than-average infection control policies". This may be a good subject for an additional table of results: A comparison of the distributions of occupation among non-HCW people who were required to be tested and those who were not. Such a table would tell the reader whether those who are required to be tested are from an unusual group of occupations, to help tell whether those occupations might be higher or lower risk than average.

- Table 3 contains a "Number of non-HCWs" column, but I don't know how to interpret this.
What does it mean to say that there were 26,805 non-HCWs in the "All HCWs" row?

- In the Limitations (page 6), the authors mention recall bias and social desirability bias as possible problems. But another key bias would be response bias: while Facebook's weights try to adjust for non-response, if they do not completely adjust for every possible factor related to non-response, there can still be bias. For example, if people who are much more concerned about COVID and take more precautions are also more likely to participate in the survey, and if Facebook does not have covariates that can predict this accurately, the survey sample can be biased relative to the population. It would be good to address this and indicate how it could affect the results.

Minor comments:

- The "Study design" subsection mentions that "Facebook also provided survey weights to adjust for the demographics of the active Facebook user population." It would be good to be explicit about what corrections are included in the weights:
  - The weights adjust for non-response, using Facebook's estimate of the probability of each sampled individual participating in the survey.
  - The weights are then post-stratified by age and gender only.

- In the "Study design" subsection, the second paragraph states "We analyzed the most recently available six weeks of data from September 6, 2020 to October 18, 2020", but Wave 4 of the survey (containing the occupation and testing questions) was only deployed on September 8, 2020. If data from September 6 and 7 was included, I assume it was left out of the study, because the respondents would not have answered the relevant questions.

- It may help readers to be explicit about the survey text and its location. The survey documentation site contains the full text of each survey wave, and referring to this could help readers who want to read the survey text and flow.

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Yes

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Yes
**Competing Interests:** I am a member of the Delphi group at Carnegie Mellon University. Delphi, in collaboration with Facebook and researchers at the University of Maryland, conducts the survey whose data is analyzed in this article, and I manage much of the process on behalf of Delphi (with assistance from Delphi team members). Delphi makes this data available to many researchers, including the authors of this article. I was not involved in the analysis conducted by the authors of this article, and have not corresponded with them about this research, so my review of the scientific merit of the work has been conducted independently. I confirm that this has not affected my ability to write an objective and unbiased review of this article.

**Reviewer Expertise:** I am a professional statistician and assistant teaching professor of Statistics & Data Science at Carnegie Mellon University. I am also a member of the Delphi group, and manage the collection of the survey data described in this article; see my Competing Interests for further details.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.