medRxiv preprint doi: https://doi.org/10.1101/2021.09.14.21263545; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Widening Disparities in Online Information Access during the COVID-19 Pandemic

Jina Suh[1,2], Eric Horvitz[1], Ryen W. White[1], and Tim Althoff[2,*]

[1]Microsoft Research, Redmond, WA, USA
[2]University of Washington, Seattle, WA, USA
[*]althoff@cs.washington.edu

ABSTRACT

The COVID-19 pandemic has stimulated a staggering increase in online information access, as digital engagement became necessary to meet the demand for health, economic, and educational resources[1,2]. We pursue insights about inequity in leveraging online information, spanning challenges with access and abilities to effectively seek and use digital information[2–4]. We observe a widening of digital inequalities through a population-scale study of 55 billion everyday web search interactions during the COVID-19 pandemic across 25,150 US ZIP codes. We observe that ZIP codes with low socioeconomic status (SES) and high racial/ethnic diversity did not leverage health information and pandemic-relevant online resources (e.g., online learning, online food delivery) as much as regions with higher SES and lower levels of diversity. We also show increased shifts in online information access to financial or unemployment assistance for ZIP codes with low SES and high racial/ethnic diversity. These findings demonstrate that the pandemic has exacerbated existing inequalities in online information access and highlight the role of large-scale, anonymized data about online search activities in digital disparities research. The results frame important questions and future research on identifying and targeting interventions for vulnerable subpopulations that could reduce further widening of digital access inequalities and associated downstream outcomes including health, education, and employment[3].

Main

Socioeconomic and environmental factors play a significant role in the health and well-being of individuals and communities[5–7]. Despite long-term and acute, pandemic-centric efforts to close the health equity gap[6], studies during the COVID-19 pandemic have demonstrated that socioeconomically and environmentally disadvantaged subpopulations have been disproportionately and negatively affected by the disease[8–10], with three-fold higher infection rates and six-fold higher death rates in predominantly black US counties than in white counties[11]. In recent decades, digital access has also gained attention as an important factor modulating health outcomes, as individuals harness the internet to seek health information and to access healthcare services (i.e., telehealth, online pharmacy)[4]. During the COVID-19 pandemic, digital engagement in resources across health, educational, economic, and social needs grew in importance because of lockdown mandates, social isolation, and economic burdens[1,2,12] as well as due to internet-based communication methods employed by public institutions, such as online dissemination of COVID-related information by the World Health Organization[2].

Unfortunately, disparities in digital access also exist across the socioeconomic and environmental factors[13]. Digital inequalities manifest not only as the differences in fundamental or infrastructure access to computers and internet but also as different abilities to harness or invest time in accessing online information and different understanding about the value of online information (i.e., second- and third-level

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2021.09.14.21263545; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

digital divide[14–16]). For example, individuals with low socioeconomic status (SES) are shown to be slower to adopt information and communication technologies (ICTs) from childhood and throughout their life course and have smaller social networks and limited employment opportunities as a result[3]. Furthermore, even after controlling for internet access, those from higher SES integrate digital resources into their lives and use the internet for more "capital-enhancing" activities that are likely to result in more upwards mobility[15,17]. Because this lack of engagement in digital resources may impact a range of downstream outcomes such as health[3,4,18], education[19], and employment[20,21], it is important to observe such behaviors across subpopulations and scrutinize the role of digital inequalities in our society. In addition, disadvantaged subpopulations are already at a higher risk of the COVID-19 infection and mortality and increased level of pandemic-induced socioeconomic burden, such that it is critical to ensure that digital inequalities do not exacerbate the disparate influences of the pandemic even further[2].

Here, we harness the centrality of web search engines for online information access to conduct a retrospective observational study on how the pandemic may have shifted people's engagement towards or away from digital resources and how such shifts may reflect societal disparities. This study extends prior work on pandemic-related disparities which are largely driven by datasets focused on hospitalization or case/fatality rates[8–11]. Leveraging web search interactions enables us to capture the use of critical digital resources such as online educational sites in response to school closures, online food delivery information in response to restaurant closures, online social interactions in response to physical distancing and travel restrictions, or online unemployment and economic assistance in response to economic instability during the pandemic. Given that the pandemic has impacted everyone's web search behaviors nationally[1,22], our goal is to identify differences across subpopulations in their behavioral responses to the pandemic and to discover potential barriers and challenges in accessing critical resources on the web.

Prior work on understanding digital disparities has relied on costly surveys, interviews, or self-reports[23–25] that require direct engagement with the study population for subjective recounting of their past behaviors rather than passively observing their actual behaviors. Datasets from specific service providers (e.g., Wikipedia[26], Zearn.org[27,28]), domain (e.g., telehealth[29], eHealth[18]) or geographic areas (e.g., Northern California[18]) do not capture digital behaviors across a broad spectrum of human needs and subpopulations and at fine geo-temporal granularities. Macroeconomic measures, such as unemployment rates, do not capture potentially unmet needs or access barriers (e.g., confusion around unemployment benefits[30–32]).

Conversely, web search logs are routinely collected at near-real time and at large scales, providing unique opportunities to unobtrusively examine digital behaviors across a wide range of topics, geographies, and subpopulations as well as highlighting potential barriers and changes to such engagement behaviors[33]. In fact, web search logs have enabled studies of human behaviors across many different domains[34–37], time[38–41], location[42,43], and to make inferences about the future or to identify risk factors[22,44–47]. In the context of the COVID-19 pandemic, such data has stimulated a prolific range of research on physical[22,48], psychological[49], and socioeconomic[50,51] well-being[1].

We analyze 55 billion everyday web search interactions during the COVID-19 pandemic across 25,150 US ZIP codes. Our dataset includes anonymized search queries to the Bing search engine and subsequently clicked web site URLs from those queries. Each search interaction is classified into categories of health, education, economic assistance, and food access, covering a range of critical resource needs (Supplementary Table S4). We link the search interactions from each United States ZIP code to their respective per-ZIP code census variables that broadly cover five social determinants of health (SDoH) categories defined by the US Department of Health[52]: (1) Healthcare Access and Quality (through health insurance coverage), (2) Education Access and Quality (through education level), (3) Social and Community Context (through race/ethnicity), (4) Economic Stability (through income and unemployment

medRxiv preprint doi: https://doi.org/10.1101/2021.09.14.21263545; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

rate), and (5) Neighborhood and Built Environment (through population density and internet access). We divide our dataset along these SDoH factors and compare the magnitude of change in search behaviors between two ZIP code groups during the pandemic, where more or less observed change in search behaviors could indicate a shift in higher or lower demand for information (e.g., health, unemployment) or a shift to digital means to accessing resources (e.g., online remote learning). For example, we split our ZIP codes into low and high income groups (below and above $55,000 median household income) and compare the magnitude of change in health condition information queries (Fig. 1a). To disentangle the confounding effects of SES and race/ethnicity on behaviors and health[53], we compare changes in search behaviors on matched pairs of ZIP codes that are highly similar across these potentially confounding factors (Methods). We isolate the relative changes in search behaviors that occur concurrently with the pandemic using difference-in-differences approach[54], adjusting for yearly and weekly seasonality and for pre-existing, pre-pandemic disparities (Fig. 1b-d, Methods). Thus, we operationalize digital disparities attributable to a single SDoH factor by quantifying the differences in these changes in search behaviors between two subpopulations delineated by that factor (Fig. 1e). Finally, we apply the same process across all SDoH factors (Fig. 1f, Methods).

Health information access

Given the higher rate of pre-existing health conditions, documented disparities in healthcare access, and higher COVID-19 case and mortality rates for low SES subpopulations[8,53], low SES subpopulations might be expected to seek information about their health conditions at higher rates during the pandemic as compared to before. Instead, we find that ZIP codes associated with lower incomes show over a 200 percentage point smaller increase (95% CI $[-287, -152]$) in health condition queries than their higher income counterparts (Fig. 1e). This means that a person who was making one health condition query a month before the pandemic makes about ten such queries a month during the pandemic, but the same person would only make about eight such queries a month if they lived in a ZIP code with lower income. We find that ZIP codes with higher Hispanic population, higher population density, and higher unemployment rate also responded to the pandemic with lower relative change in their health condition queries during the first four weeks (Fig. 1f). While ZIP codes with high (i.e., above population-average) Black population ($\geq$12%) do not seem to be affected as much as high Hispanic population groups during the first four weeks, their response is lower during the months of August to November (Supplementary Fig. S3g). On the other hand, we find that ZIP codes with low education ($\leq$21.1% with bachelor's degrees) make over 70 percentage point more (95% CI $[31, 117]$) health condition queries compared to ZIP codes with higher education (Fig. 1f).

Online health information seeking behaviors can educate and empower patients to engage in their health care, lead to better health and wellbeing, and reduce health disparities[3,4,55,56]. Prior research has shown that SES and demographics correlated with online health information seeking behaviors, highlighting the digital divide in health information access[56,57]. Unfortunately, our results suggest that racial and economic factors may have contributed to a lessened response in health condition information seeking behaviors during the pandemic, which may exacerbate health disparities down the line[3,58].

Economic assistance access

When we examine unemployment-related search interactions, we find that relative changes in unemployment related search queries (e.g., "eligible for unemployment benefits", "jobless claims") closely follow those of reported unemployment claims by the Bureau of Labor Statistics (Supplementary Fig. S1). However, the impact of the pandemic on ZIP codes with high Black population and their unemployment

medRxiv preprint doi: https://doi.org/10.1101/2021.09.14.21263545; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

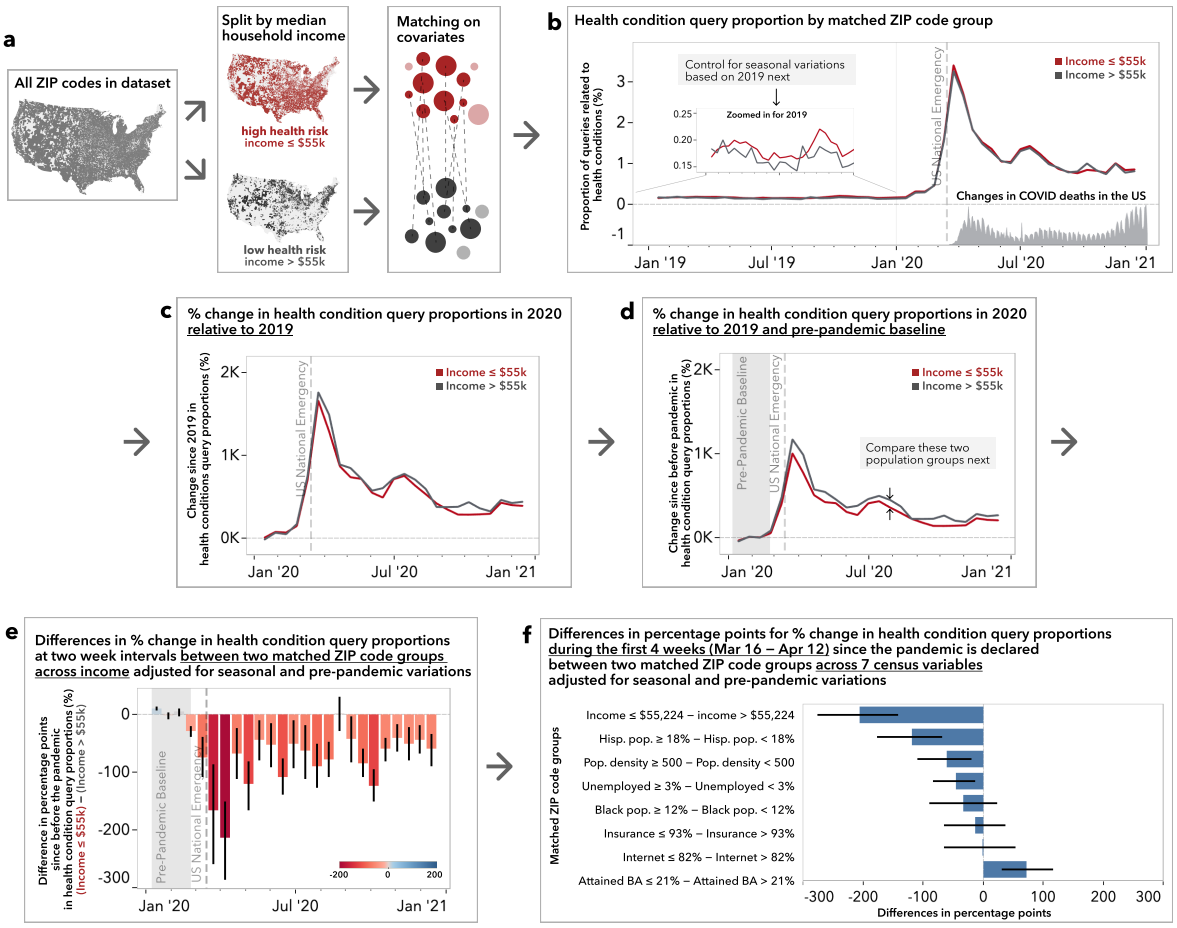


Figure 1. Illustration of the process of quantifying disparities in online health information access between high and low income groups. a, We contrast 25,150 ZIP codes above and below $55,224 median household income. Prior work suggests that ZIP codes with lower income typically have higher health risks, and that income is correlated with many other potential confounders such as race and education[53]. We disentangle the influences of other potential factors by matching each high income population ZIP code with a low income population ZIP code of similar profiles along other covariates (see Methods). **b**, From over 55 billion search queries, we examine the proportion of queries relating to health conditions (e.g., hypertension, diabetes, cancer, coronavirus), across years 2019 and 2020 and across low income population (red) and high income population (gray), using a set of regular expressions (see Supplementary Table S4). **c**, We remove seasonal and weekly variations by subtracting the proportion of queries in 2019 from 2020 while aligning the days of the week. **d**, From all data points, we further remove the seasonally adjusted pre-pandemic baseline, shaded in gray (January 6 - February 23, 2020), to quantify changes introduced during the pandemic. **e**, We compute the difference in the changes in health condition queries during the pandemic by subtracting the low income group's changes from the high income group's changes. The differences between the two groups during the pre-pandemic period average to zero because we controlled for seasonal and pre-pandemic differences between the groups in the previous steps. We observe that low income ZIP codes experienced almost 200% less change in health condition queries compared to that of the high income groups right after the US national emergency is declared. This difference already accounts for each group's own baseline and, therefore, indicates a widening disparity in digital health engagement. Error bars in all charts indicate 95% confidence intervals obtained through bootstrapping (N=500). **f**, Finally, we repeat this process across all SDoH factors. Here, blue bars indicate the differences in percentage points across two matched comparison groups during the first four weeks since the declaration of the pandemic in the US.

medRxiv preprint doi: https://doi.org/10.1101/2021.09.14.21263545; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

search queries is almost three times the amount on ZIP codes with low Black population (Fig. 2a), with 3026% increase in query proportions for ZIP codes with high Black population compared to over 1365% increase for their counterparts, resulting in a 1,661 percentage point difference (95% CI $[260, 2374]$) (Fig. 2b). We find another surge in search queries that result in over 1000% increase in the proportion of clicks into state-specific unemployment websites past July 2020, at which point the expanded federal supplement to unemployment insurance benefits expired (Fig. 2c). During the month of August, ZIP codes with higher Black and higher Hispanic population present 789 (95% CI $[595, 957]$) and 716 (95% CI $[351, 1043]$) percentage points more in their change in clicks to unemployment sites, indicating that Black and Hispanic groups may have required additional long-term unemployment benefits. Conversely, ZIP codes with lower education levels experienced 517 percentage points less (95% CI $[-1009, -81]$) in the change in state unemployment site visits (Fig. 2d).

Internet access and digital engagement is an important form of human capital that allows efficient access to information, increases in economic opportunities, and ultimately leads to better prospects and economic stability[59]. During economic hardships and especially during the pandemic, the internet can be an efficient way for government and institutions to deliver interventions and can lower barriers to accessing economic assistance or welfare services (e.g., https://www.usa.gov/food-help provides a comprehensive list of resources for food assistance). Unfortunately, the pandemic imposes multiple layers of barriers to accessing crucial economic assistance because low SES subpopulations are more likely to suffer economically from the pandemic[60] and deprioritize improving digital access as a consequence[3]. To understand the economic impact of the pandemic, we examine behaviors for accessing unemployment and financial assistance on the web.

Our results suggest that lower education level, even while controlling for other factors such as internet access and SES, is linked to barriers to accessing unemployment sites on the web. Since our difference-in-differences analysis already accounts for existing disparities through normalization, this difference in the increase in unemployment site visits may signal of widening disparities in employment. Such "interest" in digital unemployment resources is not captured in reported claims that measure unemployment claims that are actually submitted, but can be readily observed in web search logs. The discrepancy between unemployment interests expressed online and officially submitted claims may suggest potential barriers in successful submission of benefits application (e.g., confusion, eligibility[30,31]). Coupled with low recipiency rate of unemployment benefits[61] and the association between unemployment accessibility and suicide risks[62], the mismatch between demands and claims are concerning.

April of 2020 was a prime time for financial assistance related queries (e.g., "loan forgiveness", "stimulus check deposit") because the first stimulus checks were deposited on April 11, 2020 (Fig. 2e). We find that financial assistance related queries increased by over 15,000% in mid-April on average, but ZIP codes with higher Black population experience 5,119 percentage points less change (95% CI $[-8809, -1407]$) in financial assistance related queries between April 13 and May 10, 2020 (Fig. 2f). That means that if a person made one financial assistance related query per month in mid-April of 2019, that person makes 167 such queries per month in mid-April during the pandemic, but only 116 queries if that person was from a ZIP code with higher Black population. Since we successfully controlled for other potential confounding factors such as income and education in our comparison, as shown in Supplementary Table S8, our result points to race as a plausible cause for such disparity. Our finding highlights the need to further investigate potential barriers or causes that disparately prevent Black subpopulations from responding to pandemic-induced stimulus demands on the web.

medRxiv preprint doi: https://doi.org/10.1101/2021.09.14.21263545; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

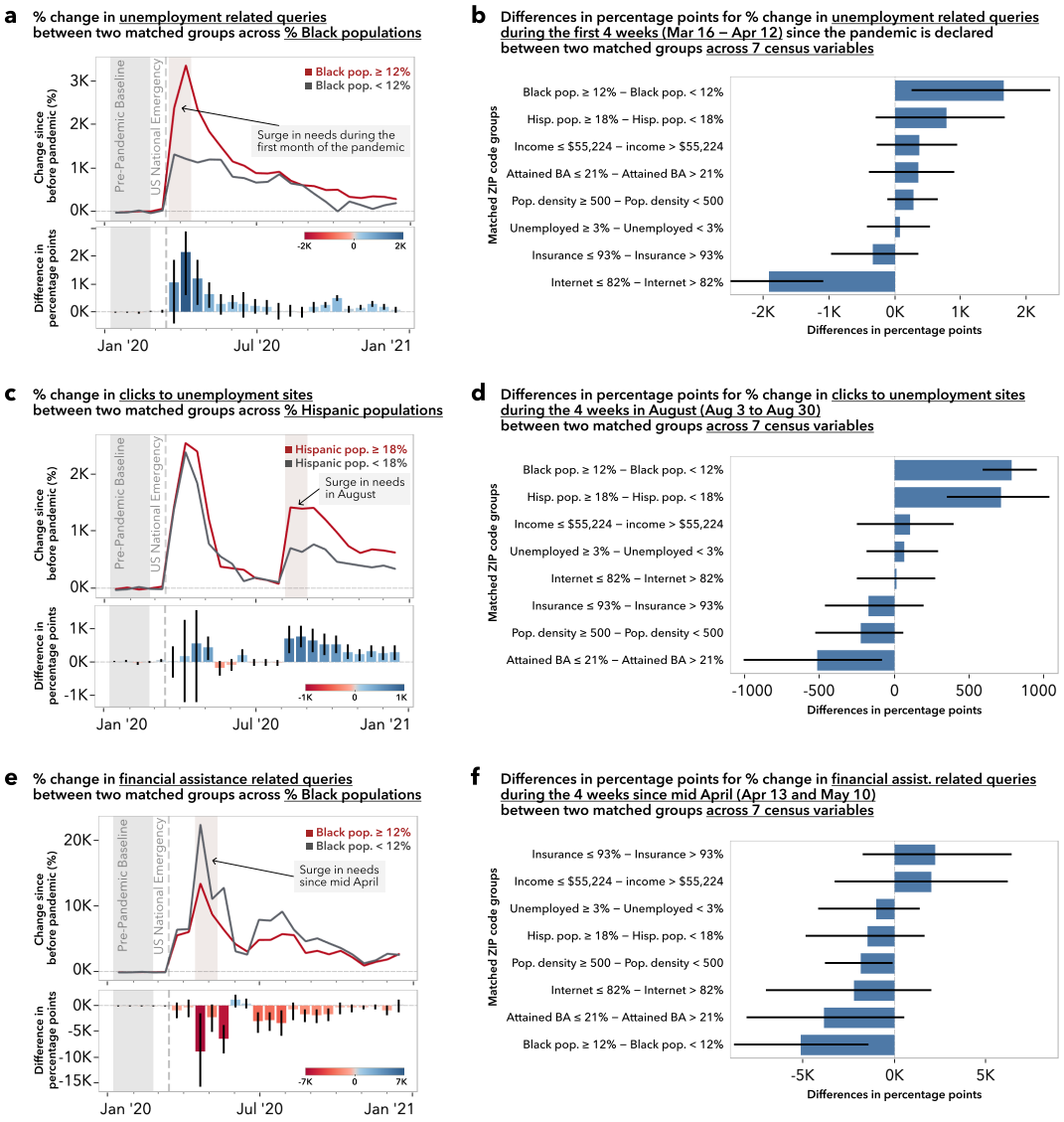


Figure 2. Disparities in online economic assistance access. a, The surge in unemployment related search queries peaks during the first month since the declaration of the pandemic and taper off over the year 2020. During this first month, ZIP codes with high Black population ($\geq 12\%$) have expressed up to 3,358% more unemployment related queries while ZIP codes with low Black population ($< 12\%$) have expressed 1,320% more. **b**, Across the seven census variables, ZIP codes with high Black, high Hispanic, and low income populations experience greater changes in unemployment related queries during this first month. **c**, When we examine search queries that led to clicks in state unemployment sites, we see a second surge in August, with ZIP codes with high Hispanic population ($\geq 18\%$) experiencing more than double the change in clicks in state unemployment sites compared to ZIP codes with low Hispanic population ($< 18\%$). **d**, We observe that ZIP codes with high Black and Hispanic populations experience greater change in clicks in unemployment sites during the month of August, but ZIP codes with low educational attainment express less change in clicks in unemployment sites. **e**, Search queries related to financial stimulus were at their peak in late April, right after the time that the first stimulus checks were deposited on April 11. **f**, However, throughout the year and especially during the four weeks since mid-April, ZIP codes with high Black population experienced smaller change in financial stimulus related queries than ZIP codes with low Black population.

medRxiv preprint doi: https://doi.org/10.1101/2021.09.14.21263545; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Shift to digital resources

The COVID-19 pandemic brought a rapid and massive digital transformation to lives as mandated lockdowns forced people to transform and reimagine traditional interpersonal connections (e.g., going to school, getting food, or meeting friends) into virtual and digital ones. Unfortunately, digital inequalities worsen social and material deprivations and perpetuate existing disadvantages into a "digital vicious cycle"[2,64]. To understand the impact of the pandemic on reinforcing this vicious cycle, we investigate two classes of digital resources particularly influenced when traditional in-person access was impossible or severely limited: online remote learning and online food delivery services.

Statewide mandates in the US required many schools to close in-person learning as early as March 16, 2020[65], and school districts scrambled to implement remote learning alternatives. Many parents, students, and teachers turned to free online resources such as Khan Academy to fill the gaps temporarily or permanently[66]. There were also reported disparities in access to technologies or live virtual learning as well as absenteeism that stymied low income students[67]. When we examined search queries that result in visits to free online learning resources (e.g., coursera.org, khanacademy.org), we found that ZIP codes with lower income and higher Hispanic population exhibited only half to two-thirds of the increase (percentage point difference 95% CI $[-227, -109]$ and $[-202, -46]$ respectively) in those queries relative to their counterpart groups (Fig. 3a). If a person made one search-led click to online learning sites per month before the pandemic, that person would make 5 such clicks per month during the pandemic, but only 3 such clicks would be made if that person was from a ZIP code with lower income or higher Hispanic population, even after controlling for internet access (Fig. 3b). ZIP codes with high Black population and high population density show a similar trend. Even though these free online learning resources are designed to be accessible and flexible, helping students to go at their own pace, we find that low income and highly diverse subpopulations did not leverage them at the same level as their counterpart groups during the pandemic. In addition, school districts in low SES neighborhoods were more likely to be closed during the pandemic and less equipped to provide remote learning or at-home assignments, greatly reducing opportunities for both in-person and online learning for students with negative educational outcomes[68,69]. Our findings suggest that there exists digital exclusion and the unintended consequences of the public health policies that perpetuate a myriad of disadvantages, as education is such a crucial factor in digital literacy, socioeconomic status, and health.

COVID-19 fundamentally changed people's purchasing and spending behaviors, as many of the restaurants, stores, and non-essential businesses were closed to in-person shopping[70]. Spending on food delivery and groceries also increased significantly during the pandemic, with more people eating at home with higher utilization of online e-commerce platforms for accessing food and groceries[70,71]. When we examine search queries for online food delivery (e.g., "grocery delivery", "deliver food"), we find that online food delivery queries increased by over 500% for ZIP codes with lower Black population while those with higher Black population only increased by over 170% (percentage point difference 95% CI $[-382, -188]$, Fig. 3c,d). We found similar lessened engagement in online food delivery searches for lower income and higher Hispanic subpopulations (95% CI $[-200, -29]$ and $[-140, -24]$ respectively, Fig. 3d). These findings could be explained by the fact that low income subpopulations receive and seek more food assistance and tend to eat food away from home less frequently[72] and that such online food delivery services may not be accessible because they accompany higher markup and delivery surcharge.

ZIP codes with low education subpopulations also experienced 301 percentage point more increase (95% CI $[167, 419]$) in queries for seeking food assistance (e.g., "Supplemental Nutrition Assistance Program", "help with food stamps", "free and reduced lunch", Fig. 2e,f) relative to their high education counterparts. Unfortunately, those that relied on these traditional food assistance programs were left with

medRxiv preprint doi: https://doi.org/10.1101/2021.09.14.21263545; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

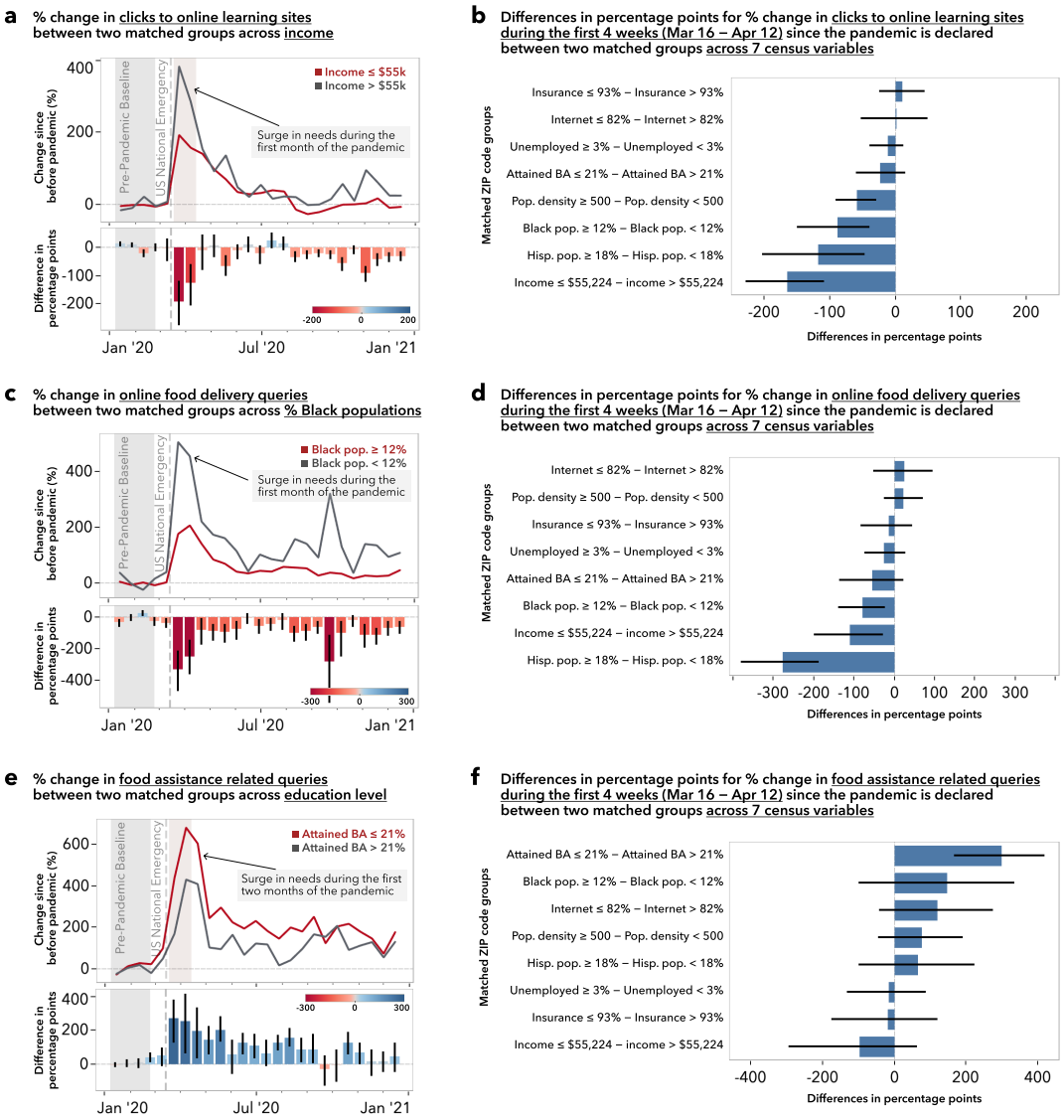


Figure 3. Disparities in shifting to digital resources. Engagement in critical online resources necessary during prolonged lockdown or school/business closures were lower for ZIP codes with low income and high racial/ethnic diversity. **a**, Online learning sites played a significant role in filling in the gaps introduced by school closures at the beginning of the pandemic with over 200% increase in engagement. **b**, However, ZIP codes with lower income and higher Black population tend to access online learning resources less. In the new academic year (after August), while low income group continued to show lower engagement, ZIP codes with higher Black population show slightly higher engagement in online learning sites. **c**, With mandated restrictions on social gatherings, populations have transitioned to online-mediated social activities during the pandemic. **d**, For ZIP codes with higher population density, where lockdown measures were more strictly enforced due to higher case and mortality rates[63], changes in online social activities search were higher. However, we see that ZIP codes with higher Hispanic population show less change in online social engagement, even after controlling for population density or internet access, indicating potential racial/ethnic barrier or preference to accessing online social activities. **e**, Online food delivery services were also in high demand due to restaurant closures, with over 250% increase at the beginning of the pandemic. **f**, However, ZIP codes with higher Black population and lower education showed smaller change in online food delivery search throughout the pandemic, regardless of population density, income, or internet access.

medRxiv preprint doi: https://doi.org/10.1101/2021.09.14.21263545; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

severely limited choices during the pandemic because these programs do not extend to online purchase or delivery services[73]. Our findings highlight a potential gap between the increased food assistance need, as illustrated by the increase in the online information seeking behavior *about* food assistance, and the ability to actually procure food goods through online food purchase and delivery services.

Discussion

This study provides quantitative evidence that the current COVID-19 pandemic was associated with a widening of disparities in access, engagement, and utilization. Prior studies have shown that access to digital resources and information and incorporation of such digital technologies in everyday lives from childhood are crucial for gaining upwards mobility[3]. Although SES is an important factor in disparities in digital access, prior research has shown that SES also impacts the levels of web expertise and utilizing those resources in more information-seeking activities[17]. Low SES populations suffer from the lack of training and educational support to build the necessary skills to make efficient use of digital access and tools[13], highlighting that simply making the internet more accessible may not level the playing field[74]. In the context of the current COVID-19 pandemic, where digital access and resources became more critical due to prolonged at-home isolation and restrictions on in-person activities, people from low SES backgrounds may experience the compounding effects of multiple potential disadvantages. Our findings confirm such compounding effects. We find that lack of internet access can be attributed to reduced engagement in web unemployment resources, implying that internet access is crucial ingredient to economic stability and most important for those experiencing significant economic burden. Furthermore, we discover that low income and high Hispanic populations are taking less advantage of web health resources, presenting similarly unfortunate consequences of eHealth initiatives that may disproportionately benefit digitally advantaged subpopulations[3]. Additionally, the findings confirm that low SES populations fell behind in the digital shift catalyzed by the pandemic[2], as low income and more diverse subpopulations did not leverage online learning or online food delivery resources as much as their high income and less diverse counterpart subpopulations.

We note the inherent limitations of studying digital engagement using digitally obtained data: This and other studies with online data can inadvertently exclude those who leave no or very little digital footprint[3]. Our information sources provide signals about levels of activity, but we cannot study details of changes in types of access if there is no engagement. Our analysis is also limited to the footprint of Bing as one of several search engines used for online information access, and Bing's user population may not be fully representative of the United States population. Our study carefully controls for internet access, as measured by the census, during the analysis such that any observed effects cannot be explained by differences in population internet access. Our approach combines search log data with socioeconomic and environmental variables that are routinely captured through census tracks to examine the influence of such census variables on potentially a diverse array of different topic areas of digital engagement at population scales. Our observed changes can only be attributed to ZIP code levels and not individuals because individual-level SDoH factors are not available and to preserve anonymity. Like any retrospective observational study, the potential for unobserved or uncontrolled confounding prevents us from making causal claims. However, we adjusted for observed confounding through a matching-based and difference-in-difference based methodology (Methods). Our data cannot be used to discern whether different access behaviors are due to the lack of web expertise, the lack of awareness of the information value, or the lack of intangible resources like time or energy. Thus, we see value in follow-up, small-scale focused studies aimed at contextualizing individuals' experiences of the crisis and measuring the effects of population-specific interventions[2]. These population-specific interventions could include education around web

medRxiv preprint doi: https://doi.org/10.1101/2021.09.14.21263545; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

expertise or digital know-how or may include non-digital methods, because traditional methods (e.g., text messaging, handouts) have been shown to work better for low SES populations[67]. Although the SDoH factors and outcomes reviewed in our analysis may not be the only variables of interest, our matching-based approach provides methodological robustness relative to traditional univariate analyses in observational studies by controlling for observed covariates[75,76]. Future research aimed at understanding digital disparities, therefore, must acknowledge the correlations between different SES, race/ethnicity, and social determinants of health[77] and leverage methods that embrace examining the intersectionality[78].

This study presents a new methodology for digital disparities research by demonstrating that web search logs can be harnessed to deliver key insights about the impact of global crises on widening of digital disparities. Our observational study design is able to scale to a large population (billions of queries by millions of people) to quantify the disparities in digital engagements. Building on prior disparities research that advocated for comprehensive look at SES factors including race/ethnicity[53,77], our study emphasizes the inclusion of a broad set of factors and outcomes representative of the SDoH. Through the lens of SDoH factors, our findings highlight under-served subpopulations that may be struggling to overcome the economic burdens through online financial assistance and unemployment resources, that may be facing barriers in maintaining the necessary level of online information access for health, education, and food, and for whom to target public health interventions to prevent further widening of digital disparities.

Methods

Data Set and Study Population

Our source data set consists of a random sample of 57 billion de-identified search interactions in the United States from years 2019 and 2020 from Microsoft's Bing search engine. Each search interaction includes the search query string, URLs of all subsequent clicks from the search result page, timestamp, and ZIP code from reverse IP lookup. We excluded search interactions from ZIP codes with less than 100 queries per month so as to preserve anonymity. All data were deidentified, aggregated to ZIP code levels or higher, and stored in a way to preserve the privacy of the users and in accordance to Bing's Privacy Policy. Our study was approved by the Microsoft Research Institutional Review Board (IRB).

Our goal is to characterize the impact of social economic status on digital engagement outcomes. To account for well-known issues associated with residential segregation and socioeconomic disparities[13,79], we use ZIP codes as our geographic unit of analysis. We leveraged available ZIP code level American Community Survey estimates using the Census Reporter API[80] in order to characterize the ZIP codes in our data set.

Because of the multidimensional nature of socioeconomic status and its association to health outcomes, it is important to include relevant socioeconomic factors[77]. Therefore, we examined eight census variables that represent all five categories of the social determinants of health (SDoH) defined by the US Department of Health[52] to cover a broad range of socioeconomic and environmental factors. Under Healthcare Access and Quality, we included the percentage of population with health insurance coverage (Table B27001). Under Education Access and Quality, we included the percentage of population that attained Bachelor's degree or higher (Table B15002). Under Social and Community Context, we included the percentage of population with Hispanic origin (Table B03003) and the percentage of population with Black or African American alone (Table B02001). Under Economic Stability, we included the median household income (Table B19013) and the percentage of civilian labor force that is unemployed (Table B23025). Under Neighborhood and Built Environment, we included the percentage of population with broadband or dial-up internet subscription (Table B28003) and the population density. We computed per ZIP code population density by joining area measurements from ZIP Code Tabulation Areas Gazetteer Files[81] and

medRxiv preprint doi: https://doi.org/10.1101/2021.09.14.21263545; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

total population (Table B01003). We joined search interaction data with the above SDoH factors on ZIP codes and excluded ZIP codes that did not have either search interactions or census data. The resulting 55 billion search interactions covered web search traffic from 25,150 ZIP codes in the US, and these ZIP codes represents 97.2% of the total US population. Supplementary Table S1 provides per-ZIP code summary statistics of our dataset.

Disparities in Digital Engagement

We leverage interactions with search engines to obtain signals about digital engagements where everyday human needs are expressed or fulfilled through a digital medium, in our case Bing[1]. To gain a nuanced understanding of these search interactions, we categorize each search interaction into topics ranging from health access, economic stability, and education access using detectors on query strings and subsequently clicked URLs based on regular expressions and basic propositional logic. For example, we compute the proportion of search queries that contain health condition keywords such as cancer, diabetes or coronavirus to quantify the level of engagement in health information seeking behaviors. In another case, we examine search queries that result in subsequent clicks to state unemployment benefit sites to quantify the level of engagement in unemployment benefits. Supplementary Table S4 enumerates the categories we examined with example query strings, URLs, and regular expressions.

From these search interactions, we want to estimate the disparate impact of the pandemic on digital engagement behaviors across SDoH factors in a way that controls for yearly and weekly seasonality and for pre-existing, pre-pandemic disparities in order to highlight where disparities have worsened throughout the pandemic period of March 16, 2020 to December 27, 2020. To do this, we first use a difference-in-differences method[54] to correct for seasonality and volume variations.

After we categorize each search interaction with our categories of interest, we count and aggregate them per calendar day and per ZIP code (Fig. 1a). We compute the proportion of the total query volume represented by each category at this daily increment to remove undesired variations in query volume over time (Fig. 1b). We denote the digital engagement at time t in category c as the fraction of the total number of queries at time t: $E(t,c) = N(t,c)/N(t)$. From this, we control for yearly seasonal variations by subtracting the digital engagements of 2019 from that of 2020: $E(t^{2020},c) - E(t^{2019},c)$. People tend to behave differently on weekends, and we observed a 7-day periodicity in our data, sometimes known as the "weekend effect"[82]. Therefore, when comparing two years, it is important to account for the weekend effect. In order to highlight actual difference that are not explained by weekend mismatches across years, we aligned the day of the week between both years (i.e., Monday, January 6, 2020 is aligned to Monday, January 7, 2019). In addition, we ensured that our comparison analysis included all seven days of the week (i.e., look at means across one or multiples of a full week) (Fig. 1c).

Finally, to compute the change in digital engagement during the pandemic since the time at which the US national emergency was declared on March 16, 2020, we subtract the query proportions between January 6, 2020 and February 23, 2020, a time period we defined as the "pre-pandemic baseline" (Fig. 1d). Even though the national emergency was declared three weeks later, we use February 23, 2020 as the cut-off because individual states declared a state of emergency at different times between February 29 and March 15 of 2020 and to avoid partial weeks in our analysis. This process results in the change in digital engagement most likely attributable to the pandemic. Our estimate of the *relative change in digital engagement* in category c between before and during the pandemic is defined as:

$$C(t_{\text{before}};t_{\text{during}},c) = \left[E(t^{2020}_{\text{during}},c) - E(t^{2019}_{\text{during}},c)\right] - \left[E(t^{2020}_{\text{before}},c) - E(t^{2019}_{\text{before}},c)\right]$$

medRxiv preprint doi: https://doi.org/10.1101/2021.09.14.21263545; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Or the relative *percentage change* in digital engagement C_{perc} is expressed as:

$$C_{\text{perc}}(t_{\text{before}};t_{\text{during}},c) = \frac{\left[E(t^{2020}_{\text{during}},c) - E(t^{2019}_{\text{during}},c)\right] - \left[E(t^{2020}_{\text{before}},c) - E(t^{2019}_{\text{before}},c)\right]}{\left[E(t^{2020}_{\text{before}},c) - E(t^{2019}_{\text{before}},c)\right]} \times 100$$

Next, we aggregate these changes in digital engagements across two comparison ZIP code groups for each SDoH factor. For example, if we are examining the impact of having low income on the changes in digital engagement during the pandemic, we compare the average change in digital engagement of the low income ZIP codes with the average change of the high income ZIP codes (Fig. 1d). Thus, we operationalize digital disparities attributable to a single socioeconomic or environmental factor by quantifying the differences in these changes in search behaviors between two subpopulations delineated by that factor (Fig. 1). In our analysis, we report the change in digital engagement as the percentages of the pre-pandemic baseline, C_{perc}, where 0% denotes no change. We report the disparities in digital engagement between two comparison ZIP code groups as the percentage point difference where 0 denotes no difference (Fig. 1e,f). We formalize disparities in digital engagement in category c during the pandemic between high-risk ZIP code group g_{high} and low-risk ZIP code group g_{low} as:

$$D_{\text{perc}}(t_{\text{before}};t_{\text{during}},g_{\text{low}};g_{\text{high}},c) = C^{g_{\text{high}}}_{\text{perc}}(t_{\text{before}};t_{\text{during}},c) - C^{g_{\text{low}}}_{\text{perc}}(t_{\text{before}};t_{\text{during}},c)$$

To obtain non-parametric confidence intervals, we conducted bootstrapping with replacement during this aggregation step (N=500). Supplementary Figures S2-S15 illustrate percent changes in each query category for each of two matched groups and their differences in percentage points across all SDoH factors.

Matching to Control for Confounding

Our goal is to quantitatively estimate the independent impact of one socioeconomic factor on digital engagements while controlling for other factors to understand if and how a single factor independently influences digital behaviors during a global crisis such as the COVID-19 pandemic. Specifically, we are interested in the impact of the eight SDoH factors: median household income, % unemployed, % with insurance, % with Bachelor's degree or higher, population density, % Black population, % Hispanic population, and % with internet access. We use % with internet access primarily to control the level of digital access because internet access is necessary for web search. Many of the socioeconomic and racial variables are known to be correlated[53,77,83]. This means that a univariate analysis of outcomes along one SDoH factor would likely be confounded by multiple other variables. In fact, within our dataset, we observed high correlation among many SDoH factors examined (Supplementary Table S3). For example, median household income of the ZIP codes in our dataset is negatively correlated with the percentage of Black population (Pearson $r = -0.23$) and is positively correlated with internet access (Pearson $r = 0.66$). Comparing high and low income groups without considering other factors would result in two groups of uneven distributions of race and internet access, among many other factors. Therefore, it is important to consider these factors jointly and adequately control for SES factors when analyzing outcome disparities[53,77].

To disentangle the individual independent contributions of these SDoH factors when measuring disparities in online information access between groups, we employ a matching-based approach[75,76], designed to create a comparable set of groups with similar covariate distributions. Because of the high degrees of spatial segregation in the US[13,79], matching every ZIP code can be challenging. For example, for every ZIP code with low income and high Black population, it is difficult to find a unique ZIP code

medRxiv preprint doi: https://doi.org/10.1101/2021.09.14.21263545; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

with high income and high Black population. Therefore, we perform one-to-one matching of ZIP codes with replacement and achieve better matches (i.e., lower bias). Theoretically, this is at the expense of higher variance, but given the size of our dataset, this downside was not a problem in practice. We use the MatchIt package[84] with the nearest neighbor method and Mahalanobis distance measure to perform the matching.

Determining Treatment and Control Groups

For each of the SDoH factors, we first split all available ZIP codes into treatment and control groups using a threshold. We use a value close to the median to split the population into two groups for median household income ($55,224), % unemployed (3.0%), % with insurance (92.7%), % with internet access (81.8%), and % with Bachelor's degree or higher (21.1%) because the mean and median of those factors across the ZIP codes are similar. In other cases, the distribution across the ZIP codes are highly skewed. For race/ethnicity, we use the rounded percentage of the national population for that race/ethnicity (12% for Black and 18% for Hispanic populations). For population density, we follow previous practices of urban-rural classification (500 people per square mile)[85]. Supplementary Tables S1 and S2 outline descriptive statistics of our ZIP codes across SDoH factors as well as the national average and our chosen cutoff thresholds.

We consistently defined the treatment group as "high-risk" – typically suffering from disadvantaged SES and digital inequalities[58], so the treatment groups are low income, high racial/ethnic diversity, low education, high unemployment rate, low insurance rate, low internet access, and high population density. For example, for income, we split the ZIP codes into a high-income group (median household income $>$ $55,224) and a low-income group (median household income $\leq$ $55,224), where the low-income group is the treatment group. Then, for each treatment ZIP code, we look for a control (i.e., "low-risk") ZIP code that closely matches it on all other SDoH factors (i.e., $|SMD| < 0.25$ to generate a matching pair of ZIP codes). We performed this matching on all ZIP codes, and we discarded ZIP codes for which we cannot find a good match. As demonstrated in Supplementary Table S6, this process retains at least 99.8% of the treatment ZIP codes in our matching process and the discarding of ZIP codes is a rare exception.

Evaluating Quality of Matching Zip Codes

To gauge whether two ZIP code groups are similar across the SDoH factors and to determine the quality of matching while minimizing potential confounding effects of these factors, we leverage Standardized Mean Difference (SMD) across ZIP code groups as our measure of comparative quality. The SMD is used to quantify the degree to which two groups are different and is computed by the difference in means of a variable across two groups divided by the standard deviation of the one group (often, the treated group)[76,86,87]. In our analysis, we use $|SMD| < 0.25$ across all our SDoH factors as a criterion to determine that two groups are comparable, following common practice[75,87]. For example, when we split our ZIP codes in half along median household income to create a high-income ZIP code group (median household income $>$ $55,224) and a low-income ZIP code group (median household income $\leq$ $55,224) and examine the SMD of other SDoH factors, we find that all SDoH factors except % Hispanic population and population density fail to achieve the necessary matching criteria of $|SMD| < 0.25$ prior to matching. This means that low-income ZIP codes are more likely to have less internet access, lower education attainment, less health insurance, more unemployment, and higher Black population. We perform this evaluation process for all comparison groups to find that correlations among all SDoH factors pose threats to validity in univariate analyses. Supplementary Table S5 summarizes mean SMD if we were to directly compare two ZIP code groups created by splitting the ZIP codes along the chosen split boundaries. Instead of such direct comparison, we perform matching and tune the caliper of the matching algorithm to determine a good match and to meet the $|SMD| < 0.25$ criterion between the two comparison

medRxiv preprint doi: https://doi.org/10.1101/2021.09.14.21263545; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

groups across all covariates. Supplementary Table S6 summarizes the result of the matching operation with the maximum $|SMD|$ being below 0.25, that is ensuring comparability across all covariates, between two ZIP code groups along all SDoH factors. Supplementary Tables S7-S22 enumerate pre- and post-matching balance assessments between groups for each SDoH factor.

Estimating the Effect Size

After identifying treatment and control ZIP code groups with comparable distributions along all SDoH factors, we compare the outcomes (i.e., constructs of digital engagement such as online access to health condition information) between the matched ZIP code groups. This matching process estimates the Average Treatment Effect on the Treated (ATT), or specifically the effects of having low income on digital engagement while removing plausible contributions from all other observed factors. Due to the segregation and inequalities revealed by these factors, estimating the Average Treatment Effect (ATE) is practically impossible. One may opt to compute a local average treatment effect (LATE) and discard a large fraction of the U.S. population. However, such local estimates are easily misleading when the underlying population is not well understoodd and they fail to capture key populations in our study (e.g., low income and high Black populations). The ATT estimates in this study provide actionable insights on the effects of being at high risk (e.g., low income, low education, high racial/ethnic diversity) that can be used to suggest interventions to mitigate or reduce risk.

Data availability

Raw US census data are publicly available through the Census Reporter API (https://censusreporter.org/). Geographical area measurements are available through the US Census Bureau (https://www.census.gov/geographies/reference-files/2010/geo/state-area.html). Seasonally adjusted US unemployment claims data for 2020 is available through the US Department of Labor (https://oui.doleta.gov/unemploy/claims.asp). The Bing data that support the findings of this study are available on request from the corresponding author with a clear justification and a license agreement. The Bing data are not publicly available.

Code availability

Source code used for processing and analysis of the data is available on request from the corresponding author.

Acknowledgments

We thank E. Pierson, the University of Washington Behavioral Data Science Group, Microsoft Research Human Understanding and Empathy Group, participants at seminars and talks for their support and comments. This research was supported by Microsoft Research. T.A. was supported by NSF grant IIS-1901386, NSF grant CNS-2025022, NIH grant R01MH125179, Bill & Melinda Gates Foundation (INV-004841), and the US Office of Naval Research (#N00014-21-1-2154).

Author affiliations

Allen School of Computer Science & Engineering, University of Washington (Jina Suh, Tim Althoff); Microsoft Research (Jina Suh, Eric Horvitz, Ryen W. White)

medRxiv preprint doi: https://doi.org/10.1101/2021.09.14.21263545; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Author contributions statement

J.S., E.H., R.W., and T.A. were involved with the conceptualization of the study, and contributed to the design and refinement of the methodology. J.S. conducted data collection and analysis. All authors interpreted the data, drafted the manuscript, and critically contributed to the important intellectual content of the manuscript.

Ethics declarations

The authors declare no competing interests. The study was approved by the Microsoft Research Institutional Review Board (IRB).

References

1. Suh, J., Horvitz, E., White, R. W. & Althoff, T. Population-scale study of human needs during the covid-19 pandemic: Analysis and implications. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, 4–12 (2021).
2. Beaunoyer, E., Dupéré, S. & Guitton, M. J. COVID-19 and digital inequalities: Reciprocal impacts and mitigation strategies. *Comput. human behavior* **111**, 106424 (2020).
3. Robinson, L. *et al.* Digital inequalities and why they matter. *Information, communication & society* **18**, 569–582 (2015).
4. Hale, T. M., Goldner, M., Stern, M., Drentea, P. & Cotten, S. R. Patterns of online health searching 2002–2010: Implications for social capital, health disparities and the de-professionalization of medical knowledge. In *Technology, communication, disparities and government options in health and health care services* (Emerald Group Publishing Limited, 2014).
5. Feinstein, J. S. The relationship between socioeconomic status and health: a review of the literature. *The Milbank Q.* 279–322 (1993).
6. Solar, O. & Irwin, A. A conceptual framework for action on the social determinants of health. (2010).
7. Williams, D., Costa, M., Odunlami, A. O. & Mohammed, S. Moving upstream: how interventions that address the social determinants of health can improve health and reduce disparities. *J. public health management practice : JPHMP* **14**, S8–17 (2008).
8. Chowkwanyun, M. & Reed, A. Racial Health Disparities and Covid-19 - Caution and Context. *The New Engl. journal medicine* (2020).
9. Yehia, B. R. *et al.* Association of race with mortality among patients hospitalized with coronavirus disease 2019 (COVID-19) at 92 US hospitals. *JAMA network open* **3**, e2018039–e2018039 (2020).
10. Chang, S. *et al.* Mobility network models of COVID-19 explain inequities and inform reopening. *Nature* **589**, 82–87 (2021).
11. Yancy, C. W. COVID-19 and african americans. *Jama* **323**, 1891–1892 (2020).
12. Anderson, M. & Vogels, E. A. Americans turn to technology during COVID-19 outbreak, say an outage would be a problem. *Pew Res. Cent.* (2020).
13. Fong, E., Wellman, B., Kew, M. & Wilkes, R. Correlates of the digital divide: Individual, household and spatial variation. In *Office of Learning Technologies, Human Resources Development* (Citeseer, 2001).

medRxiv preprint doi: https://doi.org/10.1101/2021.09.14.21263545; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

14. Hargittai, E. Second-level digital divide: Mapping differences in people's online skills. *arXiv preprint cs/0109068* (2001).

15. Hargittai, E. & Hinnant, A. Digital inequality: Differences in young adults' use of the Internet. *Commun. research* **35**, 602–621 (2008).

16. Van Deursen, A. J. & Helsper, E. J. The third-level digital divide: Who benefits most from being online? In *Communication and information technologies annual* (Emerald Group Publishing Limited, 2015).

17. Hargittai, E. Digital na(t)ives? Variation in internet skills and uses among members of the "net generation". *Sociol. inquiry* **80**, 92–113 (2010).

18. Sarkar, U. *et al.* Social disparities in internet patient portal use in diabetes: evidence that the digital divide extends beyond access. *J. Am. Med. Informatics Assoc.* **18**, 318–321 (2011).

19. Bulman, G. & Fairlie, R. W. Technology and education: Computers, software, and the internet. In *Handbook of the Economics of Education*, vol. 5, 239–280 (Elsevier, 2016).

20. Kuhn, P. & Mansour, H. Is internet job search still ineffective? *The Econ. J.* **124**, 1213–1233 (2014).

21. DiMaggio, P. & Bonikowski, B. Make money surfing the web? The impact of Internet use on the earnings of US workers. *Am. Sociol. Rev.* **73**, 227–250 (2008).

22. Lampos, V. *et al.* Tracking COVID-19 using online search. *NPJ digital medicine* **4**, 1–11 (2021).

23. Allen, W. E. *et al.* Population-scale longitudinal mapping of COVID-19 symptoms, behaviour and testing. *Nat. Hum. Behav.* **4**, 972–982 (2020).

24. Hale, T. M., Cotten, S. R., Drentea, P. & Goldner, M. Rural-urban differences in general and health-related Internet use. *Am. Behav. Sci.* **53**, 1304–1325 (2010).

25. Kontos, E., Blake, K. D., Chou, W.-Y. S. & Prestin, A. Predictors of eHealth usage: insights on the digital divide from the Health Information National Trends Survey 2012. *J. Med. Internet Res.* **16**, e3117 (2014).

26. Hargittai, E. & Shaw, A. Mind the skills gap: the role of Internet know-how and gender in differentiated contributions to Wikipedia. *Information, communication & society* **18**, 424–442 (2015).

27. Zearn.org. Press Release: Zearn Provides Real-Time Snapshot on the State of U.S. Math Education Through New OI Economic Tracker by Opportunity Insights. https://about.zearn.org/press-releases/zearn-provides-real-time-snapshot-on-the-state-of-u-s-math-education-through-new-oi-economic-tracker-by-opportunity-insights (Accessed 2021-08-17).

28. Yglesias, M. Reopening schools safely is going to take much more federal leadership. https://www.vox.com/2020/7/8/21314563/school-reopening-testing-money (Accessed 2021-08-17) (2020).

29. Dorsey, E. R. & Topol, E. J. State of telehealth. *New Engl. J. Medicine* **375**, 154–161 (2016).

30. Schwartz, N. D., Hsu, T. & Cohen, P. Stymied in Seeking Benefits, Millions of Unemployed Go Uncounted. https://www.nytimes.com/2020/04/30/business/economy/coronavirus-unemployment-claims.html (Accessed 2021-08-17) (2020).

31. Bhardwaj, P. Most Job Search Requirements for Unemployment Benefits Are Waived. So Why Do State Websites Say Otherwise? https://money.com/unemployment-benefits-job-search-requirements-coronavirus/ (Accessed 2021-08-17) (2020).

medRxiv preprint doi: https://doi.org/10.1101/2021.09.14.21263545; this version posted September 17, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

32. Desilver, D. Not all unemployed people get unemployment benefits; in some states, very few do. https://www.pewresearch.org/fact-tank/2020/04/24/not-all-unemployed-people-get-unemployment-benefits-in-some-states-very-few-do/ (Accessed 2021-08-17) (2020).

33. Dumais, S., Jeffries, R., Russell, D. M., Tang, D. & Teevan, J. Understanding user behavior through log data and analysis. In *Ways of Knowing in HCI* (Springer, 2014).

34. Althoff, T., Horvitz, E. & White, R. W. Psychomotor function measured via online activity predicts motor vehicle fatality risk. *NPJ Digit. Med.* **1** (2018).

35. Ginsberg, J. *et al.* Detecting influenza epidemics using search engine query data. *Nature* **457** (2009).

36. West, R., White, R. W. & Horvitz, E. From cookies to cooks: Insights on dietary patterns via analysis of web usage logs. In *Proceedings of the 22nd international conference on World Wide Web*, 1399–1410 (2013).

37. Weber, I., Garimella, V. R. K. & Borra, E. Mining web query logs to analyze political issues. In *Proceedings of the 4th annual acm web science conference*, 330–334 (2012).

38. Althoff, T., White, R. W. & Horvitz, E. Influence of Pokémon Go on physical activity: study and implications. *J. Med. Internet Res.* **18**, e315 (2016).

39. Althoff, T., Horvitz, E., White, R. W. & Zeitzer, J. Harnessing the web for population-scale physiological sensing: A case study of sleep and performance. In *Proceedings of the 26th international conference on World Wide Web*, 113–122 (2017).

40. Paul, M. J., White, R. W. & Horvitz, E. Search and breast cancer: On episodic shifts of attention over life histories of an illness. *ACM Transactions on Web (TWEB)* **10**, 1–27 (2016).

41. Fourney, A., White, R. W. & Horvitz, E. Exploring time-dependent concerns about pregnancy and childbirth from search logs. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 737–746 (2015).

42. West, R., White, R. W. & Horvitz, E. Here and there: Goals, activities, and predictions about location from geotagged queries. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*, 817–820 (2013).

43. Sadilek, A. *et al.* Machine-learned epidemiology: real-time detection of foodborne illness at scale. *NPJ digital medicine* **1**, 1–7 (2018).

44. D'Amuri, F. & Marcucci, J. 'Google it!' Forecasting the US unemployment rate with a Google job search index. (2010).

45. Choi, H. & Varian, H. Predicting the present with Google Trends. *Econ. record* **88**, 2–9 (2012).

46. Paparrizos, J., White, R. W. & Horvitz, E. Screening for pancreatic adenocarcinoma using signals from web search logs: Feasibility study and results. *J. oncology practice* **12**, 737–744 (2016).

47. White, R. W. & Horvitz, E. Evaluation of the feasibility of screening patients for early signs of lung carcinoma in web search logs. *JAMA oncology* **3**, 398–401 (2017).

48. Lin, Y.-H., Liu, C.-H. & Chiu, Y.-C. Google searches for the keywords of "wash hands" predict the speed of national spread of COVID-19 outbreak among 21 countries. *Brain, Behav. Immun.* (2020).

49. Tubadji, A., Boy, F. & Webber, D. J. Narrative economics, public policy and mental health. *Covid Econ.* **20** (2020).

medRxiv preprint doi: https://doi.org/10.1101/2021.09.14.21263545; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

50. Gupta, S. *et al.* Effects of social distancing policy on labor market outcomes. *NBER Work. paper* (2020).

51. Abay, K. A., Tafere, K. & Woldemichael, A. Winners and losers from COVID-19: Global evidence from Google Search. *World Bank Policy Res. Work. Pap.* (2020).

52. Social Determinants of Health. https://health.gov/healthypeople/objectives-and-data/social-determinants-health (Accessed 2021-08-17).

53. Adler, N. E. & Rehkopf, D. H. US disparities in health: descriptions, causes, and mechanisms. *Annu. Rev. Public Heal.* **29**, 235–252 (2008).

54. Lechner, M. The Estimation of Causal Effects by Difference-in-Difference Methods. *Foundations Trends Econom.* **4**, 165–224, DOI: 10.1561/0800000014 (2011).

55. Hall, A. K., Bernhardt, J. M., Dodd, V. & Vollrath, M. W. The digital health divide: evaluating online health information access and use among older adults. *Heal. Educ. & Behav.* **42**, 202–209 (2015).

56. Cotten, S. R. & Gupta, S. S. Characteristics of online and offline health information seekers and factors that discriminate between them. *Soc. science & medicine* **59**, 1795–1806 (2004).

57. Goldner, M. How health status impacts the types of information consumers seek online. *Information, Community & Soc.* **9**, 693–713 (2006).

58. DiMaggio, P., Hargittai, E. *et al.* From the 'digital divide' to 'digital inequality': Studying Internet use as penetration increases. *Princeton: Cent. for Arts Cult. Policy Studies, Woodrow Wilson Sch. Princet. Univ.* **4**, 4–2 (2001).

59. Dimaggio, P. J. & Bonikowski, B. Make Money Surfing the Web? The Impact of Internet Use on the Earnings of U.S. Workers. *Am. Sociol. Rev.* **73**, 227 – 250 (2008).

60. Fernandes, N. Economic effects of coronavirus outbreak (COVID-19) on the world economy. *Available at SSRN 3557504* (2020).

61. McHugh, R. & Kimball, W. How Low Can We Go? State Unemployment Insurance Programs Exclude Record Numbers of Jobless Workers. *EPI, March* **9** (2015).

62. Kaufman, J. A., Livingston, M. D. & Komro, K. A. Unemployment insurance program accessibility and suicide rates in the United States. *Prev. Medicine* **141**, 106318 (2020).

63. Roy, S. & Ghosh, P. Factors affecting covid-19 infected and death rates inform lockdown-related policymaking. *PloS one* **15**, e0241165 (2020).

64. Baum, F., Newman, L. & Biedrzycki, K. Vicious cycles: digital technologies and determinants of health in Australia. *Heal. promotion international* **29**, 349–360 (2012).

65. Raifman, J. *et al.* COVID-19 US state policy database. (2020).

66. Patel, N. & Erickson, S. Remote learning is here to stay - can we make it better? https://www.theverge.com/21570482/remote-learning-khan-academy-interview-decoder-podcast (Accessed 2021-05-30) (2020).

67. Herold, B. The disparities in remote learning under coronavirus (in charts). *Educ. Week* **10** (2020).

68. Parolin, Z. & Lee, E. K. Large socio-economic, geographic and demographic disparities exist in exposure to school closures. *Nat. human behaviour* **5**, 522–528 (2021).

69. Engzell, P., Frey, A. & Verhagen, M. D. Learning loss due to school closures during the COVID-19 pandemic. *Proc. Natl. Acad. Sci.* **118** (2021).

medRxiv preprint doi: https://doi.org/10.1101/2021.09.14.21263545; this version posted September 17, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

70. Baker, S. R., Farrokhnia, R. A., Meyer, S., Pagel, M. & Yannelis, C. How does household spending respond to an epidemic? Consumption during the 2020 COVID-19 pandemic. *The Rev. Asset Pricing Stud.* **10**, 834–862 (2020).

71. Chang, H.-H. & Meyerhoefer, C. D. Covid-19 and the demand for online food shopping services: Empirical evidence from Taiwan. *Am. J. Agric. Econ.* **103**, 448–465 (2021).

72. Saksena, M. J. *et al.* America's eating habits: food away from home. Tech. Rep., United States Department of Agriculture, Economic Research Service (2018).

73. Bublitz, M. G., Czarkowski, N., Hansen, J., Peracchio, L. A. & Tussler, S. Pandemic reveals vulnerabilities in food access: confronting hunger amidst a crisis. *J. Public Policy & Mark.* **40**, 105–107 (2021).

74. Ramsetty, A. & Adams, C. Impact of the digital divide in the age of COVID-19. *J. Am. Med. Informatics Assoc.* **27**, 1147–1148 (2020).

75. Stuart, E. A. Matching methods for causal inference: A review and a look forward. *Stat. science: a review journal Inst. Math. Stat.* **25**, 1 (2010).

76. Austin, P. C. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivar. behavioral research* **46**, 399–424 (2011).

77. Braveman, P. A. *et al.* Socioeconomic status in health research: one size does not fit all. *Jama* **294**, 2879–2888 (2005).

78. Williams, D. R., Mohammed, S. A., Leavell, J. & Collins, C. Race, socioeconomic status and health: Complexities, ongoing challenges and research opportunities. *Annals New York Acad. Sci.* **1186**, 69 (2010).

79. Williams, D. R. & Collins, C. Racial residential segregation: a fundamental cause of racial disparities in health. *Public health reports* (2016).

80. U.S. Census Bureau. American Community Survey 5-year estimates (2014-2018). https://censusreporter.org (Accessed 2020-06-08).

81. U.S. Census Bureau. ZIP Code Tabulation Areas. https://www.census.gov/geographies/reference-files/time-series/geo/gazetteer-files.2019.html (Accessed 2020-06-08).

82. Sanderson, M. & Dumais, S. Examining repetition in user search behavior. In *European Conference on Information Retrieval*, 597–604 (Springer, 2007).

83. Kaufman, J. S., Cooper, R. S. & McGee, D. L. Socioeconomic status and health in blacks and whites: the problem of residual confounding and the resiliency of race. *Epidemiology* 621–628 (1997).

84. Ho, D. E., Imai, K., King, G., Stuart, E. A. *et al.* MatchIt: nonparametric preprocessing for parametric causal inference. *J Stat Softw* **42**, 1–28 (2011).

85. Ratcliffe, M., Burd, C., Holder, K. & Fields, A. Defining rural at the US Census Bureau. *Am. community survey geography brief* **1** (2016).

86. Rubin, D. B. Using propensity scores to help design observational studies: application to the tobacco litigation. *Heal. Serv. Outcomes Res. Methodol.* **2**, 169–188 (2001).

87. Stuart, E. A., Lee, B. K. & Leacy, F. P. Prognostic score–based balance measures can be a useful diagnostic for propensity score methods in comparative effectiveness research. *J. clinical epidemiology* **66**, S84–S90 (2013).

medRxiv preprint doi: https://doi.org/10.1101/2021.09.14.21263545; this version posted September 17, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

88. U.S. Census Bureau. State Area Measurements and Internal Point Coordinates. https://www.census.gov/geographies/reference-files/2010/geo/state-area.html (Accessed 2020-06-08).

medRxiv preprint doi: https://doi.org/10.1101/2021.09.14.21263545; this version posted September 17, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Supplementary Information

Table S1. Means and medians of eight SDoH factors across 25,801 ZIP codes in our dataset. Population density is computed from the total U.S. population[80] divided by the total U.S. land area[88].

SDoH Factor	ZIP Code Mean	ZIP Code Median	U.S. Mean	ACS Table
Median Household Income	$60,819.849	$55,224	$65,712	B19013
% Unemployed	3.4%	3.0%	2.9%	B23025
% Black or African American Alone	8.3%	1.6%	12.8%	B02001
% Hispanic or Latino	10.1%	4.0%	18.4%	B03003
% with Bachelor or Higher Degree	25.9%	21.1%	33.1%	B15002
Population Density (per sq mile)	1573.966	108.421	92.830*	B01003
% with Health Insurance Coverage	91.1%	92.7%	98.4%	B27001
% with Internet Access	80.5%	81.8%	85.8%	B28003

Table S2. Split boundaries used to generate high-risk (treated) and low-risk (control) comparison groups based on ZIP code medians and U.S. national means of eight SDoH factors. Values that informed the choice of the split boundaries are bolded. For population density, we follow previous practices of urban-rural classification[85].

SDoH Factor	ZIP Code Median	U.S. Mean	Boundary	Treated	Control
Median Household Income	**$55,224**	$65,712	$55,224	$\leq$ $55,224	$>$ $55,224
% Unemployed	**3.0%**	2.9%	3.0%	$\leq$ 3.0%	$>$ 3.0%
% Black or African American Alone	1.6%	**12.8%**	12%	$\geq$ 12%	$<$ 12%
% Hispanic or Latino	4.0%	**18.4%**	18%	$\geq$ 18%	$<$ 18%
% with Bachelor or Higher Degree	**21.1%**	33.1%	21.1%	$\leq$ 21.1%	$>$ 21.1%
Population Density (per sq mile)	108.421	92.830	500	$\geq$ 500	$<$ 500
% with Health Insurance Coverage	**92.7%**	98.4%	92.7%	$\leq$ 92.7%	$>$ 92.7%
% with Internet Access	**81.8%**	85.8%	81.8%	$\leq$ 81.8%	$>$ 81.8%

medRxiv preprint doi: https://doi.org/10.1101/2021.09.14.21263545; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Table S3. Pearson correlation coefficients between SDoH factors across 25,801 ZIP codes in our dataset. These correlations reflect socioeconomic segregration and are controlled for through a matching-based approach (Methods). (* = attained Bachelor's degree or higher)

SDoH factor	% Unemp.	% Black	% Hispanic	% BA*	Pop. Density	% Insurance	% Internet
M. H. Income	-0.25	-0.23	-0.07	0.74	0.10	0.41	0.66
% Unemp.		0.32	0.15	-0.18	0.09	-0.23	-0.20
% Black			0.03	-0.09	0.15	-0.19	-0.23
% Hispanic				-0.07	0.21	-0.35	-0.04
% BA*					0.23	0.36	0.62
Pop. Density						0.00	0.10
% Insurance							0.41

Table S4. Seven search categories examined and their example search strings and regular expressions.

Search categories	Example search string	Example regular expression
Health condition related queries	arthritis; diabetes	\b(stroke|diabetes|covid|salmonella| autism|asthma|cancer)\b
Unemployment related queries	im unemployed; jobless benefits	\b((i m|im|i am) (unemployed|laid off|furloughed))\b
Clicks to state unemployment sites	click on www.michigan.gov/uia from search results	[\/.](www\.michigan\.gov\/uia| www\.in\.gov\/dwd\/2362\.htm)
Financial assistance related queries	relief fund; financial assistance	\b((relief|stimulus) (funds?|package|checks?))\b
Clicks to online learning sites	click on coursera.org from search results	[\/.](quizlet\.com|mooc-list\.com|coursera\.org|khanacademy\.org)
Food delivery related queries	grocery delivery; deliver meal	\b((food|grocery|meal) delivery)\b
Food assistance related queries	food stamps; snap program	b(food assistance|snap program|food stamps|food bank|food pantry)\b

Table S5. Summary of *pre-matching* balance assessment and sample sizes across SDoH factors. $|SMD| > 0.25$ indicates that the two ZIP code groups created along the split boundary are different and, thus, not comparable.

SDoH Factor	Split Boundary	Treated Count	Control Count	Max SMD	Mean SMD
Median Household Income	$55,224	12637	12637	1.018	0.459
% Unemployed	3.0%	12637	12637	0.385	0.269
% Black or Afr. Am. Alone	12%	4945	20329	0.467	0.300
% Hispanic or Latino	18%	4028	21246	0.591	0.218
% with Bachelor or Higher	21.1%	12637	12637	0.975	0.416
Population Density (per mi^2)	500	8037	17237	0.689	0.392
% with Health Ins. Cov.	92.7%	12666	12608	0.629	0.396
% with Internet Access	81.8%	12637	12637	0.974	0.429

medRxiv preprint doi: https://doi.org/10.1101/2021.09.14.21263545; this version posted September 17, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Table S6. Summary of *post-matching* balance assessment and sample sizes across SDoH factors.

SDoH Factor	Max SMD	Mean SMD	Tr. Matched	Tr. Unmat.	Ctrl. Matched	Ctrl. Unmat.
Median Household Income	0.236	0.105	12555	21	3854	8720
% Unemployed	0.088	0.026	12544	31	5917	6658
% Black or Afr. Am. Alone	0.144	0.072	4932	0	1603	18615
% Hispanic or Latino	0.131	0.082	4018	0	2247	18885
% with Bachelor or Higher	0.113	0.069	12572	3	3981	8594
Population Density (per mi^2)	0.181	0.114	8022	6	2836	14286
% with Health Ins. Cov.	0.067	0.030	8508	0	4339	12303
% with Internet Access	0.176	0.075	12575	0	3623	8952

medRxiv preprint doi: https://doi.org/10.1101/2021.09.14.21263545; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Table S7. Balance assessment between unmatched high (Treated) and low (Control) '% Black or African American Alone' Groups

SDoH Factor	Means Treated	Means Control	Mean Diff.	Std. Mean Diff.
distance	0.590	0.100	0.490	1.606
Median Household Income	49810.138	63497.945	-13687.808	-0.680
Unemployed	0.045	0.031	0.015	0.628
Hispanic or Latino	0.125	0.095	0.030	0.201
With Bachelor or Higher Degree Attained	0.232	0.265	-0.034	-0.244
Population Density (per sq mile)	2975.905	1172.126	1803.778	0.245
With Health Insurance Coverage	0.883	0.918	-0.034	-0.575
With Internet Access	0.766	0.816	-0.050	-0.414
Black or African American Alone	0.337	0.021	0.316	1.517

Table S8. Balance assessment between matched high (Treated) and low (Control) '% Black or African American Alone' groups

SDoH Factor	Means Treated	Means Control	Mean Diff.	Std. Mean Diff.
distance	0.887	0.791	0.096	0.043
Median Household Income	49835.515	52509.535	-2674.020	-0.133
Unemployed	0.045	0.043	0.002	0.097
Hispanic or Latino	0.125	0.126	-0.001	-0.004
With Bachelor or Higher Degree Attained	0.232	0.240	-0.008	-0.059
Population Density (per sq mile)	2981.898	2250.314	731.584	0.099
With Health Insurance Coverage	0.884	0.881	0.003	0.044
With Internet Access	0.766	0.775	-0.009	-0.072
Black or African American Alone	0.337	0.041	0.297	1.425

medRxiv preprint doi: https://doi.org/10.1101/2021.09.14.21263545; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Table S9. Balance assessment between unmatched high (Treated) and low (Control) '% Hispanic or Latino' Groups

SDoH Factor	Means Treated	Means Control	Mean Diff.	Std. Mean Diff.
distance	0.368	0.120	0.249	1.023
Median Household Income	57217.731	61502.769	-4285.038	-0.209
Unemployed	0.041	0.032	0.009	0.398
Black or African American Alone	0.099	0.080	0.019	0.145
With Bachelor or Higher Degree Attained	0.234	0.264	-0.030	-0.224
Population Density (per sq mile)	4021.757	1051.698	2970.060	0.333
With Health Insurance Coverage	0.864	0.920	-0.056	-0.751
With Internet Access	0.806	0.806	-0.000	-0.001
Hispanic or Latino	0.399	0.044	0.355	1.736

Table S10. Balance assessment between matched high (Treated) and low (Control) '% Hispanic or Latino' groups

SDoH Factor	Means Treated	Means Control	Mean Diff.	Std. Mean Diff.
distance	-0.639	-0.699	0.060	0.042
Median Household Income	57234.061	58095.893	-861.832	-0.042
Unemployed	0.041	0.038	0.002	0.105
Black or African American Alone	0.099	0.085	0.014	0.111
With Bachelor or Higher Degree Attained	0.234	0.239	-0.006	-0.042
Population Density (per sq mile)	4029.507	2986.171	1043.336	0.117
With Health Insurance Coverage	0.864	0.865	-0.001	-0.018
With Internet Access	0.806	0.817	-0.010	-0.096
Hispanic or Latino	0.399	0.086	0.313	1.530

medRxiv preprint doi: https://doi.org/10.1101/2021.09.14.21263545; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Table S11. Balance assessment between unmatched high (Control) and low (Treated) 'Median Household Income' Groups

SDoH Factor	Means Treated	Means Control	Mean Diff.	Std. Mean Diff.
distance	0.725	0.275	0.450	1.855
Unemployed	0.039	0.028	0.011	0.412
Black or African American Alone	0.117	0.049	0.068	0.343
Hispanic or Latino	0.110	0.091	0.019	0.102
With Bachelor or Higher Degree Attained	0.177	0.341	-0.164	-1.751
Population Density (per sq mile)	1266.027	1784.065	-518.038	-0.107
With Health Insurance Coverage	0.887	0.934	-0.047	-0.692
With Internet Access	0.744	0.868	-0.125	-1.257
Median Household Income	43356.749	78282.949	-34926.200	-4.189

Table S12. Balance assessment between matched high (Control) and low (Treated) 'Median Household Income' groups

SDoH Factor	Means Treated	Means Control	Mean Diff.	Std. Mean Diff.
distance	1.567	1.449	0.119	0.060
Unemployed	0.039	0.036	0.003	0.113
Black or African American Alone	0.117	0.092	0.025	0.126
Hispanic or Latino	0.111	0.091	0.020	0.107
With Bachelor or Higher Degree Attained	0.177	0.180	-0.003	-0.029
Population Density (per sq mile)	1272.878	747.235	525.643	0.108
With Health Insurance Coverage	0.888	0.889	-0.001	-0.021
With Internet Access	0.745	0.746	-0.001	-0.011
Median Household Income	43401.084	62615.121	-19214.037	-2.306

medRxiv preprint doi: https://doi.org/10.1101/2021.09.14.21263545; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Table S13. Balance assessment between unmatched high (Control) and low (Treated) '% with Bachelor or Higher Degree Attained' Groups

SDoH Factor	Means Treated	Means Control	Mean Diff.	Std. Mean Diff.
distance	0.714	0.286	0.429	1.869
Median Household Income	47879.339	73760.359	-25881.020	-2.046
Unemployed	0.037	0.029	0.008	0.308
Black or African American Alone	0.095	0.072	0.023	0.125
Hispanic or Latino	0.109	0.093	0.016	0.086
Population Density (per sq mile)	788.145	2261.947	-1473.803	-0.441
With Health Insurance Coverage	0.890	0.931	-0.041	-0.571
With Internet Access	0.745	0.867	-0.121	-1.222
With Bachelor or Higher Degree Attained	0.141	0.376	-0.235	-5.334

Table S14. Balance assessment between matched high (Control) and low (Treated) '% with Bachelor or Higher Degree Attained' groups

SDoH Factor	Means Treated	Means Control	Mean Diff.	Std. Mean Diff.
distance	1.336	1.242	0.094	0.058
Median Household Income	47895.952	48454.763	-558.811	-0.044
Unemployed	0.037	0.036	0.002	0.073
Black or African American Alone	0.095	0.084	0.011	0.060
Hispanic or Latino	0.109	0.097	0.013	0.066
Population Density (per sq mile)	793.735	751.143	42.592	0.013
With Health Insurance Coverage	0.891	0.897	-0.007	-0.097
With Internet Access	0.746	0.751	-0.005	-0.050
With Bachelor or Higher Degree Attained	0.142	0.269	-0.128	-2.903

medRxiv preprint doi: https://doi.org/10.1101/2021.09.14.21263545; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Table S15. Balance assessment between unmatched high (Treated) and low (Control) 'Population Density (per sq mile)' Groups

SDoH Factor	Means Treated	Means Control	Mean Diff.	Std. Mean Diff.
distance	0.645	0.166	0.479	1.701
Median Household Income	71129.001	56013.058	15115.943	0.462
Unemployed	0.038	0.031	0.006	0.322
Black or African American Alone	0.144	0.055	0.089	0.450
Hispanic or Latino	0.169	0.069	0.100	0.522
With Bachelor or Higher Degree Attained	0.364	0.210	0.154	0.805
With Health Insurance Coverage	0.915	0.909	0.006	0.097
With Internet Access	0.860	0.781	0.079	0.877
Population Density (per sq mile)	4599.740	91.425	4508.315	0.510

Table S16. Balance assessment between matched high (Treated) and low (Control) 'Population Density (per sq mile)' groups

SDoH Factor	Means Treated	Means Control	Mean Diff.	Std. Mean Diff.
distance	0.882	0.763	0.118	0.067
Median Household Income	71169.714	72899.376	-1729.662	-0.053
Unemployed	0.038	0.035	0.002	0.132
Black or African American Alone	0.144	0.124	0.020	0.099
Hispanic or Latino	0.169	0.154	0.015	0.079
With Bachelor or Higher Degree Attained	0.363	0.371	-0.007	-0.038
With Health Insurance Coverage	0.915	0.925	-0.010	-0.156
With Internet Access	0.860	0.873	-0.013	-0.147
Population Density (per sq mile)	4598.958	227.548	4371.410	0.494

medRxiv preprint doi: https://doi.org/10.1101/2021.09.14.21263545; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Table S17. Balance assessment between unmatched high (Treated) and low (Control) '% Unemployed' Groups

SDoH Factor	Means Treated	Means Control	Mean Diff.	Std. Mean Diff.
distance	0.581	0.419	0.162	0.781
Median Household Income	55178.441	66461.257	-11282.816	-0.503
Black or African American Alone	0.125	0.042	0.083	0.421
Hispanic or Latino	0.132	0.070	0.062	0.328
With Bachelor or Higher Degree Attained	0.230	0.288	-0.058	-0.416
Population Density (per sq mile)	2062.020	988.072	1073.948	0.173
With Health Insurance Coverage	0.898	0.924	-0.026	-0.404
With Internet Access	0.791	0.821	-0.029	-0.272
Unemployed	0.049	0.018	0.030	1.401

Table S18. Balance assessment between matched high (Treated) and low (Control) '% Unemployed' groups

SDoH Factor	Means Treated	Means Control	Mean Diff.	Std. Mean Diff.
distance	0.460	0.423	0.037	0.032
Median Household Income	55192.003	55360.831	-168.828	-0.008
Black or African American Alone	0.124	0.122	0.002	0.010
Hispanic or Latino	0.131	0.125	0.006	0.030
With Bachelor or Higher Degree Attained	0.230	0.225	0.004	0.030
Population Density (per sq mile)	1902.049	1476.827	425.222	0.069
With Health Insurance Coverage	0.898	0.898	-0.001	-0.010
With Internet Access	0.791	0.791	0.001	0.006
Unemployed	0.048	0.021	0.028	1.303

medRxiv preprint doi: https://doi.org/10.1101/2021.09.14.21263545; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Table S19. Balance assessment between unmatched high (Control) and low (Treated) '% with Health Insurance Coverage' Groups

SDoH Factor	Means Treated	Means Control	Mean Diff.	Std. Mean Diff.
distance	0.528	0.242	0.286	1.140
Median Household Income	48183.387	67291.556	-19108.168	-1.336
Unemployed	0.040	0.030	0.010	0.374
Black or African American Alone	0.133	0.058	0.075	0.378
Hispanic or Latino	0.167	0.067	0.101	0.458
With Bachelor or Higher Degree Attained	0.187	0.296	-0.109	-1.076
Population Density (per sq mile)	1529.525	1522.752	6.774	0.001
With Internet Access	0.752	0.834	-0.082	-0.737
With Health Insurance Coverage	0.841	0.947	-0.106	-1.770

Table S20. Balance assessment between matched high (Control) and low (Treated) '% with Health Insurance Coverage' groups

SDoH Factor	Means Treated	Means Control	Mean Diff.	Std. Mean Diff.
distance	0.255	0.199	0.055	0.038
Median Household Income	48177.005	48170.216	6.789	0.000
Unemployed	0.040	0.039	0.001	0.025
Black or African American Alone	0.133	0.126	0.007	0.035
Hispanic or Latino	0.168	0.158	0.009	0.043
With Bachelor or Higher Degree Attained	0.187	0.187	0.000	0.005
Population Density (per sq mile)	1537.624	1250.643	286.982	0.060
With Internet Access	0.752	0.753	-0.001	-0.008
With Health Insurance Coverage	0.841	0.932	-0.092	-1.530

medRxiv preprint doi: https://doi.org/10.1101/2021.09.14.21263545; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Table S21. Balance assessment between unmatched high (Control) and low (Treated) '% with Internet Access' Groups

SDoH Factor	Means Treated	Means Control	Mean Diff.	Std. Mean Diff.
distance	0.749	0.251	0.497	2.260
Median Household Income	46716.098	74923.599	-28207.501	-2.274
Unemployed	0.037	0.030	0.007	0.256
Black or African American Alone	0.107	0.060	0.047	0.244
Hispanic or Latino	0.102	0.099	0.003	0.019
With Bachelor or Higher Degree Attained	0.168	0.350	-0.182	-2.326
Population Density (per sq mile)	1003.589	2046.502	-1042.913	-0.225
With Health Insurance Coverage	0.889	0.932	-0.043	-0.603
With Internet Access	0.723	0.889	-0.166	-1.980

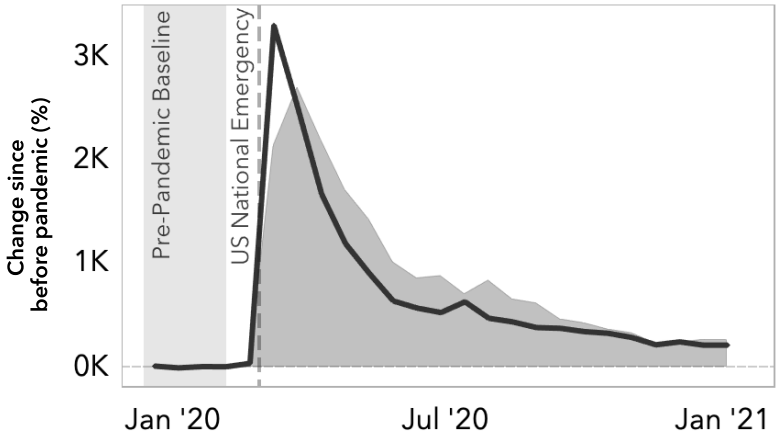
Table S22. Balance assessment between matched high (Control) and low (Treated) '% with Internet Access' groups

SDoH Factor	Means Treated	Means Control	Mean Diff.	Std. Mean Diff.
distance	1.436	1.234	0.202	0.136
Median Household Income	46721.584	48710.356	-1988.773	-0.160
Unemployed	0.037	0.035	0.001	0.056
Black or African American Alone	0.107	0.096	0.011	0.059
Hispanic or Latino	0.103	0.093	0.009	0.050
With Bachelor or Higher Degree Attained	0.168	0.173	-0.005	-0.066
Population Density (per sq mile)	1009.178	898.001	111.178	0.024
With Health Insurance Coverage	0.889	0.893	-0.003	-0.049
With Internet Access	0.724	0.854	-0.130	-1.559

medRxiv preprint doi: https://doi.org/10.1101/2021.09.14.21263545; this version posted September 17, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

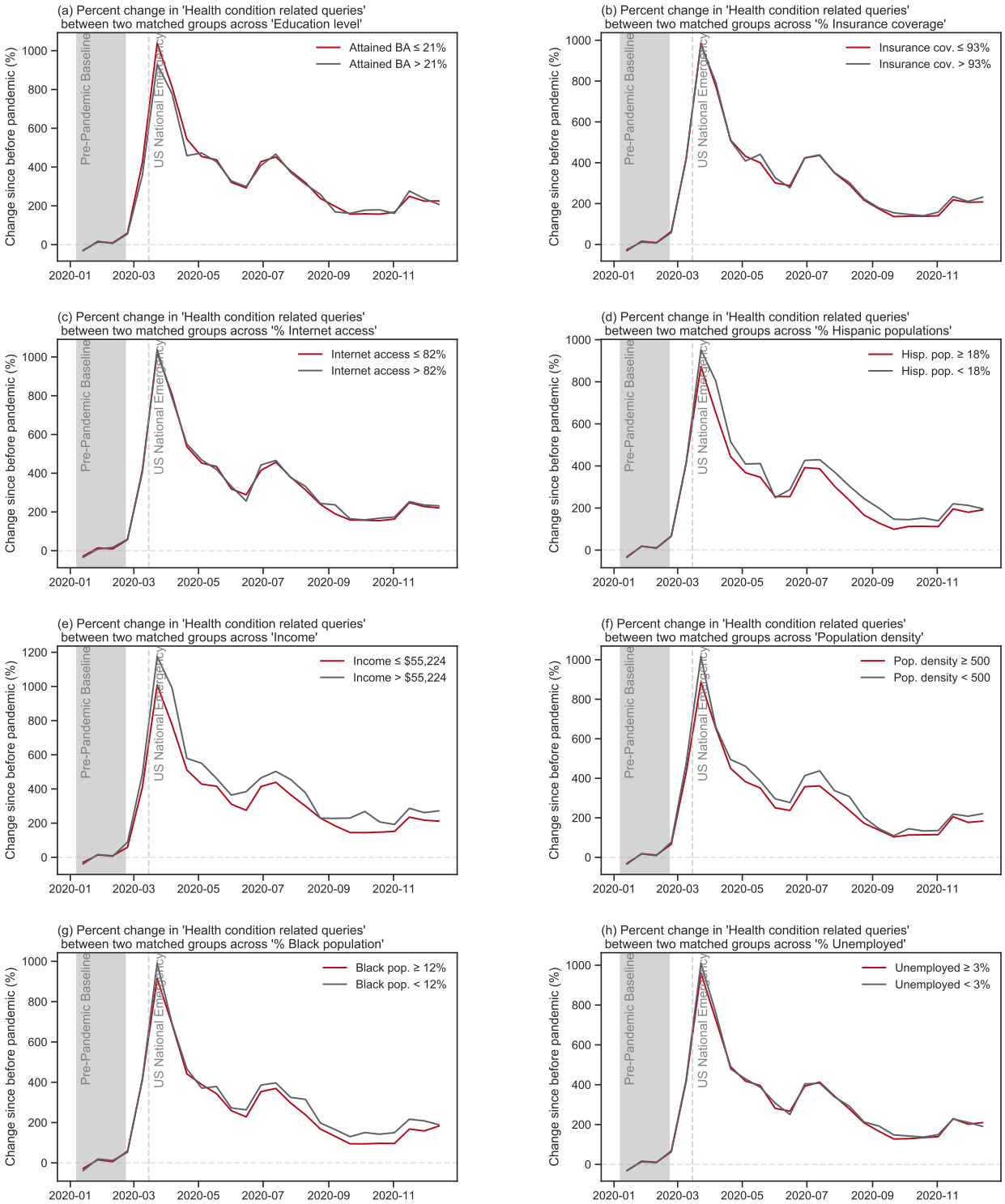
Figure S1. Percent change in the unemployment related queries in Bing (shaded) and the reported unemployment claims from the US Department of Labor (line, https://oui.doleta.gov/unemploy/claims.asp) compared to pre-pandemic baseline.



medRxiv preprint doi: https://doi.org/10.1101/2021.09.14.21263545; this version posted September 17, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

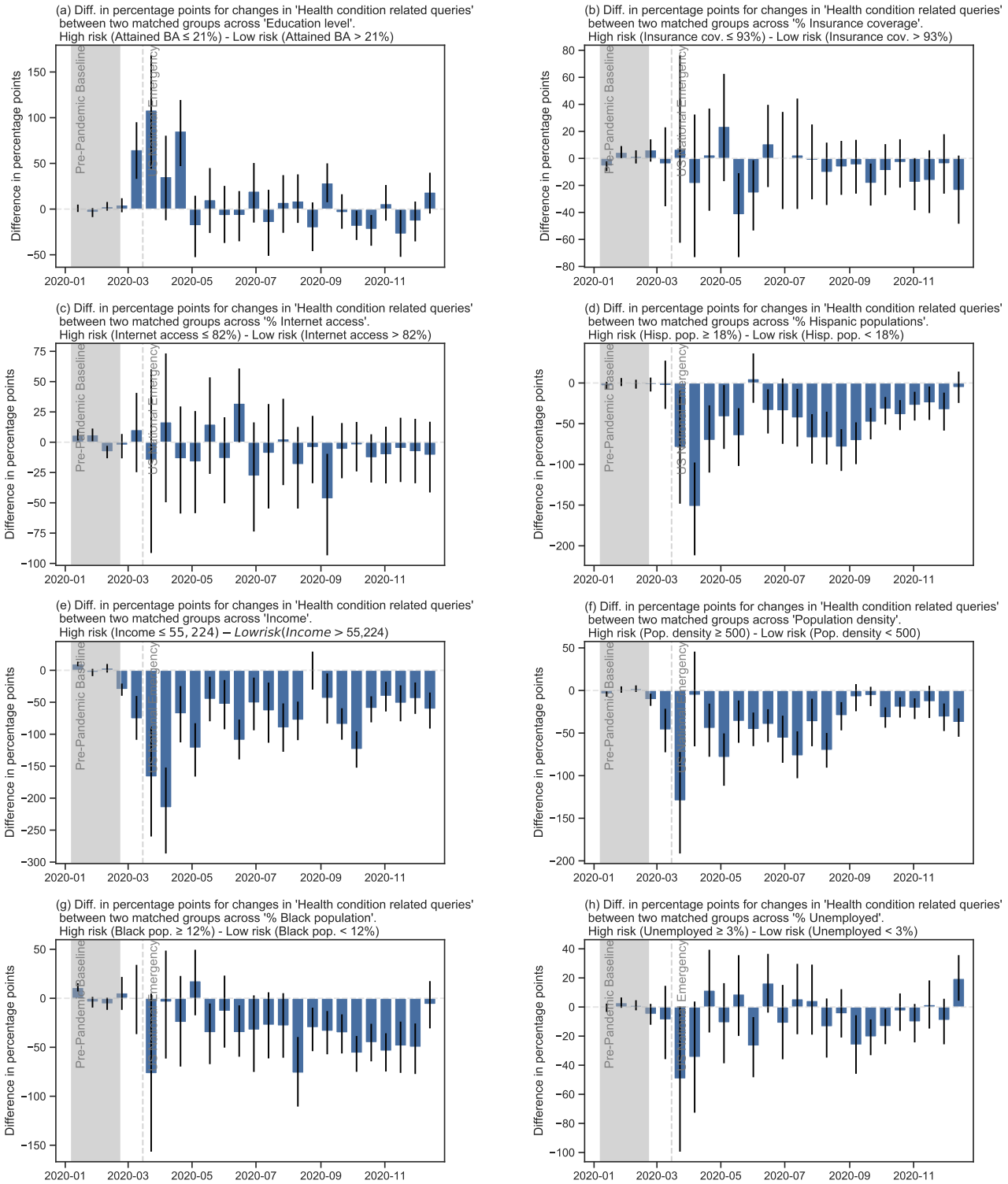
Figure S2. Percent change in 'Health condition related queries' between two matched groups across eight SDoH factors.



medRxiv preprint doi: https://doi.org/10.1101/2021.09.14.21263545; this version posted September 17, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

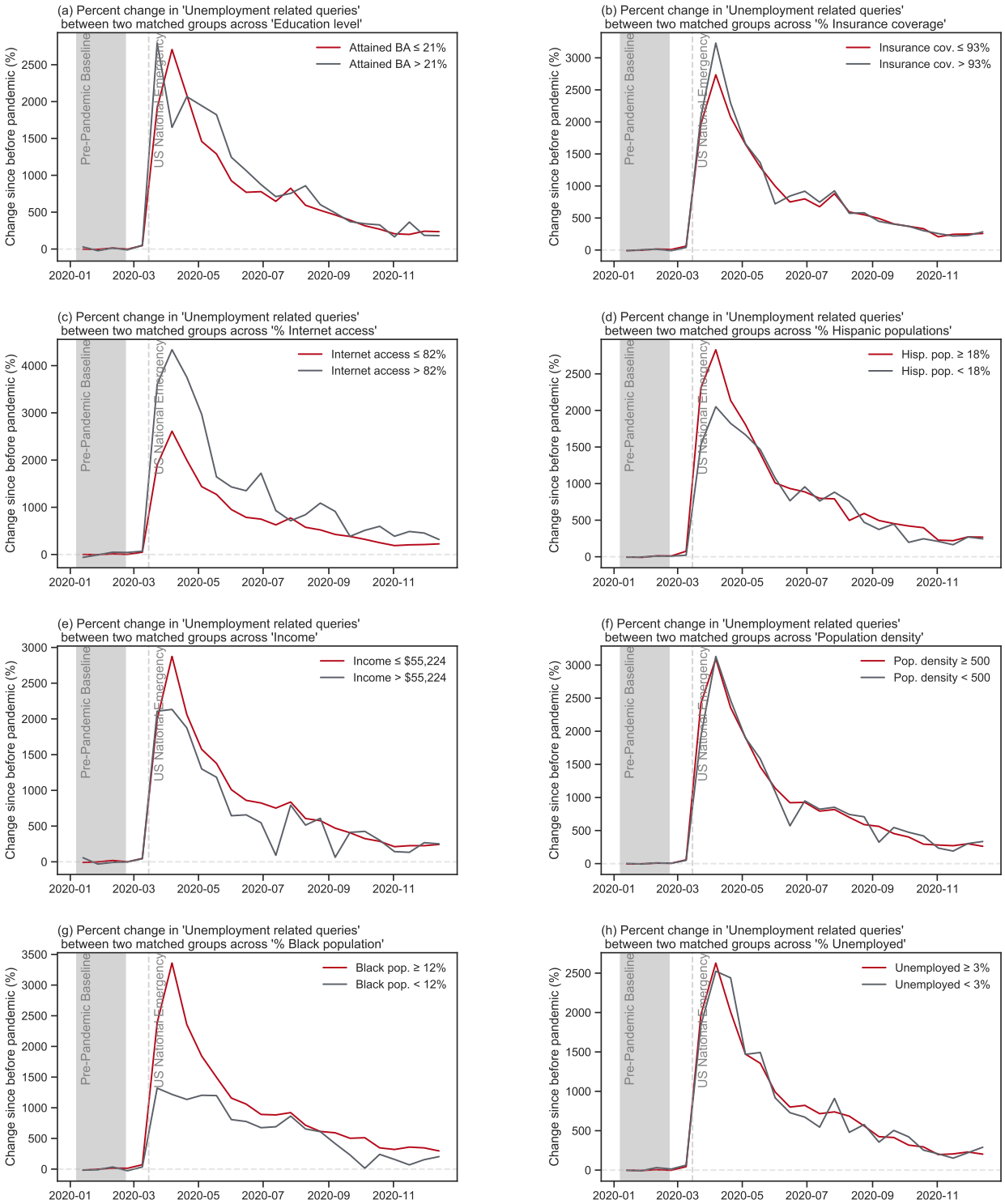
Figure S3. Differences in percentage points for changes in 'Health condition related queries' between two matched groups across eight SDoH factors.



medRxiv preprint doi: https://doi.org/10.1101/2021.09.14.21263545; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

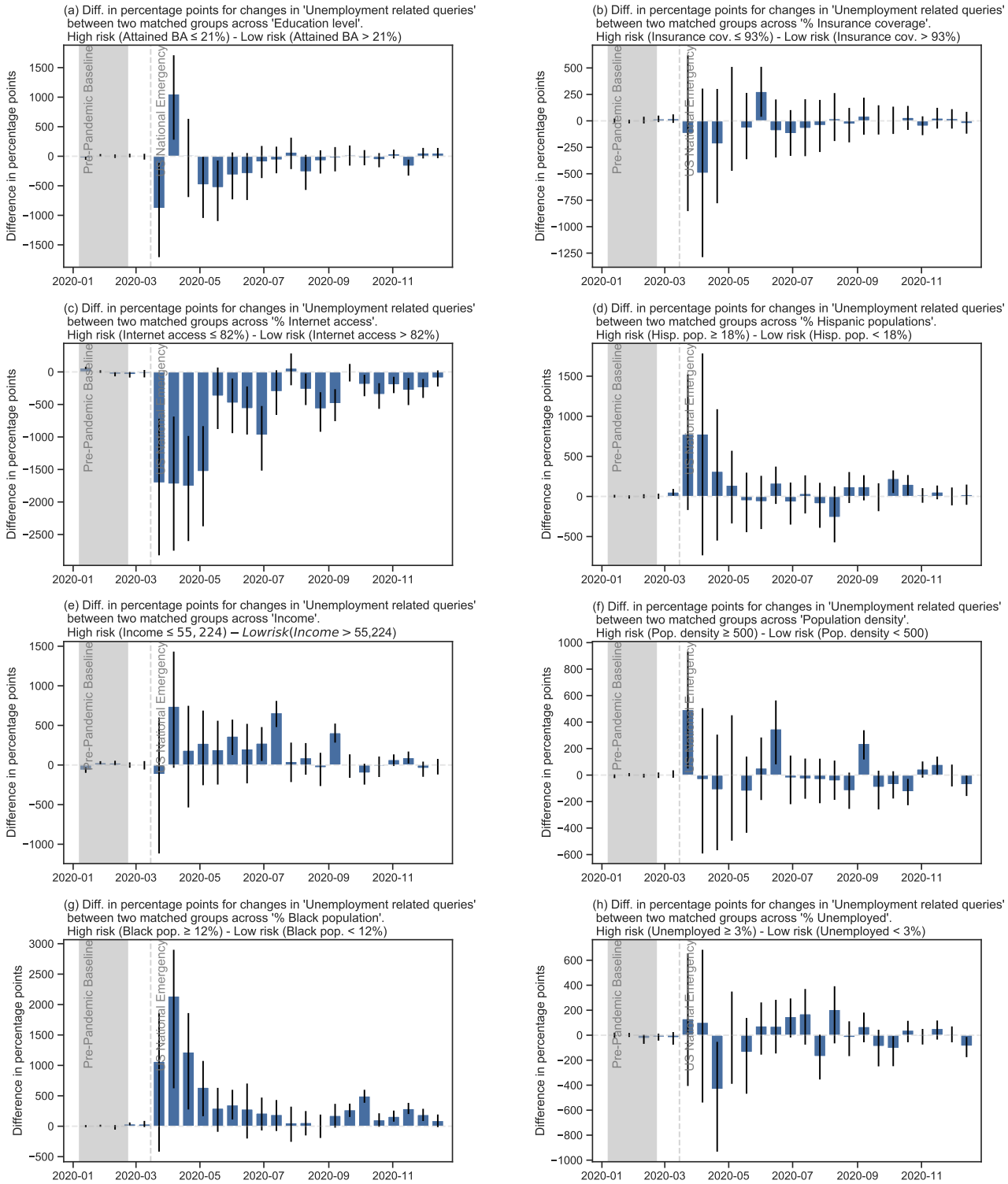
Figure S4. Percent change in 'Unemployment related queries' between two matched groups across eight SDoH factors.



medRxiv preprint doi: https://doi.org/10.1101/2021.09.14.21263545; this version posted September 17, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

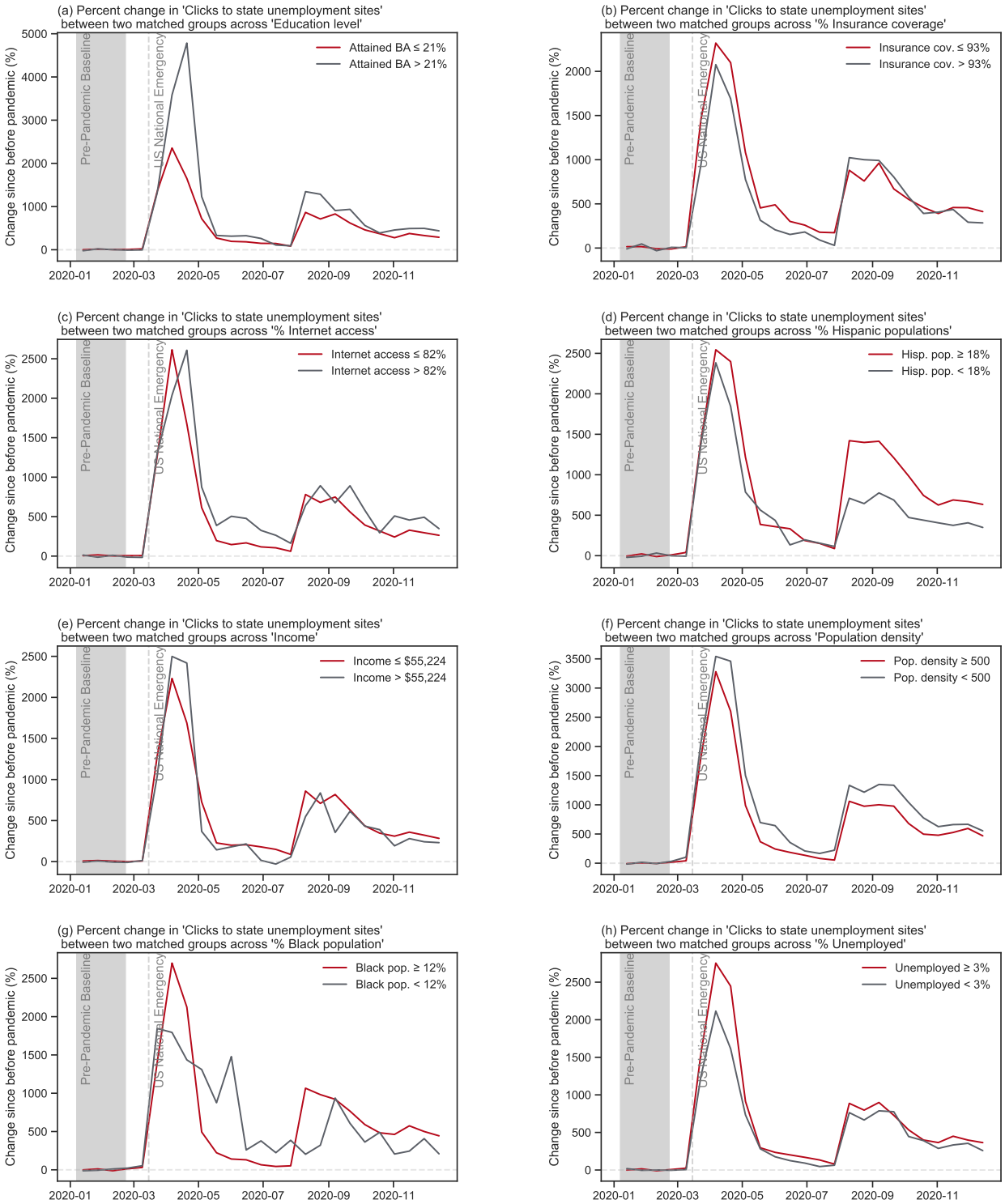
Figure S5. Differences in percentage points for changes in 'Unemployment related queries' between two matched groups across eight SDoH factors.



medRxiv preprint doi: https://doi.org/10.1101/2021.09.14.21263545; this version posted September 17, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

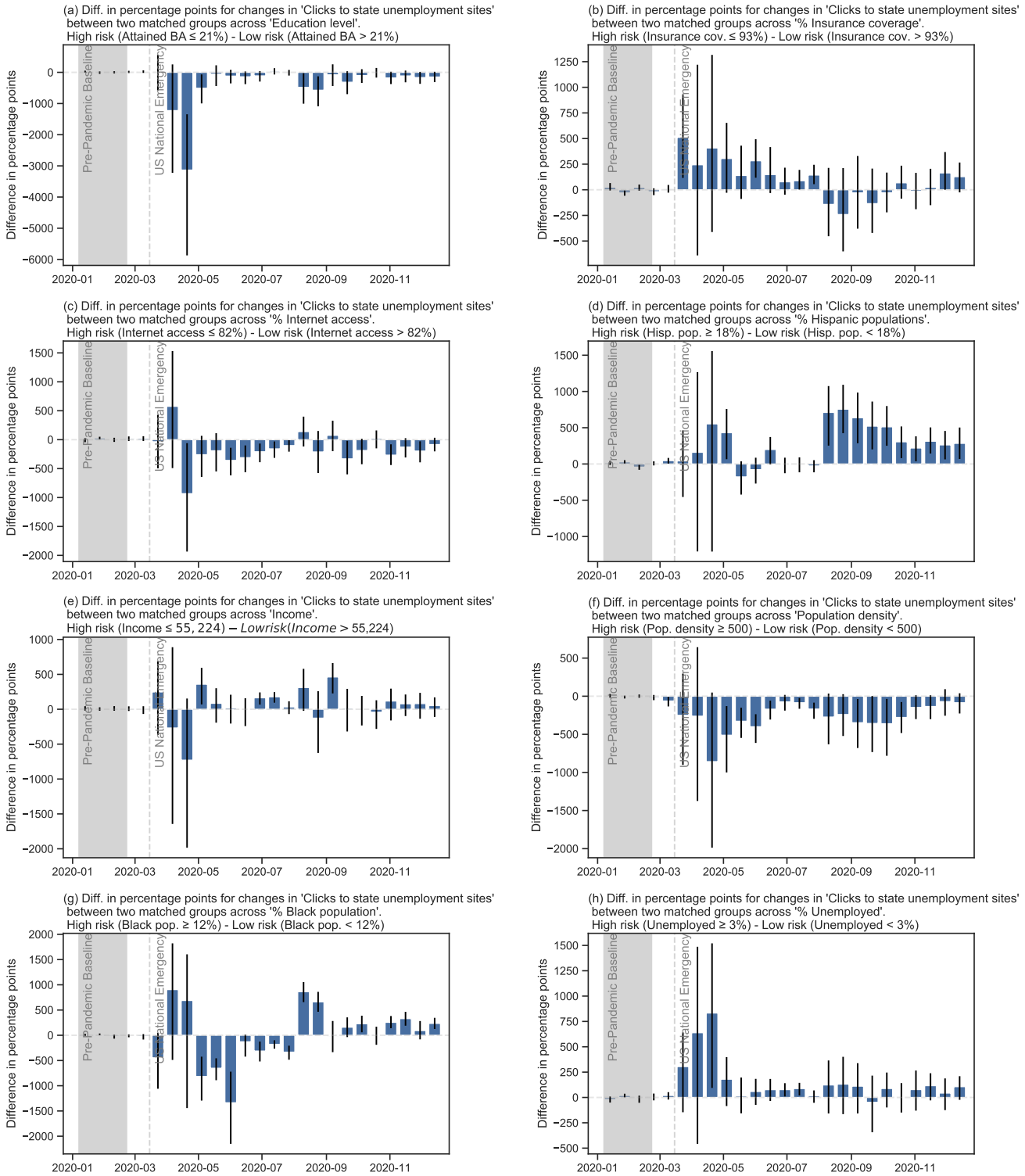
Figure S6. Percent change in 'Clicks to state unemployment sites' between two matched groups across eight SDoH factors.



medRxiv preprint doi: https://doi.org/10.1101/2021.09.14.21263545; this version posted September 17, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

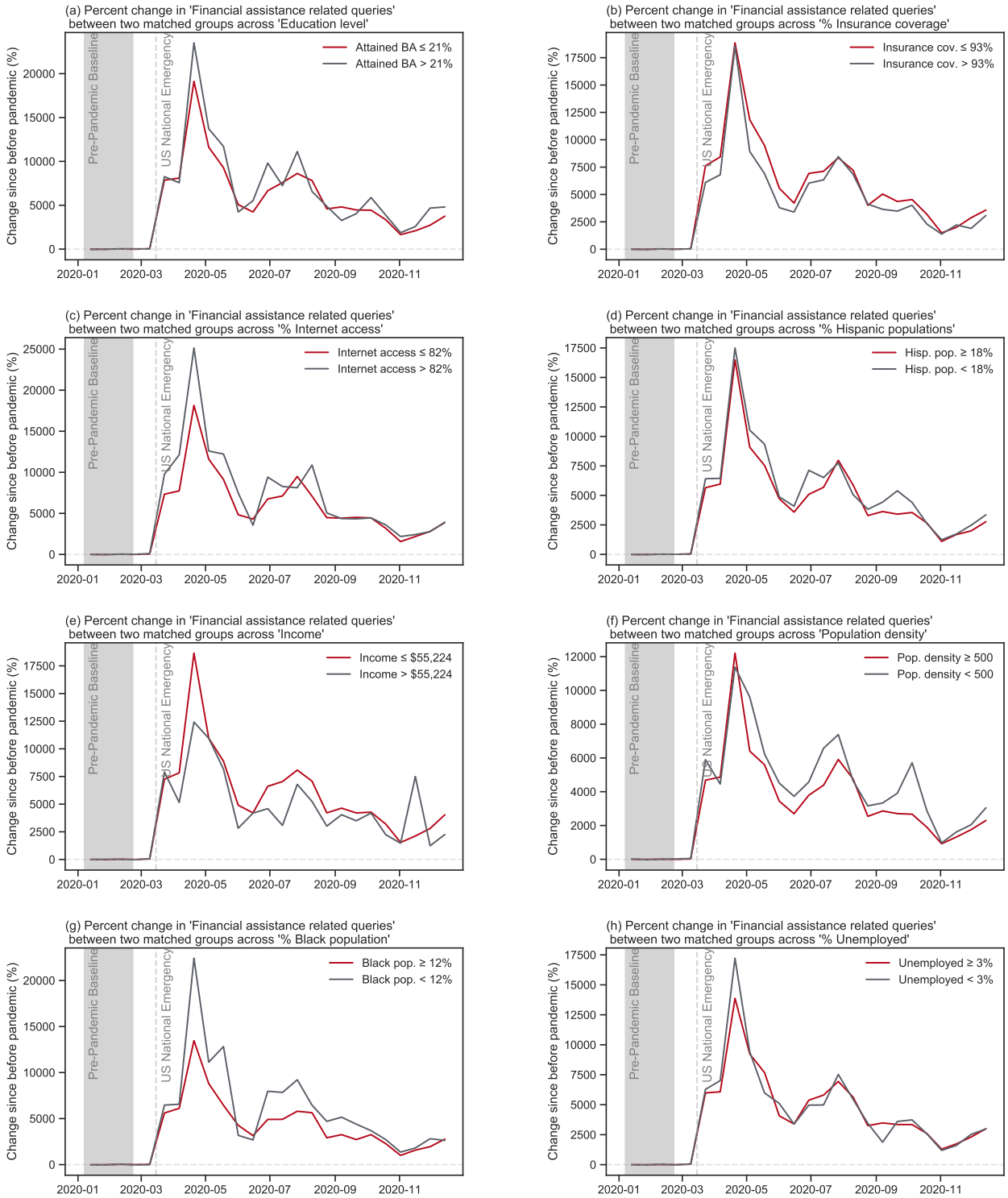
Figure S7. Differences in percentage points for changes in 'Clicks to state unemployment sites' between two matched groups across eight SDoH factors.



medRxiv preprint doi: https://doi.org/10.1101/2021.09.14.21263545; this version posted September 17, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

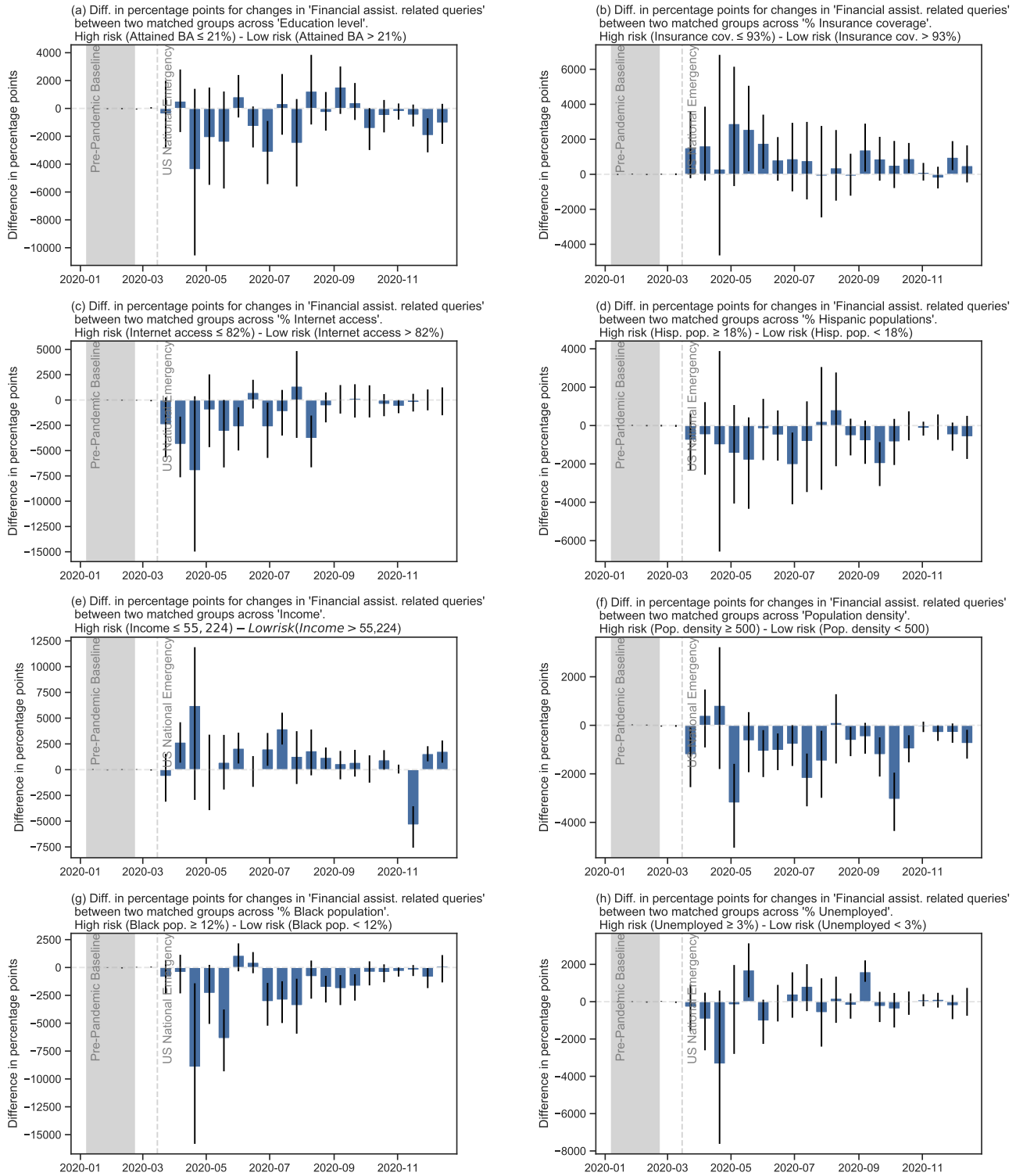
Figure S8. Percent change in 'Financial assistance related queries' between two matched groups across eight SDoH factors.



medRxiv preprint doi: https://doi.org/10.1101/2021.09.14.21263545; this version posted September 17, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

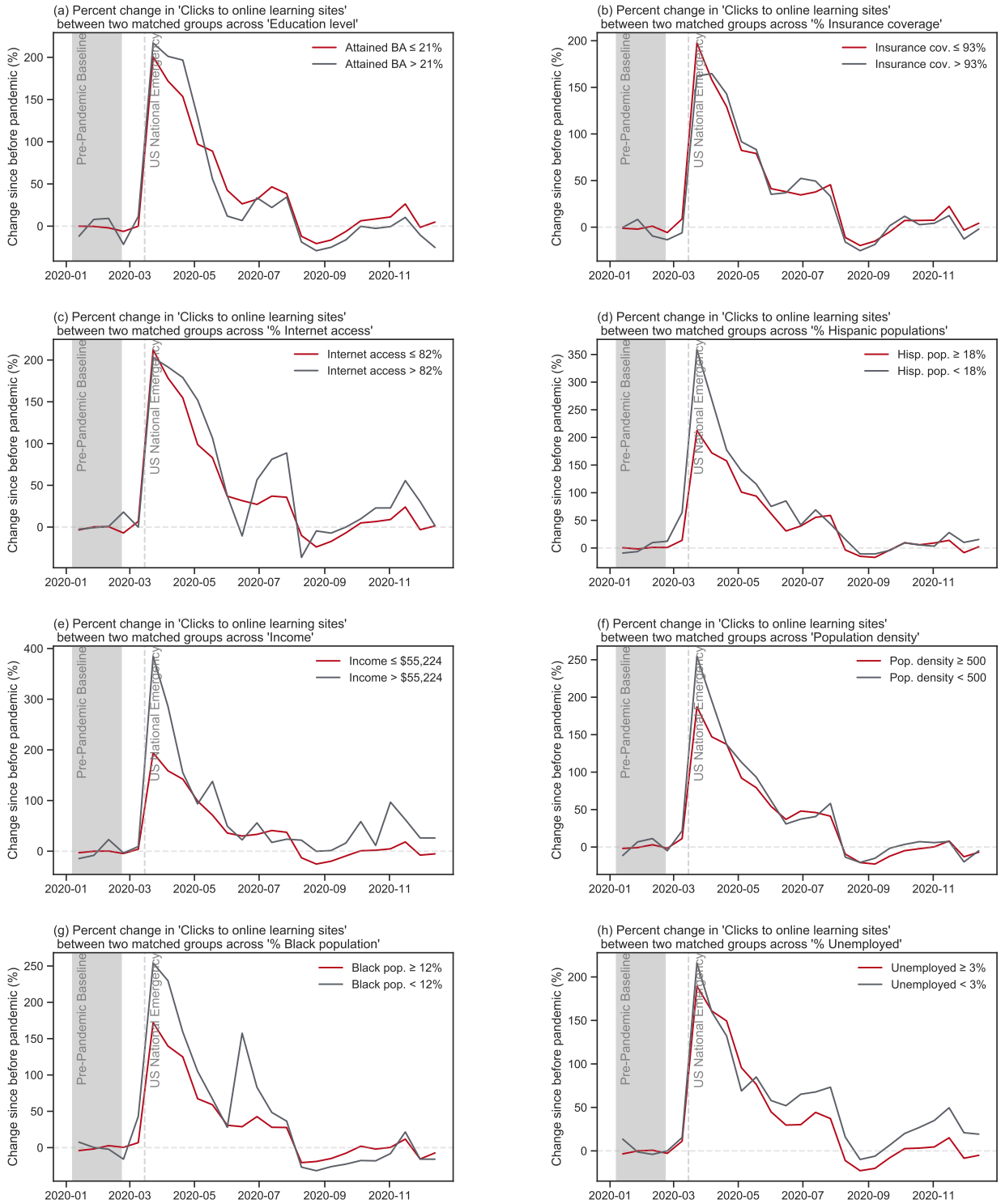
Figure S9. Differences in percentage points for changes in 'Financial assistance related queries' between two matched groups across eight SDoH factors.



medRxiv preprint doi: https://doi.org/10.1101/2021.09.14.21263545; this version posted September 17, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

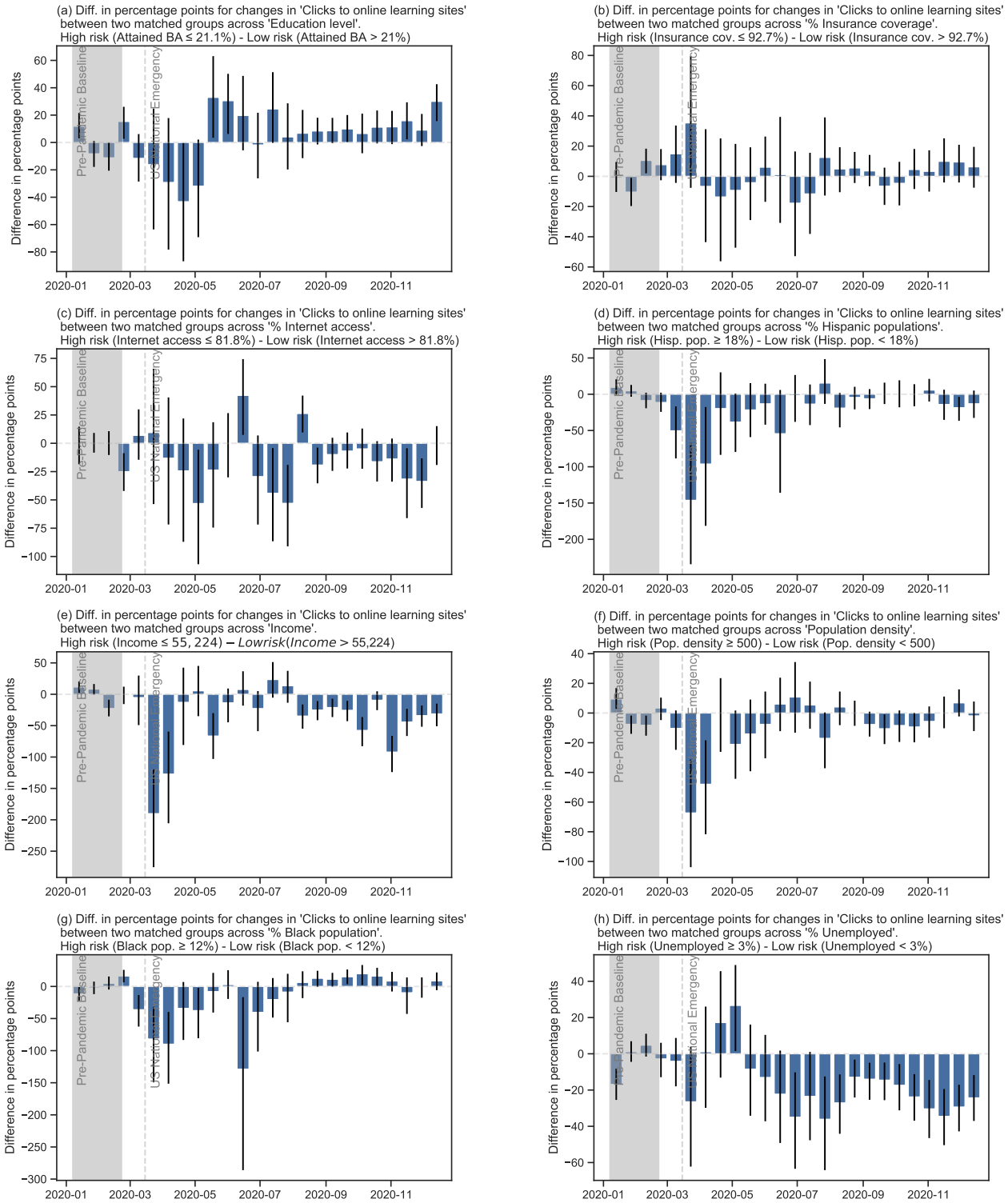
Figure S10. Percent change in 'Click to online learning sites' between two matched groups across eight SDoH factors.



medRxiv preprint doi: https://doi.org/10.1101/2021.09.14.21263545; this version posted September 17, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

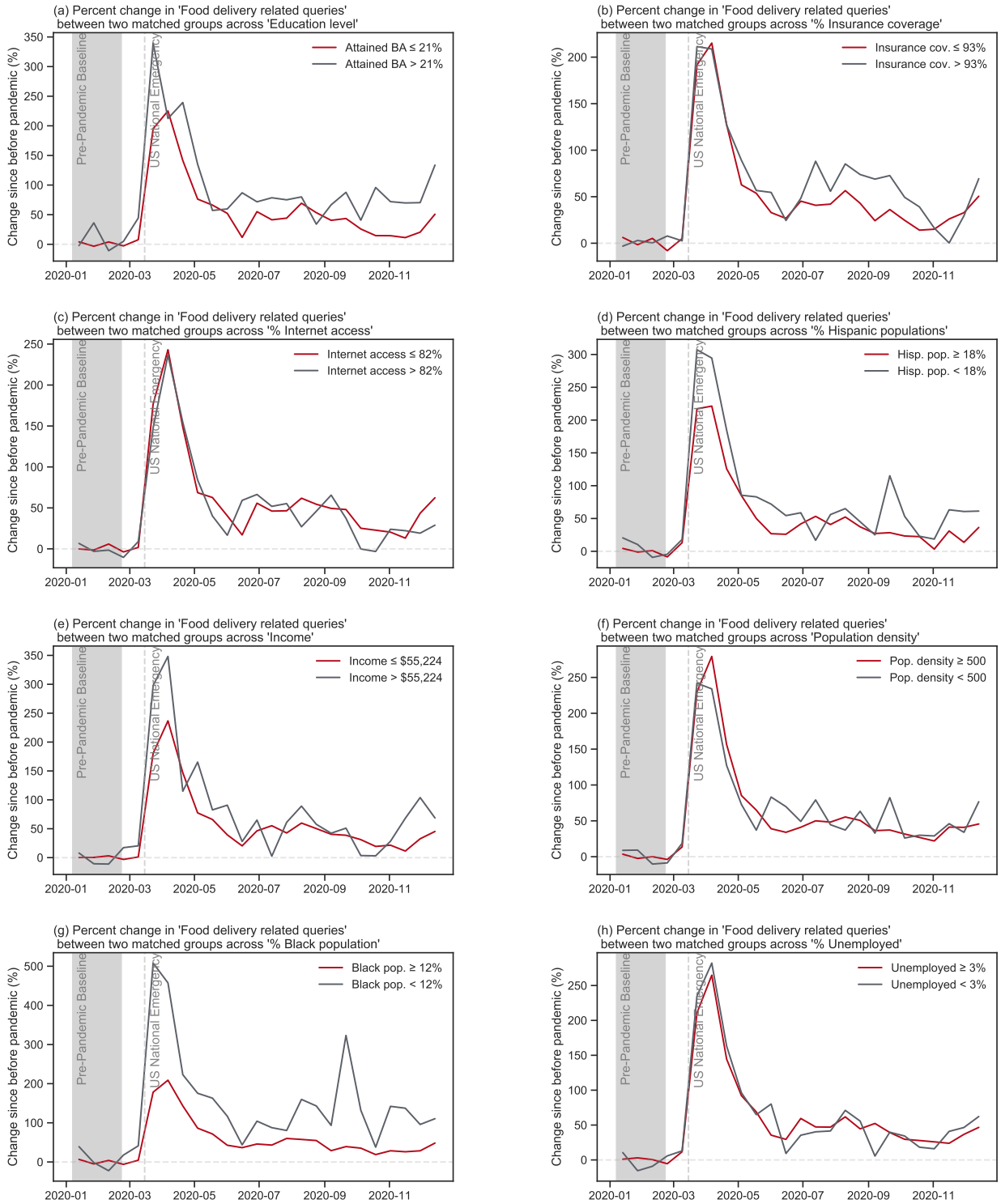
Figure S11. Differences in percentage points for changes in 'Click to online learning sites' between two matched groups across eight SDoH factors.



medRxiv preprint doi: https://doi.org/10.1101/2021.09.14.21263545; this version posted September 17, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

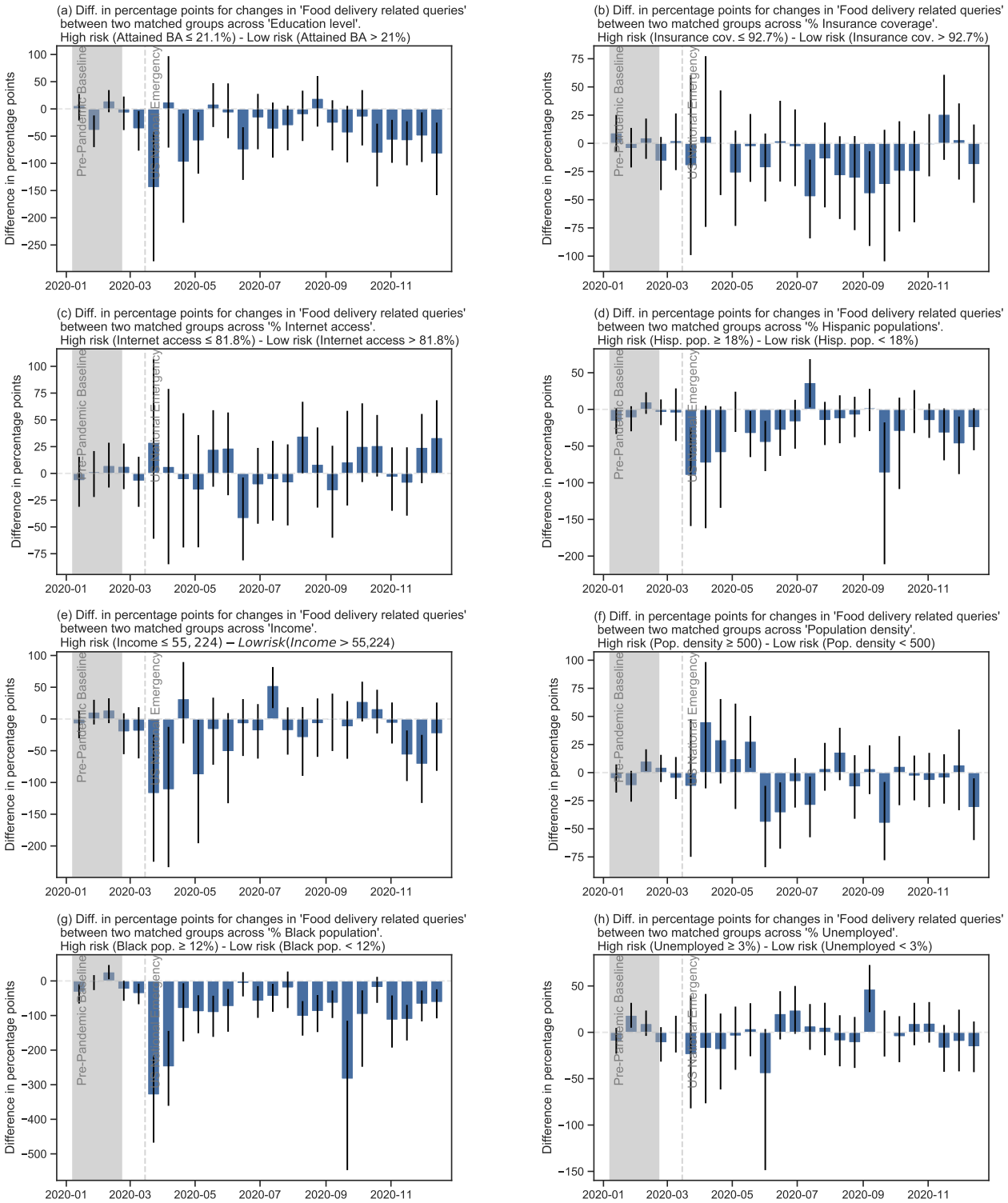
Figure S12. Percent change in 'Food delivery related queries' between two matched groups across eight SDoH factors.



medRxiv preprint doi: https://doi.org/10.1101/2021.09.14.21263545; this version posted September 17, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

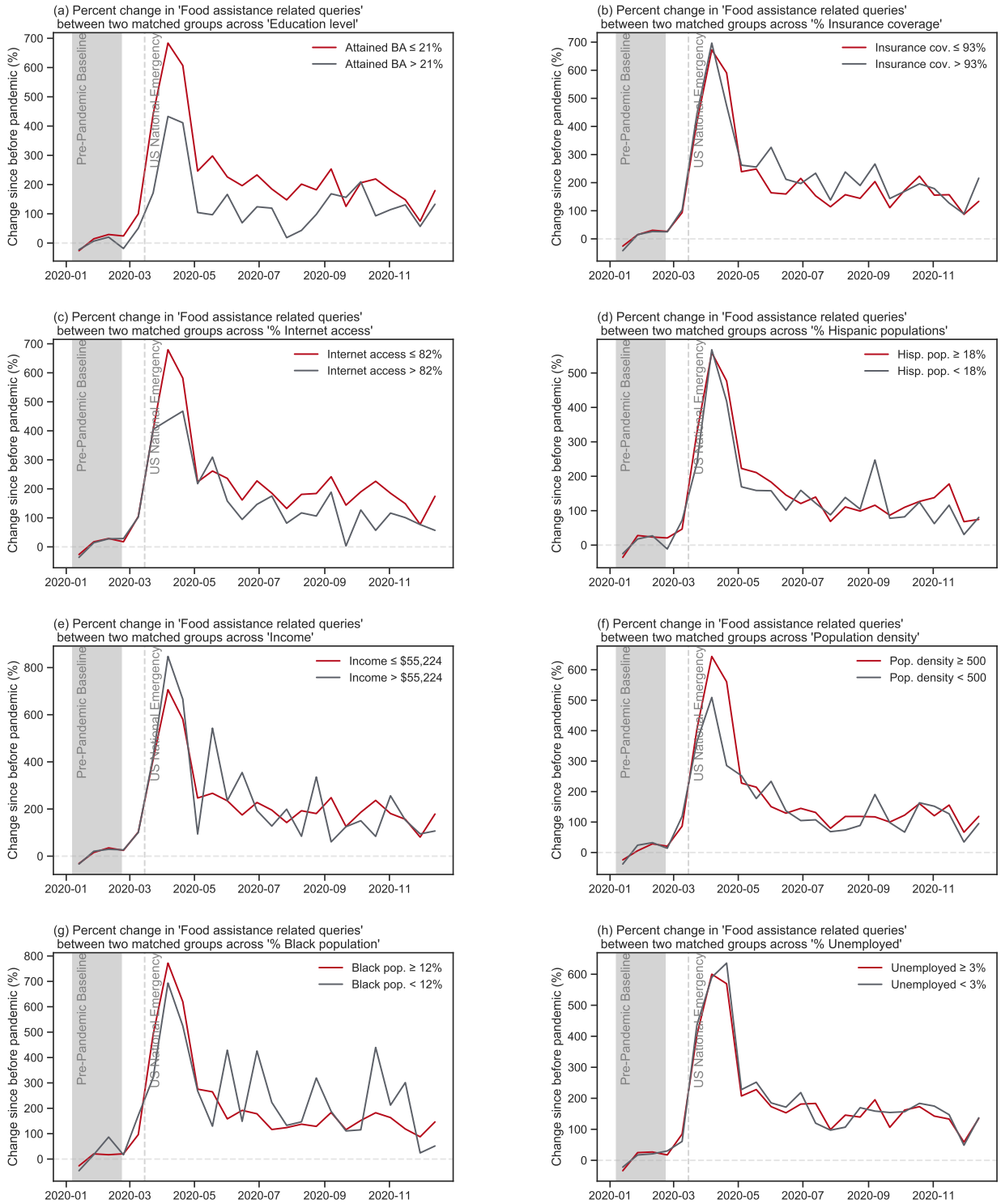
Figure S13. Differences in percentage points for changes in 'Food delivery related queries' between two matched groups across eight census variables.



medRxiv preprint doi: https://doi.org/10.1101/2021.09.14.21263545; this version posted September 17, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Figure S14. Percent change in 'Food assistance related queries' between two matched groups across eight SDoH factors.



medRxiv preprint doi: https://doi.org/10.1101/2021.09.14.21263545; this version posted September 17, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

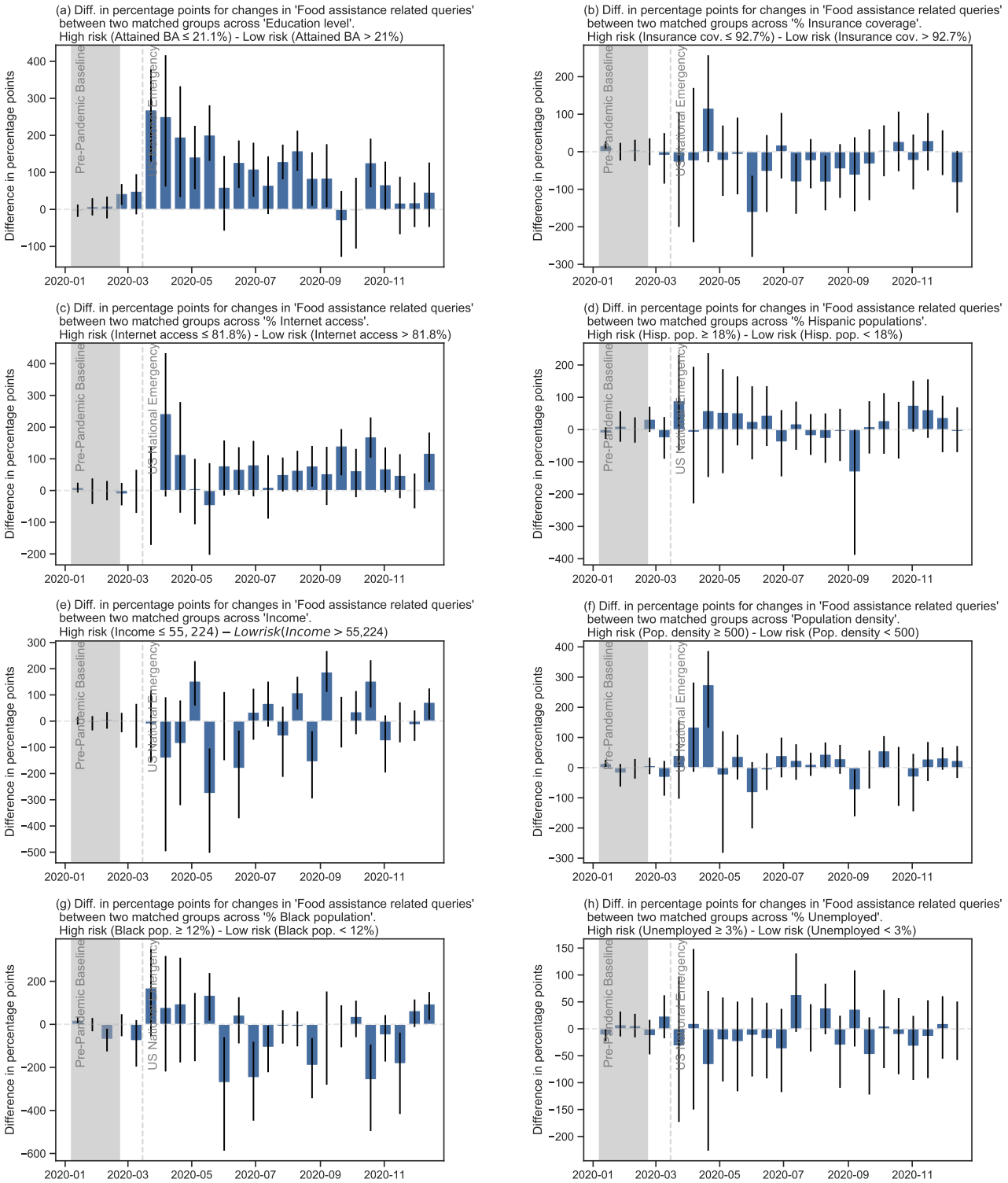



Figure S15. Differences in percentage points for changes in 'Food assistance related queries' between two matched groups across eight SDoH factors.