Contagion at Work: Occupations, Industries and Human Contact*

Anna Houštecká
Institute of Employment Research (IAB)

Dongya Koh
University of Arkansas

Raül Santaeulàlia-Llopis
UAB, Barcelona GSE and CEPR

May 7, 2021

Abstract

Using nationally representative micro panel data on flu incidence from the Medical Expenditure Panel Survey in the United States, we show that employed individuals are on average 35.3% more likely to be infected with the flu virus. Our results are robust to individual characteristics including vaccinations, health insurance and individual fixed effects. Within the employed, we find significant differences in flu incidence by occupation (e.g., sales occupations show 40.5% higher probability of infection than farmers) and by industry (e.g., education, health and social services show 52.2% higher probability of infection than mining). Further, we show that the interaction between occupations and industries is important to understand contagion. Indeed, cross-industry differences in flu incidence cannot be fully explained by differences in the within-industry occupation structure. As a potential mechanism for contagion, we study how flu incidence varies with the extent of human contact interaction at work—with an occupation-industry-specific score that we construct based on O'NET occupational characteristics. We find that the higher the human contact at work, the greater are the odds of infection. Our results are larger in years of high aggregate flu incidence and robust to firm size, a number of jobs and hours worked.

Keywords: Contagion, Flu, Employment, Unemployment, Occupations, Industry, Human Contact, Vaccines, Lockdown, Policy, Macroeconomics

JEL classification: E06, J01, J06

*We thank Chris Busch, Alex Ludwig and Ludo Visschers for their comments and suggestions at early stages of this project. We also thank Gabriel A. Facchini, Etienne Wasmer and the participants of the UAB Student and Alumni Webinar and of the Institute for Employment Research (IAB) Brownbag Seminar for their comments. Raül Santaeulàlia-Llopis acknowledges financial support from the AGAUR 2020PANDE00036 "Pandemies 2020" Grant and from the Spanish Ministry of Economy and Competitiveness, through the Proyectos I+D+i 2019 Retos Investigacion PID2019-110684RB-I00 Grant and the Severo Ochoa Programme for Centres of Excellence in R&D (CEX2019-000915-S).

1 Introduction

In pandemics in which a virus spreads via droplets, fomites or close *physical* human contact interaction (e.g., the 1918 influenza, SARS and Covid-19),[1] the public health response typically involves the enactment of economic lockdowns. The immediate goal of these policies is to reduce the spread of the virus generated from economic activity (e.g. Adda, 2016; Aleman et al., 2020a; Alvarez et al., 2020; Atkeson, 2020; Barro et al., 2020). That is, at the core of lockdown policies there is a trade-off between public health and economic activity. Therefore, the assessment of these policies inevitably requires knowledge about the contagion rate associated with different forms of economic activity. Unfortunately, since representative data on actual infections is rather limited, contagion rates at work (and elsewhere, for that matter) are often not directly observable.[2]

In this paper, we exploit a rare opportunity to study virus contagion at work. We focus on influenza (henceforth, flu) infections using nationally representative micro panel data from the Medical Expenditure Panel Survey (MEPS). The MEPS collects individual flu infection—together with a wide range of individual characteristics, including labor market variables—for a span of twenty years in the United States. First, we study contagion rates by employment status. Within the employed, we focus on flu incidence by occupations, industries and their interaction. Second, we assess how actual flu incidence varies with specific occupation-industry exposure to human contact interaction at work—a score that we construct based on Occupational Information Network (O'NET) work requirements. Several findings arise.[3]

First, employed individuals are more likely to be infected with the flu than the rest of the population controlling for year and season (month). Since the employed can differ from the non-employed by a set of individual characteristics, we estimate differences in flu incidence by employment status controlling for age, gender, marital status, households size and medical conditions. In our benchmark specification, the employed are 35.3% more likely to be infected with the virus than the non-employed. Unpacking the employed and non-employed, wage earners are 8.9% more likely to be infected than the self-employed, 30.1% more likely to be infected than the unemployed and 40.8% more likely to be infected than individuals out of the labor force. Additional controls on vaccination and access to health insurance do not alter our results.[4] With individual

[1] See a detailed comparison among these viruses in the online discussion at the World Health Organization.

[2] For long-standing epidemics such as the human immunodeficiency virus (HIV) epidemic there is larger availability of nationally representative micro data; see Iorio and Santaeulàlia-Llopis (2016) and Aleman et al. (2020b).

[3] Our data and results are available in this permanent link: "ContagionAtWork" GitHub repository.

[4] The CDC monitors influenza vaccine effectiveness (VE) every year through the US Flu VE Network. Across recent years the CDC reports VE of 60% in 2010-11, 47% in 2011-12, 49% in 2012-13, 52% in 2013-14, 19% in 2014-15, 48% in 2015-2016 and 40% in 2016-17. The medical literature suggests that lower VE is partially associated with a mismatch of the vaccine and the active strand of influenza. For example, in a recent meta-analysis of observational VE studies conducted in ambulatory care settings from 2004 to 2015, Belongia et al.

fixed effects, the difference in flu incidence between the employed and the non-employed slightly increases to 38.4%. Further, employed individuals that work more hours are more likely to be infected. That is, contagion increases with both extensive and intensive margins of labor supply.

Within the employed, we find significant differences in contagion across occupations and industries.[5] Across occupations, we find that relative to the occupation with the lowest flu incidence, "farming, fishing and forestry," there are 40.5% higher odds of infection for the occupation with the highest flu incidence, "sales and related occupations." This is followed by "professional and related occupations", "office and administrative support" and "management, business, and financial operations" that show differences in flu incidence of, respectively, 35.6%, 31.4% and 31.1% with respect to the occupation with the lowest flu incidence. This figure is 23.1%, 10.3% and 9.9% for, respectively, "service occupations", "construction, extraction, and maintenance" and "production, transportation, and material moving operations." Across industries, we find that relative to the industry with the lowest flu incidence, "mining," the odds of infection are 52.2% larger for the industry with the highest flu incidence, "education, health and social services" (EHSS). This is followed by "public administration", "professional/business" and "other services" with a difference in flu incidence of, respectively, 44.0%, 43.9% and 41.2% with respect to the industry with the lowest flu incidence. This figure is 3.5%, 12.8% and 21.7%, respectively, for the industries of "natural resources", "transportation" and "construction."

Although there are cross-industry differences in occupational structure (e.g., there are more "professionals and related occupations" in EHSS than in the rest of the economy), the cross-industry differences cannot be fully accounted for by the within-industry differences in occupation structure. This is due to the fact that occupation-specific flu incidence differs by industry. For example, "management, business, and financial operations" has the largest contagion rate in the EHSS industry but fares as the lowest contagion rate in "wholesale and retail trade." Similarly, the contagion rate of "service occupations" ranks highest in "wholesale/retail trade" and lowest in "information." This non-monotonicity in flu incidence across occupations and industries implies that the interaction between occupations and industries is important to understand contagion rates—that is, separately focusing on occupations or industries shows only a partial view of the patterns of contagion. Indeed, with a simple decomposition exercise that compares the industry with the largest flu incidence, EHSS, with the rest of the economy, we find that differences in

(2016) show that VE against influenza B viruses was (on average) 54%, against A(H1N1)pdm09 viruses 61% and against H3N2 viruses 33%. Indeed, in the season with the lowest annual VE effectiveness over the past years (19%, in 2013-14) the CDC reports that the predominant virus was H3N2.

[5]MEPS provides individual data for eight occupation groups and thirteen industries. For the entire sample period, MEPS also specifies military specific occupations, which are by far the least affected by the flu. We also observe that 85% of those in military occupations had been vaccinated for the flu. We drop these military specific occupation and also "unclassifiable occupations", which together consist of less than 2% of our sample.

the within-industry occupation structure only partially explain—by approximately one fourth—the total cross-industry difference in flu incidence.

Second, we explore a potential mechanism for the occupation-industry-specific differences in flu incidence. Our idea is that occupations and industries with more human contact interaction at work are subject to larger contagion risk and, hence, flu incidence. To address this question, we construct an occupation-specific human contact score—from a set of O'NET descriptors—that captures the extent of human contact interaction at work for a rich gradation of occupations. We construct this human contact score as the first principal component of all the O'NET descriptors from "Work Activities: Interacting with Others" plus two additional descriptors from "Work Context" that include "Physical Proximity" and "Contact with Others". Then, we aggregate this O'NET occupation-specific human contact score up to a set of broader MEPS occupation-industry groups using as weights fine detailed occupation-industry employment shares from the Occupational Employment Statistics (OES). With this procedure we construct the occupation-industry human contact score that we use to assess whether flu incidence varies with the extent of human contact at work and by how much. Our main finding is that, for all specifications, the higher the human contact score, the higher is the flu incidence. In our benchmark specification that controls for individual characteristics, we find that a one percent increase in the human contact score increases the probability of infection by 0.448 p.p. We also split the sample in years of high annual aggregate flu incidence (above median) and low annual aggregate flu incidence (below median). The human contact score increases the probability of infection in years of high annual aggregate incidence. Our results are robust to firm size—i.e., number of employees, number of jobs and hours worked.

Related literature Our work relates to a strand of work in epidemiology that estimates contagion rates at work and other settings of human interaction. Mossong et al. (2008) and Klepac et al. (2020) measure setting-specific contagion (e.g. home, work and school, among others) constructing large-scale social mixing patterns from primary data on human contacts. Alternatively, some epidemiological studies, such as Ferguson et al. (2005) which also focuses on the flu, are carried out under the assumption that the levels of contagion are approximately equal across settings (i.e., at work, school and the general community). This is also the case of Halloran et al. (2008) who, for example, write: *"In the absence of data to inform the choice, transmission in other contexts was arbitrarily partitioned to give levels of within-place transmission comparable with household transmission, namely 33% of transmission was assumed to occur in schools and workplaces, and 37% in the wider community (i.e., in contexts other than households, schools, and workplaces)."* We contribute to this literature focusing on a specific setting—work—by providing estimates on how actual contagion varies by employment status, occupations and industries.

There is also a body of epidemiological literature that carefully studies the relationship between occupations and influenza-like illness (e.g. Anderson et al., 2012; Luckhaupt et al., 2014; Østergaard et al., 2021). We add to this literature by studying the interaction between occupations and industries that can be relevant for the assessment of public health policy against the spread of a pandemic.[6] Further, we combine our estimates of actual occupation-industry-specific flu incidence with a score that summarizes the extent of physical human interaction at the workplace from O'NET descriptors. Then, we empirically assess how contagion varies with the extent of human contact at work.

In that direction, closest to our work is Almagro and Orane-Hutchinson (2020) that relates Covid-19 hospitalizations and occupation structure at the zip code level in New York.[7] In contrast, in the context of the flu, we directly link occupations and incidence at the individual level without requiring any geographical (or other types of) aggregation.[8] In addition, concurrent work also uses O'NET occupation information providing alternative measures of human contact interaction at the occupation level (Mongey et al., 2020) or the industry level (Azzimonti et al., 2020) to proxy for the exposure to the Covid-19 risk of infection. Our main point of departure from these papers is that we directly link—at the individual level—our measure of occupation-industry-specific human contact interaction with actual contagion. This allows us to assess, for the first time, how actual contagion risk differs by the extent of human contact at work.

Our work is also related to the macroeconomic literature that assesses optimal lockdown policies in so far as policy assessment requires estimates of the contagion rate associated with economic activity. This type of policy analysis has recently regained a new and growing interest due to the Covid-19 pandemic (see Alvarez et al. (2020), Aspri et al. (2021), Atkeson (2020), Bognanni et al. (2020), Casares et al. (2020), Eichenbaum et al. (2020), Farboodi et al. (2020), Fajgelbaum et al. (2020), Garibaldi et al. (2020), Glover et al. (2020) and Kaplan et al. (2020), among many others).[9] However, as far as we know, the current assessments of policy to fight against the Covid-19 epidemic do not incorporate information from actual occupation and industry contagion rates at work. In this context, our estimates for the flu—which pinpoint the contagion risk of employment and highlight the fact that this risk is heterogeneous across occupations

[6]The interaction between occupations and industries has proven useful to understand other aspects of the labor market, such as occupational mobility (Carrillo-Tudela and Visschers (2020) and Darougheh (2019)).

[7]More generally, the effects of Covid-19 on the labor market have been extensively studied; see, among many others, Adams-Prassl et al. (2020), Cajner et al. (2020), Chetty et al. (2020) and Eyméoud et al. (2021).

[8]In the context of the Covid-19, Pollan et al. (2020) uses a serological survey in Spain to document incidence by employment status during a period of national lockdown, which may underestimate the impact of employment on infection rates. Our study provides estimates during times of no lockdowns and includes an occupation and industry analysis to assess whether contagion varies with human contact interaction.

[9]This growing literature also includes testing and quarantine policies; see Berger et al. (2020), Obiols-Homs (2020) and Piguillem and Shi (2020), among others.

and industries—opens the door for the assessment of optimal nonpharmaceutical policies against a pandemic (e.g. stay-home policies and economic lockdowns) that are potentially shaped by occupation-industry-specific contagion risk.

Finally, our work also speaks to recent empirical assessment of sick-leave mandates. In particular, Pichler and Ziebarth (2017) introduce the notion of contagious presenteeism and show that sick-leave mandates contribute to reducing the spread of the flu.[10] More recently, Pichler et al. (2020) shows how emergency sick leaves reduce the spread of Covid-19 through the U.S. Families First Coronavirus Response Act. In this context, our work can prove useful for the potential assessment of optimal—and potentially targeted—sick-leave mandates by providing an additional dimension—the contagion risk by occupations and industries—to that type of analysis.

2 Data

We use the Medical Expenditure Panel Survey (MEPS) as our main data to study individual flu incidence in relation with a wide range of labor market variables that include employment status, occupations, and industries. We also use the Occupational Information Network (O'NET) and Occupational Employment Statistics (OES) to construct an occupation-industry-specific measure of physical human contact in order to assess how the extent of physical human contact required by industry and occupation relates to contagion.[11]

2.1 Medical Expenditure Panel Survey (MEPS)

The Medical Expenditure Panel Survey (MEPS) is a set of surveys of families and individuals, their medical providers and employers across the United States.[12] We use the Household Component of the Medical Expenditure Panel Survey (MEPS-HC) which is a nationally representative survey of the U.S. civilian noninstitutionalized population. Data collection has been running since 1996 and the data set consists of yearly panels, in which data are collected for two calendar years for the same individual. Within those 2 years, MEPS conducts 5 rounds of interviews on each individual.[13] The average number of families surveyed per year is 12,500 and the average number

[10]A large set of epidemiological studies also focus on sick leaves. For example, Zhai et al. (2018) investigates sociodemographic characteristics of receiving paid sick leave benefits due to the flu. Vahtera et al. (1999) shows that blue-collar workers use longer sick leaves due to infection than white-collar workers using three towns data in Finland. Alternatively, using a modelling approach Potter et al. (2015) studies workplace contact networks and find, through simulations, that face-to-face social contacts are important for the transmission of respiratory infections. A meta analysis of these two specific approaches to study the flu is in Edwards et al. (2016). (This footnote in the first paragraph?)

[11]More details on O'NET, OES and how we merge these to the MEPS can be found in Section 4.1.

[12]The data can be found at `https://www.meps.ahrq.gov/mepsweb/`.

[13]Each interview round collects information corresponding to a reference period, which is household-specific. The first round of each panel begins on the January 1 of the year where the panel started and ends at the date

of individuals is 31,000. This implies a total amount of approximately 65,000 interviews per year and a total of approximately 4 million observations throughout. Further, note that the data from a particular round spans several months. To understand the seasonality of the flu in our data we create monthly data based on the interview month.

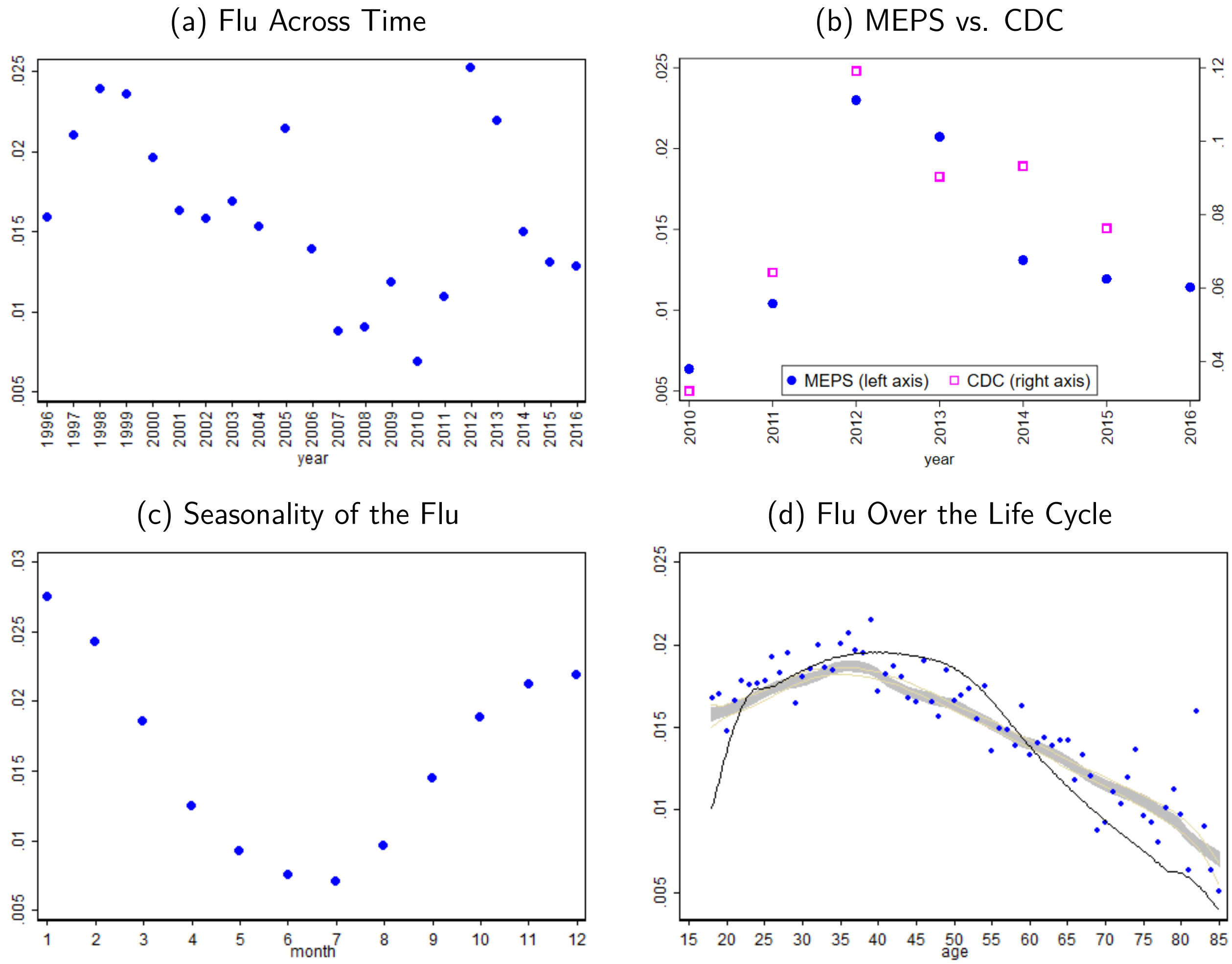
The Household Component includes data on health conditions, services, and insurance together with information on individual employment (e.g., employment status, occupation, industry, hours worked, number of employed, whether or not an individual was self-employed, etc.) and other individual characteristics. The Medical Conditions files which are at the event-level give us information about all medical conditions that the individual reported in a particular round. Hence, we know in which round individuals first report the flu per year and what other medical conditions they report. In addition, the Preventive Care modules, which take place only in rounds 3 and 5, report information on the time since the individual received a flu shot.

Flu Incidence in MEPS An annual flu incidence is constructed as total number of individuals infected with the flu divided by the entire population in a given calendar year. We find that the annual flu incidence is 1.6% on average between the years 1996 and 2016. The average flu incidence has large annual changes; see panel (a) of Figure 1. For example, in late 1990s the flu incidence reached average values for the entire population of approximately 2.4%. In contrast, in the late 2000s the average incidence was approximately 1%. The maximum average prevalence in our sample year reaches approximately 2.5% in 2012. The behavior of the flu incidence that we construct using MEPS is compared to that reported by the Center for Disease Control (CDC) in panel (b) of Figure 1. Note that the flu incidence that we calculated using MEPS is the total number of individuals infected with the flu divided by the entire population, while the CDC reports the ratio between the number of positive tests divided by the total number of specimens tested. This explains the higher values in the CDC compared with MEPS. Nevertheless, the behavior of the time series of flu incidence is similar between MEPS and CDC after 2010.[14]

of the first interview. The reference periods for Rounds 2, 3, and 4 varied from household to household and covered the time between interview dates of the previous round and the current round. In our sample 80% of the first interviews were conducted between February and May of the first year, 80% of the second interviews were conducted between August and October of the first year. In the same way, 80% of the third and fourth interviews conducted during the second year of the panel were between January and April and between July and October, respectively. The last reference period ends on December 31 of the second year of the panel. A reference period thus corresponds, on average, to 5 months. See more details in Appendix A.

[14]Unfortunately, we cannot reproduce the CDC statistic using MEPS data because we have no information about whether an individual was tested for the flu in MEPS. In addition, the CDC data are not engineered to be necessarily representative, while MEPS is constructed with a sample design that aims to be representative. Instead, the CDC data are collected using approximately 270 National Respiratory and Enteric Virus Surveillance System laboratories and 110 United States World Health Organization Collaborating Laboratories located in the United States. These laboratories include public health laboratories and clinical laboratories. Since public health laboratories often receive samples from a clinical laboratory to test the types of influenza virus, the results could

Figure 1: Flu Incidence across Years, Seasons and Age, MEPS 1996-2016



Notes: This figure shows flu incidence, i.e., the total number of individuals infected with the flu divided by the total population, constructed using the MEPS 1996-2016 panels. In panel (a) we show how the flu incidence varies across years. In panel (b) we compare the flu incidence constructed from MEPS (left axis) with the number of specimens that are tested positive for flu (conditional on testing) from the Center for Disease Control (CDC). In panels (c) and (d) we show, respectively, the (monthly) seasonality of flu incidence and the behavior of flu incidence over the life cycle. In panel (d) we include the empirical density of age in our MEPS sample (black line).

Although the flu is most common during the fall and winter, it is actually detected year-round.[15] The exact month at which the flu peaks in a year can change across years, but most

be duplicated prior to year 2015 where the test results are not reported separately from each laboratory.

[15] This includes different strands of influenza of type A (e.g., H1N1pdm09, H3N2 and others) and of type B (e.g., Yamagata lineage, Victoria lineage, and others).

of the flu activity peaks between October and March, see panel (c) in Figure 1.[16] In that figure, we show the patterns of flu incidence across months pooling the information from all our MEPS sample years, 1996-2016. The seasonal pattern that emerges is clear. The flu reaches its maximum incidence from October to March with the highest level of incidence at 2.56% in January and the lowest level of incidence close to 0.71% in June and July. Note that survey retrieves information with an average of five-month recall across interview rounds, which tends to mitigate the differences in incidence between months of high and low incidence; see Appendix A for further details on how the panel rounds are designed in the MEPS.

The CDC estimates that during the 2018–2019 season there were 35.5 million people getting sick with flu, 16.5 million people visiting a health care provider for their illness, 490,600 hospitalizations and 34,200 deaths from flu.[17,18] More than 46,000 hospitalizations occurred in children (aged $<$18 years); however, 57% of hospitalizations occurred in older adults aged $\geq$ 65 years. Older adults also accounted for 75% of flu-associated deaths, highlighting that older adults are particularly vulnerable to severe outcomes resulting from a flu virus infection. An estimated number of 8,100 deaths occurred among working age adults (aged 18–64 years), an age group that often has low flu vaccination uptake. For this reason, we now also assess how flu incidence varies with age in our sample; see panel (d) of Figure 1 where we plot the age profile of the flu together with the age density in our sample. The flu incidence significantly differs by ages increasing from approximately 1.7% in the late teens to above 2% in the mid 30s to 40. After the early 40s, the flu incidence relentlessly declines to values below 1.3% after retirement and below 1% for individuals who are in their 70s. Despite the lower rate of flu infections in the older age groups, this is the population that is suffers more deaths due to flue according to the CDC. Hence, the old are the population that suffers higher flu fatality understood as the rate at which infections lead to deaths. It is clear, however, that for individuals above 18, most of the infections occur to the working age population.

Benchmark Sample. To conduct our analysis, we use the first 20 panels of the MEPS, which span the years 1996-2016. In our benchmark sample we keep all individuals in the working age population, i.e., above 18 years old and below 65 years of age. This leaves us with 180,737 individuals and 3,932,565 monthly observations in total. Of these observations, 2,773,444 (70.5%)

[16] Recently, the CDC reports similar peak months of flu activity from a longer sample period, 1982-2018. See here: https://www.cdc.gov/flu/about/season/flu-season.htm.

[17] See https://www.cdc.gov/flu/about/burden/2018-2019.html.

[18] The number of flu-associated illnesses that occurred last season was similar to the estimated number of flu-associated illnesses during the 2012–2013 flu season when estimated 34 million people had symptomatic flu illness.

Table 1: MEPS Sample Statistics

	Full Sample	Flu=1	Flu=0	Diff.	t
Employed	0.75	0.80	0.75	-0.05***	(-33.34)
Age	40.29	39.54	40.31	0.77***	(15.40)
Female	0.52	0.54	0.52	-0.02***	(-13.02)
Never Married	0.30	0.29	0.30	0.01**	(2.57)
HH Size	5.47	5.51	5.47	-0.03**	(-2.76)
N. Other Conditions	0.92	1.09	0.91	-0.18***	(-29.45)
Student Status	0.06	0.06	0.06	0.00	(0.57)
Flu Shot	0.49	0.49	0.49	-0.00	(-0.33)
N. Employees	101.23	107.09	101.13	-5.96***	(-9.13)
Observations	3,932,565	65,651	3,866,914		
Observations (Vaccines)	3,704,726				
% Of Total	(94.2%)	(1.67%)	(98.33%)		

Notes: This table shows the summary statistics of our benchmark sample (age 18-64), the sample of individuals infected with the flu (Flu=1) and not infected with the flu (Flu=0). We use sampling weights. The flu shot is observed for a subset of the individuals in the benchmark sample. The number of employees is set to zero for the non-working age population.

were employmed individuals, while the rest were not working.[19] In Table 1 we report the sample statistics for the full sample and across individuals who report a flu infection and those who report no flu infection. The average age is 40 years old, 52% are females and 30% have never been married, on average 5.7 people cohabitate and report less than one medical condition other than the flu (examples of other conditions include cancer, diabetes, sexually transmitted diseases or pneumonia).[20] Moreover, 6% are students, 50% of the observations had a flu shot, and the average number of employees is approximately 100.[21]

For the sample that reports a flu infection, we find that 78% of the individuals are employed, 54% are female, 29% were never married, 6% are students, and 50% had a flu shot in the past. As compared to the sample of individuals not infected with the flu, the proportion of employed individuals and females, the number of other conditions reported, the number of employees of the firm and the average size of the household are significantly higher, whereas the average age and the proportion of never married individuals are significantly lower. The proportion of students and those who had a flu shot does not differ significantly between the flu and no-flu populations.

[19] In MEPS the employment status variables do not distinguish between unemployed workers and those out of the labor force. For individuals who have worked before, we observe the reason for not working which we can use to further decompose the non-working population.

[20] A detailed list of all medical conditions included in the MEPS can be found in Appendix 3 of `https://meps.ahrq.gov/data_stats/download_data/pufs/h180/h180app3.html`.

[21] We set the number of employees to zero for those who are not working.

3 Contagion at Work

We use the MEPS to study flu infection at work. First, we focus on how flu incidence varies with employment status in Section 3.1. Second, within the employed, we study how flu incidence varies across industries and occupations in Section 3.2.

3.1 Employment Status and the Flu

We now study the how flu incidence varies with employment status for individuals of 18 years of age and above.[22] Specifically, we run the following linear probability model specification:

$$flu_{it} = cons + \gamma e_{i,t} + \sum_t \omega_t \mathbf{1}_t + \sum_x \beta x_{it} + u_{it} \quad (1)$$

where flu_{it} is an indicator equal to one if an individual i reports a flu infection at month-year observation t and zero otherwise. If individuals are employed we set $e_{i,t}$ equal to one and zero otherwise; therefore, γ captures how the odds of infection vary with employment status. We capture the effects of years and seasons with a set of month-year dummies, $\mathbf{1}_t$. In our benchmark specifications we also control for a set of individual characteristics that we denote as x_{it}. The set of individual characteristics includes age (a cubic), gender, family size, marital status, and health status including a number of medical conditions reported in the same round as a proxy for the general health status of the individual. The residual is captured by u_{it}.

We first run our specification controlling for month-year effects but without individual controls. This gives us a measure of unconditional association between employment status and the probability of being infected with the flu; see column (1) in Table 2. We find that, unconditionally, employed individuals have higher flu infection rates than non-employed individuals. In particular, the estimated coefficient on employment implies a higher probability of infection of 0.461 percentage points (p.p.). That is, employed individuals are 33.3% more likely to get infected with the virus than the non-employed individuals.[23] The association between employment and the prob-

[22] In the MEPS, the labor market status tells us whether the individual has been employed during the round of reference or has not worked during the round. We do not consider the two other cases in which the individual reports that he 1) had a job to return to or 2) was employed sometime during the reference period.

[23] To see this, first we solve for the conditional flu prevalence of the non-employed individuals, x, in

$$\frac{\sum_i - \sum_i \mathbf{1}_{e_i}}{\sum_i} x + \frac{\sum_i \mathbf{1}_{e_i}}{\sum_i}(x+\gamma) \equiv 0.246x + 0.754 \times (x + 0.461) = 1.73 \equiv \frac{\sum_i flu_i}{\sum_i}, \quad (2)$$

which implies that $x = 1.383$. Then, the employed are ((1.383+0.461)/1.383-1)×100=33.3% more likely to be infected than the non-employed population. Note that in equation (2) the term $(\sum_i - \sum_i \mathbf{1}_{e_i})/\sum_i$ is the share of non-employed in the population, $\sum_i \mathbf{1}_{e_i}/\sum_i$ is the share of employed in the population and $\sum_i flu_i/\sum_i$ is the aggregate (average) flu prevalence in the economy.

ability of infection slightly increases to 0.483 p.p. after controlling for individual characteristics; see column (2) in Table 1. This implies that, after controlling for individual characteristics, the employed are 35.3% more likely to get infected with the virus than the non-employed population. Turning into the effects of individual characteristics, we find that flu incidence follows a hump shaped pattern over the life cycle captured by a cubic function of age; women have significantly higher chances of infection by 0.167 p.p.; higher number of medical conditions are associated with significantly higher flu incidence; and being a student also increases the odds of flu infection by 0.204 p.p. The size of the household does not significantly affect the probability of infection.[24]

Vaccination and health insurance. Whether or not an individual gets infected is potentially affected by flu vaccination—vaccine effectiveness against the flu is approximately 45% across recent seasons (2010-2017). Hence, it is important to control for vaccinations since our previous findings could be affected by employed individuals having better access to health insurance and vaccination than non-employed individuals. Individuals in the MEPS are asked how long it had been since they had a flu shot which allows us to control for flu shots.[25] We obtain similar insights to our benchmark; see column (3) of Table 2. While the coefficient on ever getting vaccinated is positive but not significant,[26] the coefficient on the time since vaccine is positive and significant suggesting that an additional year since the last time the individual had a vaccination is associated with 0.034 p.p. higher chances of having the flu; see column (4) of Table 2. Our results stand when we control directly for whether an individual had health insurance in the same round; see column (5) of Table 2. Health insurance status is negatively but not significantly associated with flu incidence.

Subjective Health Measures In our benchmark specification, we control for the number of other medical conditions reported by the individual in the same round (apart from flu, if reported) to capture differences in individual health. We also run a specification where instead of using the number of other medical conditions reported we control for a self-reported general health variable; see column (6) in Table 2. The association between flu incidence and employment increases slightly, and all other coefficients are close in magnitude to those in the benchmark specification.

[24]When we re-run our benchmark specification using the entire population, we find similar results with the employed being 0.470 p.p. more likely to be infected with the flu.

[25]In interview rounds three and five, the MEPS provides retrospective information about how long ago individuals have been last vaccinated.

[26]This can be a result of having a large number of individuals reporting that they had a flu shot during the last year. However, we do not observe when the shot occurs and, hence, it could be before or after getting the flu.

Table 2: Employment Status and Flu Incidence

VARIABLES	(1) Year/Month FE	(2) Benchmark	(3) Vaccines	(4) Vaccines	(5) Insurance	(6) Subj. Health	(7) MSA	(8) Class of Worker	(9) Hours Worked
Employed	0.00461***	0.00483***	0.00478***	0.00491***	0.00470***	0.00519***	0.00476***	0.00514***	0.00275**
	(0.000417)	(0.000440)	(0.000452)	(0.000631)	(0.000452)	(0.000460)	(0.000488)	(0.000472)	(0.00112)
Self-employed								0.00369***	
								(0.000770)	
Unemployed								0.00104	
								(0.00117)	
Employment*Hours									6.10e-05**
									(2.75e-05)
Flu Shot			0.000610						
			(0.000427)						
Time Since Vaccine				0.000356**					
				(0.000145)					
Health Insurance					-0.000188				
					(0.000494)				
Self-reported Bad Health						0.00341***			
						(0.000648)			
MSA							0.000657		
							(0.000559)		
Age		0.00135**	0.00135**	0.00187**	0.00182***	0.00140**	0.00131**	0.00138**	0.00132**
		(0.000587)	(0.000606)	(0.000853)	(0.000601)	(0.000587)	(0.000658)	(0.000592)	(0.000605)
Age^2		-3.22e-05**	-3.22e-05**	-4.51e-05**	-4.43e-05***	-3.38e-05**	-3.08e-05*	-3.24e-05**	-3.24e-05**
		(1.45e-05)	(1.50e-05)	(2.08e-05)	(1.50e-05)	(1.46e-05)	(1.63e-05)	(1.47e-05)	(1.51e-05)
Age^3		2.24e-07*	2.22e-07*	3.22e-07**	3.22e-07***	2.42e-07**	2.08e-07	2.23e-07*	2.33e-07*
		(1.15e-07)	(1.19e-07)	(1.63e-07)	(1.19e-07)	(1.15e-07)	(1.28e-07)	(1.16e-07)	(1.20e-07)
Female		0.00167***	0.00156***	0.00154***	0.00140***	0.00220***	0.00169***	0.00173***	0.00196***
		(0.000407)	(0.000419)	(0.000589)	(0.000422)	(0.000406)	(0.000450)	(0.000411)	(0.000441)
Never Married		-0.000573	-0.000764	-0.000321	-0.000708	-0.000686	-8.40e-05	-0.000513	-0.000865
		(0.000569)	(0.000583)	(0.000811)	(0.000581)	(0.000570)	(0.000642)	(0.000575)	(0.000591)
HH Size		-4.71e-05	-9.48e-05	-7.75e-05	-8.49e-05	-7.06e-05	-3.89e-05	-3.27e-05	-2.02e-05
		(7.03e-05)	(7.19e-05)	(0.000104)	(7.27e-05)	(7.01e-05)	(7.86e-05)	(7.09e-05)	(7.41e-05)
N. Other Conditions		0.00150***	0.00145***	0.00156***	0.00165***		0.00121***	0.00147***	0.00143***
		(0.000159)	(0.000164)	(0.000208)	(0.000169)		(0.000176)	(0.000161)	(0.000165)
Student Status		0.00204*	0.00215*	0.00102	0.00266**	0.00223*	0.00199	0.00212*	0.00244**
		(0.00119)	(0.00123)	(0.00180)	(0.00119)	(0.00119)	(0.00135)	(0.00120)	(0.00121)
Constant	0.0257***	0.00646	0.00652	-0.00124	0.0230***	0.00735	0.00609	0.00561	0.00701
	(0.000895)	(0.00762)	(0.00788)	(0.0113)	(0.00827)	(0.00762)	(0.00857)	(0.00771)	(0.00781)
Observations	3,932,565	3,932,565	3,704,726	1,746,050	3,343,721	3,932,565	3,212,569	3,880,624	3,584,277
R-squared	0.005	0.006	0.006	0.006	0.005	0.005	0.006	0.006	0.006
Year/Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pseudo R2	0.00513	0.00550	0.00554	0.00550	0.00527	0.00535	0.00624	0.00552	0.00558

Notes: This table shows results from a linear probability model for flu incidence on MEPS data 1996-2016. Robust standard errors clustered at the individual level are in parentheses *** p<0.01, ** p<0.05, * p<0.1. In Column (8), the coefficient for "Employed" refers to "Wage Earners" only.

Metropolitan Area The MEPS includes data on whether or not an individual lives in a metropolitan area.[27] This is a potential confounder with employment because both employment and the spread of flu can be larger in metropolitan areas due to the larger density of people engaging in work. The coefficient for employment remains significant even when we explicitly control for individual's residing area; see column (7) in Table 2. Our results remain unaltered if we instead control for urban-rural residence differences.

Wage-Earners, Self-employed, Unemployed and Out of the Labor Force In our analysis we have so far computed the focused on the differences in flu incidence between the employed and the non-employed. Unpacking the non-employed we can group individuals into unemployed and those out of the labor force.[28] In the same way, we also split the employed into self-employed and wage earners. We re-run our benchmark specification with associated dummies for unemployed, self-employed, and wage earners using individuals out of the labor force as a reference group; see column (8) in Table 2. We find that the flu incidence of the unemployed is larger by 0.104 p.p. but not significantly different from that of the individuals who are out of the labor force. Compared with the individuals out of the labor force, the self-employed and the employed wage earners are, respectively, 0.369 p.p. and 0.514 p.p. more likely to be infected with the flu. These results imply that wage earners are 8.9% more likely to be infected than the self-employed, 30.1% more likely to be infected than the unemployed, and 40.8% more likely to be infected than individuals out of the labor force.

Hours Worked The results above demonstrate that the extensive margin of labor supply is significantly related to flu prevalence at work. The MEPS data also contains information on usual hours worked per week and thus allows us to determine whether the intensive margin of labor supply also matters for flu incidence at work. We run our benchmark regression that additionally includes an interaction of employment with the hours worked,[29] restricting our attention to wage-earners; see column (9) in Table 2.[30] We find that there is a positive and significant association between hours worked and the probability of catching the flu.

Individual Fixed Effects. Given the panel structure of the MEPS and the fact that we have enough variation in employment status within individuals we also run a specification including individual fixed effects in order to control for potential permanent unobserved individual charac-

[27]The MSA variable is only available up to 2015; thus, we run our regression on a smaller sample.

[28]We define unemployed as those who report the reason for not working as 1) Could not find work, 2) On temporary layoff, and 3) Waiting to start new job.

[29]We trim the upper and lower 1% of reported weekly hours in the MEPS, corresponding to less than 5 and more than 72 weekly hours worked.

[30]Since the sample includes employed and non-employed, the hours worked of the non-employed are zero.

Table 3: Employment Status and Flu Incidence: Individual Fixed Effects

VARIABLES	(1) Benchmark	(2) + Indiv. FE	(3) + Vaccine
Employed	0.00483***	0.00496***	0.00515***
	(0.000440)	(0.00110)	(0.00117)
Flu Shot			-0.00145
			(0.00101)
Age	0.00135**	0.00297	0.00284
	(0.000587)	(0.00302)	(0.00314)
Age^2	-3.22e-05**	-0.000117	-0.000112
	(1.45e-05)	(7.70e-05)	(8.05e-05)
Age^3	2.24e-07*	1.03e-06*	9.85e-07
	(1.15e-07)	(6.17e-07)	(6.47e-07)
Female	0.00167***		
	(0.000407)		
Never Married	-0.000573	-0.000649	-0.000236
	(0.000569)	(0.00253)	(0.00270)
HH Size	-4.71e-05		
	(7.03e-05)		
N. Other Conditions	0.00150***	-0.00153***	-0.00139***
	(0.000159)	(0.000205)	(0.000210)
Student Status	0.00204*	0.00317*	0.00423**
	(0.00119)	(0.00189)	(0.00186)
Constant	0.00646	0.0707*	0.0595
	(0.00762)	(0.0396)	(0.0408)
Observations	3,932,565	3,932,565	3,704,726
R-squared	0.006	0.005	0.006
Year/Month FE	Yes	Yes	Yes
Individual FE	No	Yes	Yes
Pseudo R2	0.00550	0.00548	0.00551
Number of id		180,737	177,586

Notes: This table shows results from a linear probability model for flu incidence on MEPS data 1996-2016. Robust standard errors clustered at the individual level are in parentheses *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

teristics. The coefficient for employment stays remarkably stable even if we use within-individual variation to study the relationship between employment status and flu incidence. If anything, controlling for time-invariant characteristics increases the association between the odds of infection and employment to 0.496 p.p.; see column (2) of Table 3. In addition to controlling for individual fixed effects, if we control for vaccinations, we find even a slightly larger employment coefficient of 0.515 p.p., which implies a 38.4% higher probability of being infected with the flu; see column (3) of Table 3. It is interesting to note that the sign of the coefficient of the number of other conditions changes. This points to the fact that variation in the number of other conditions across individuals summarizes well differences in permanent health across indi-

Table 4: Employment Status and Flu Incidence: High versus Low Aggregate Incidence

VARIABLES	(1) All years	(2) High Incidence	(3) Low Incidence
Employed	0.00483***	0.00574***	0.00393***
	(0.000440)	(0.000701)	(0.000498)
Age	0.00135**	0.00110	0.00153**
	(0.000587)	(0.000906)	(0.000681)
Age^2	-3.22e-05**	-2.24e-05	-4.01e-05**
	(1.45e-05)	(2.25e-05)	(1.69e-05)
Age^3	2.24e-07*	1.23e-07	3.09e-07**
	(1.15e-07)	(1.78e-07)	(1.34e-07)
Female	0.00167***	0.00239***	0.000963**
	(0.000407)	(0.000629)	(0.000480)
Never Married	-0.000573	0.000469	-0.00159**
	(0.000569)	(0.000887)	(0.000666)
HH Size	-4.71e-05	5.17e-05	-0.000128
	(7.03e-05)	(0.000114)	(8.11e-05)
N. Other Conditions	0.00150***	0.00112***	0.00191***
	(0.000159)	(0.000243)	(0.000198)
Student Status	0.00204*	0.00139	0.00256*
	(0.00119)	(0.00184)	(0.00135)
Constant	0.00646	0.0101	0.00465
	(0.00762)	(0.0117)	(0.00878)
Observations	3,932,565	1,995,272	1,937,293
R-squared	0.006	0.006	0.003
Year/Month FE	Yes	Yes	Yes
Pseudo R2	0.00550	0.00553	0.00329

Notes: This table shows results from a linear probability model for flu incidence on MEPS data 1996-2016. High incidence years are those where flu incidence was higher than 1.67%. Robust standard errors clustered at the individual level are in parentheses *** $p<0.01$, ** $p<0.05$, * $p<0.1$

viduals and that those individuals more prone to having other medical conditions also catch the flu. However, when using time variation within individual, flu rounds are associated with a lower number of other reported conditions.[31] Including vaccinations, though insignificant, we find a negative relationship between vaccinations and flu incidence within individuals.

Aggregate Flu Incidence. Across years, the median annual aggregate flu incidence was of 1.67%. Using the median aggregate flu incidence as a threshold, we split our sample between years above and below the median aggregate incidence which we define, respectively, as high and low incidence years. We find that the employment coefficient is increasing with the aggregate

[31] Since for each medical condition in the MEPS we observe the first round it was recorded, it is possible that an individual reports less new conditions in the same round as the flu as they are less likely to contract other contagious illnesses and less likely to get preventive care for other conditions and get diagnoses of longer term conditions such as cancer or diabetes.

flu incidence in the economy. In high aggregate flu incidence years the employed have higher infection rates than the non-employed by 0.574 p.p.; see column (2) in Table 4. This estimate is of 0.393 p.p. in years with low aggregate flu incidence; see column (3) in Table 4. The difference in the employment coefficient is significant between high and low aggregate incidence years.

3.2 Flu Incidence by Occupation and Industry

In this section we focus on employed individuals and explore the flu incidence by occupation, by industry and by occupation-industry mix.

Occupation-specific flu incidence. We restrict our attention to the 2002-2016 sample because the MEPS categorization of occupations changed in 2002. After 2002, the occupational groups are condensed census codes and correspond to the following eight occupations: management, business, and financial operations; professional and related occupations; service occupations; sales and related occupations; office and administrative support; farming, fishing and forestry; construction, extraction, and maintenance; and production, transportation, and material moving operations.[32] The two largest occupations are "professional and related occupations" and "service occupations" that, respectively, account for 23% and 16.4% of the total employment in the MEPS for 2002-2016. In terms of employment size, they are followed by "management, business, and financial operations","office and administrative support" and "production, transportation, and material moving operations" that account for, respectively, 15.6%, 13.2% and 12.5% of total employment. The occupations of "sales and related occupations" and "construction, extraction, and maintenance" represent, respectively, 9.7% and 8.9% of total employment. The occupation with the lowest share of employment is "farming, fishing and forestry" that represents 0.7% of total employment.

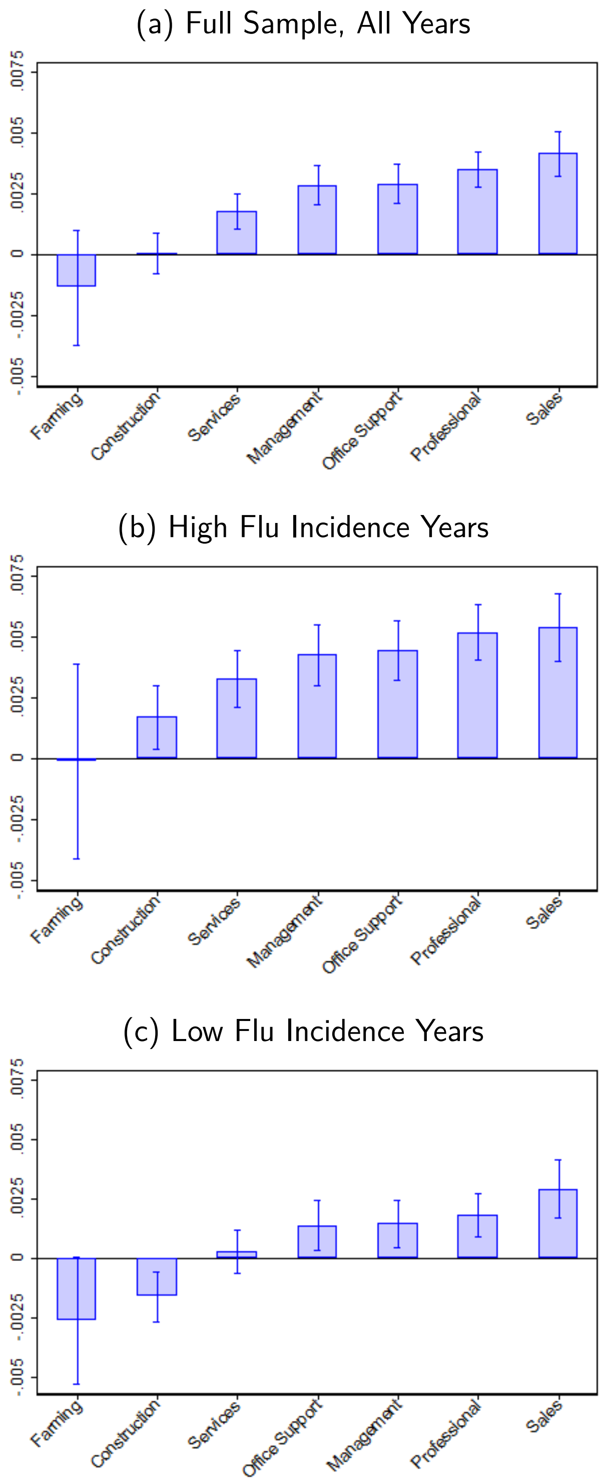
We first estimate how flu incidence differs by occupations. Precisely, we compute the occupation-specific effects on flu incidence estimates, β_{occ}, from the following regression specification,

$$flu_i = cons + \sum_{occ \neq 1} \beta_{occ} \mathbf{1}_{occ} + \sum_t \omega_t \mathbf{1}_t + u_{it}, \tag{3}$$

where "production, transportation and material moving operations" is the reference occupation indexed as 1. We control for year and month. Our results are in panel (a) of Figure 2.

[32] For the entire sample period, MEPS also specifies military specific occupations, which are by far the least affected by the flu. We also observe that 85% of those in military occupations had been vaccinated for the flu. We drop these military specific occupation and also "unclassifiable occupations," which together consist of less than 2% of our sample.

Figure 2: Occupation-Specific Flu Incidence, MEPS 2002-2016



Notes: The occupations in MEPS are (1) management, business, and financial operations; (2) professional and related occupations; (3) service occupations; (4) sales and related occupations; (5) office and administrative support; (6) farming, fishing and forestry; (7) construction, extraction, and maintenance; and (8) production, transportation, and material moving operations. Our data is retrieved from MEPS, 2002-2016.

We find significant differences in flu incidence by occupations. Relative to the occupation with the lowest flu incidence, "farming, fishing and forestry," the occupation with the highest flu incidence, "sales and related occupations," shows 40.5% higher odds of infection.[33] Further conditioning on our benchmark set of individual characteristics does not alter this result with an estimated coefficient of 34.2%. The difference between "farming, fishing and forestry" and the second occupation with highest flu incidence, "professional and related occupations," is 35.6%. Next, "office and administrative support" and "management, business, and financial operations" show difference in flu incidence of, respectively, 31.4% and 31.1%. This is followed by "service occupations" with a difference of 23.1% in flu incidence with respect to "farming, fishing and forestry." Last, "construction, extraction, and maintenance" and "production, transportation, and material moving operations" show a lower difference in flu incidence of, respectively, 10.3% and 9.9%. We reproduce the same exercise for years above and below the median annual aggregate flu incidence, respectively, in panel (b) and (c) of Figure 2. The main takeaway is that the differences in flu incidence across occupations is significant in both low and flu incidence years and particularly accentuated in years when the aggregate flu incidence is high.

Industry-specific flu incidence. Next, we study flu incidence by industries. We follow the classification of industries in MEPS: "natural resources", "mining", "construction", "manufacturing", "wholesale and retail trade", "transportation and utilities, "information", "financial activities", "professional and business services", "education, health, and social services", "leisure and hospitality", "other services", and "public administration". We drop military and "unclassifiable" industries. Pooling all sample years, we find that the employment share of"natural resources" represents 1.3% of the economy, "mining" 0.4%, "construction" 7.1%, "manufacturing" 11.3%, "wholesale and retail trade" 13.4%, "Transportation and utilities" 4.9%, "information" 2.4%, "financial activities" 6.7%, "professional and business services" 11.3%, "education, health and social services" 22.7%, "leisure and hospitality" 8.4%, "other services" 4.9% and "public administration" 5.2%.

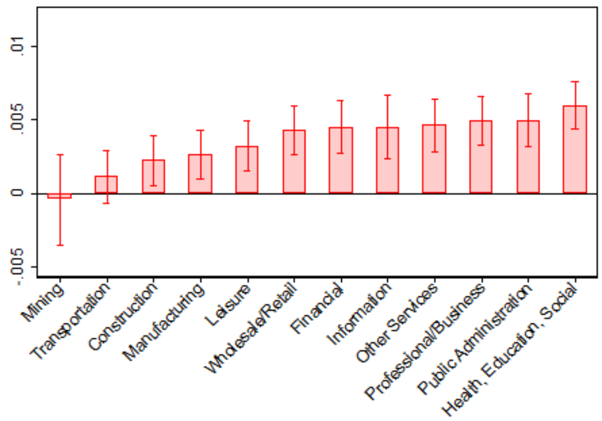
Panel (a) of Figure 3 shows the flu incidence by industry for all our sample years. Precisely, we show the β_{ind}'s from the following regression with "natural resources" as our reference industry:

$$flu_i = cons + \sum_{ind \neq 10} \beta_{ind} \mathbf{1}_{ind} + \sum_t \omega_t \mathbf{1}_t + u_{it}. \tag{4}$$

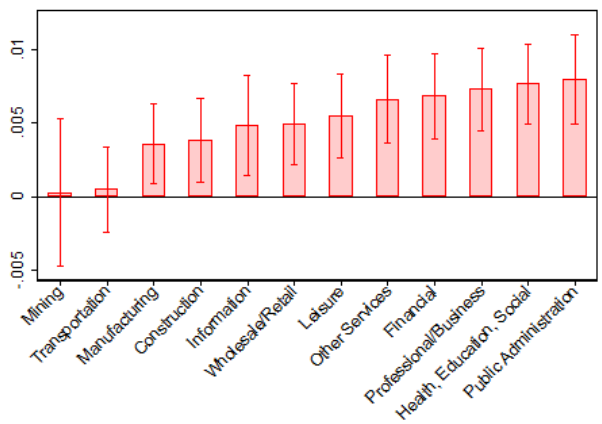
[33] In order to compute these differences in flu incidence by occupation, note that the average flu incidence in the our reference occupation, "production, transportation and material moving operations" is 0.0149. Then, based on the results of our regression the occupation with the lowest flu incidence is "farming, fishing and forestry," with a conditional flu incidence of 0.01355. For each occupation j we first compute its conditional flu incidence x_j as $0.0149 + \beta_j$ from Equation (3) and then the percentage difference with respect to the lowest flu incidence occupation as $\frac{x_j - 0.01355}{0.01355} \times 100$.

Figure 3: Industry-Specific Flu Incidence, MEPS 2002-2016

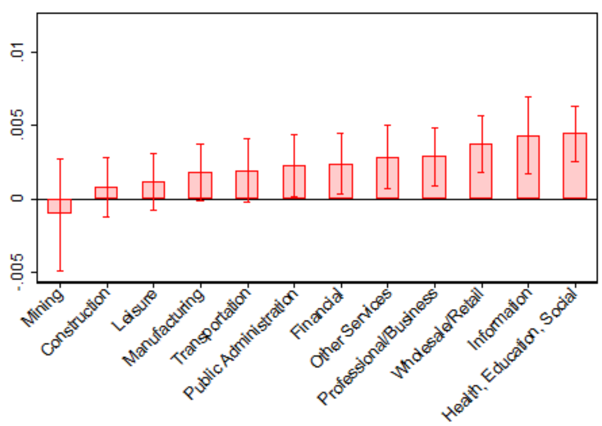
(a) Full Sample, All Years



(b) High Flu Incidence Years



(c) Low Flu Incidence Years



Notes: The industries in MEPS are categorized into twelve groups: (1) natural resources, (2) mining, (3) construction, (4) manufacturing, (5) wholesale and retail trade, (6) transportation and utilities, (7) information, (8) financial activities, (9) professional and business services, (10) education, health, and social services, (10) education, health, and social services, (11) leisure and hospitality, and (12) other services.

Relative to "mining", which is the industry with the lowest flu incidence, we find a difference in flu incidence of 52.2% for the industry with the highest flu incidence, "education, health and social services" (EHSS).[34] Further conditioning on our benchmark set of individual characteristics results in a similar coefficient of 47.7%. Focusing on the industries that follow EHSS at the top of the industry distribution in terms of flu incidence, we find that "public administration", "professional/business" and "other services" show differences in flu incidence of, respectively, 44.0%, 43.9% and 41.2%. Focusing on the industries at the bottom of the industry distribution in terms of contagion, we find that, "natural resources", "transportation" and "construction" show differences in flu incidence of, respectively, 3.5%, 12.8% and 21.7%. As it was the case for occupations, the cross-industry differences in flu incidence are steeper in high flu incidence years (see panel (b) of Figure 3) than in low flu incidence years (see panel (c) of Figure 3).

Occupation-industry specific flu incidence. A potential reason for the industry differentials in flu incidence is the within-industry occupation structure. We thus estimate how flu incidence differs by the occupation-industry mix. Precisely, we compute the occupation-industry-specific effects on flu incidence estimates, β, from the following regression specification,

$$flu_i = cons + \sum_{ind\neq 1} \sum_{occ\neq 8} \beta(\mathbf{1}_{ind} \times \mathbf{1}_{occ}) + \sum_t \omega_t \mathbf{1}_t + \sum_x \gamma x_{it} + u_{it}, \tag{5}$$

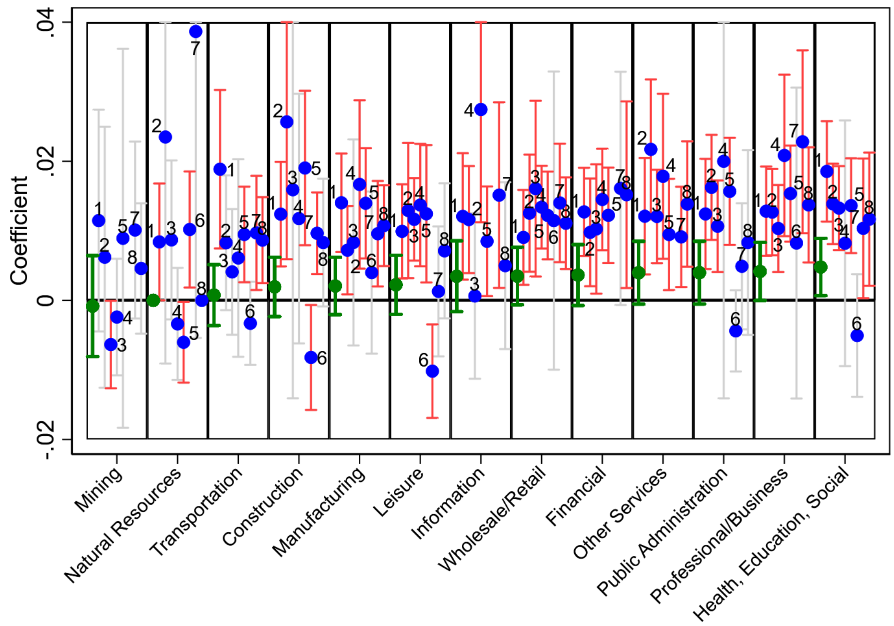
with "production, transportation and material moving operations"-"natural resources" as a reference occupation-industry pair. We include year and month fixed effects and all our benchmark individual controls as in equation (1). The estimated coefficients for each occupation-industry pair together with the 95% confidence interval (in red when statistically significant) sorted across industries and occupations are, respectively, in panels (a) and (b) of Figure 4. We further include the average effects in green by industry and occupations.

Mainly, we find that differences in industry infection rates cannot be fully accounted for by differences in the occupation structure across industries. This is due to the fact that the ranking of flu incidence by occupation changes depending on the industry. An example of such an occupation is "management, business, and financial operations," which has the largest infection rate in "education, health, and social services," but fares as the lowest in "wholesale and retail trade." Similarly, the contagion rate of "service occupations" ranks highest in "wholesale/retail trade" and lowest in "information." This non-monotonicity in flu incidence across occupations and industries makes the occupation-industry interaction important to understand contagion.

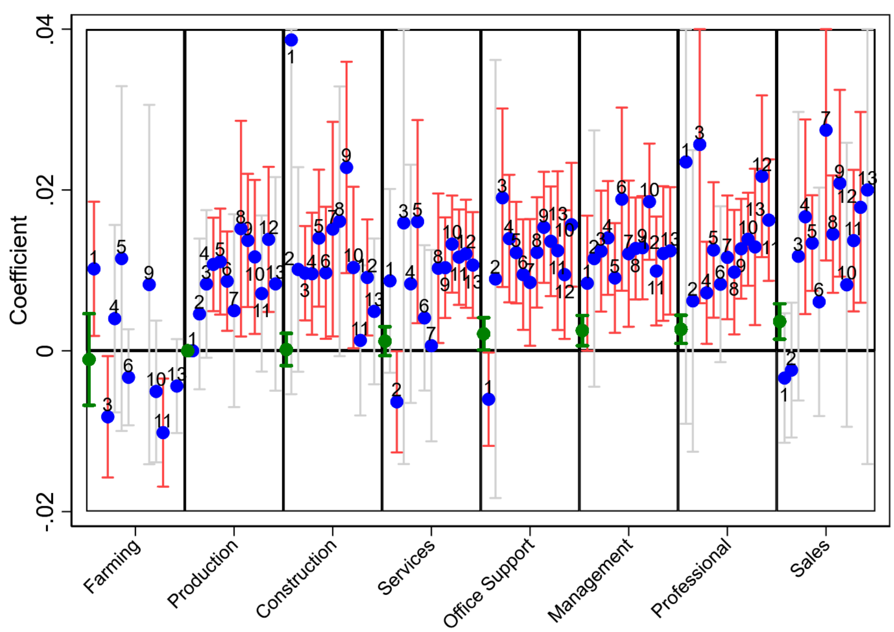
[34] Here, note that the average flu incidence in our base category, "natural resources," is 0.0126. The industry with the lowest flu incidence is "mining" with a conditional flu incidence of 0.0122. For each industry k we first compute its conditional flu incidence x_k as $0.0126 + \beta_k$ from Equation (4) and then the percentage difference with respect to the lowest flu incidence industry as $\frac{x_k - 0.0122}{0.0122} \times 100$.

Figure 4: Occupation-Industry-Specific Flu Incidence, MEPS 2002-2016

(a) Within-Industry Differences in Flu Incidence by Occupation



(b) Within-Occupation Differences in Flu Incidence by Industries



Notes: For each estimate we show the 95% confidence interval. The confidence interval is in red when the estimate is statistically significant, and in grey otherwise. In panel (a), occupations are labeled numerically as: (1) management, business, and financial operations; (2) professional and related occupations; (3) service occupations; (4) sales and related occupations; (5) office and administrative support; (6) farming, fishing and forestry; (7) construction, extraction, and maintenance; and (8) production,transportation, and material moving operations. In panel (b), industries are labeled numerically as: (1) natural resources, (2) mining, (3) construction, (4) manufacturing, (5) wholesale and retail trade, (6) transportation and utilities, (7) information, (8) financial activities, (9) professional and business services, (10) education, health, and social services, (10) education, health, and social services, (11) leisure and hospitality, and (12) other services.

Further, we note that there are some occupations that can be labeled as high or low-infection rate occupations regardless of the industry. This is, for example, the case for "professionals and related" and "sales and related" occupations which tend to be the highest occupation in terms of flu incidence across industries and often significantly, and for "farming, fishing and forestry," which is the lowest in terms of flu incidence in most industries; see panel (a) of Figure 4. At the same time, we observe that even within the high-infection occupations, e.g., professionals and sales, there is a large variation in the infection rate across industries; see panel (b) of Figure 4.

Finally, we conduct a simple decomposition exercise to assess the role of occupations in accounting for cross-industry differences in flu incidence. We focus on a comparison between the industry with the largest flu incidence, EHSS, and the rest of the economy (RoE), with, respectively, incidence levels of $flu_{EHSS} = 1.89\%$ and $flu_{RoE} = 1.66\%$. First, we impose the occupation structure of EHSS onto the rest of the economy, keeping everything else constant. We find that occupation structure explains $\frac{(flu_{RoE}|occ.\ structure_{EHSS})-flu_{RoE}}{flu_{EHSS}-flu_{RoE}} =$ 20.3% of the difference in flu incidence between EHSS and the rest of the economy. Second, we impose the probability of being infected with the flu by occupation in EHSS onto the rest of the economy, keeping everything else constant. The fact that occupations in EHSS show a different probability of being infected with the flu explains $\frac{(flu_{RoE}|flu\ by\ occ_{EHSS})-flu_{RoE}}{flu_{EHSS}-flu_{RoE}} =$ 62.6% of the differences in flu incidence between EHSS and the rest of the economy. Splitting in two the unexplained cross-industry differences, we find that differences in occupation structure explain 28.9% of the total difference in flu incidence between EHSS and the rest of the economy.[35] This implies that separately studying occupations (or industries) delivers only a partial view of the patterns of contagion. Thus, as in the previous analysis, we conclude that occupation-industry interactions are relevant to understand contagion.

4 Human Contact Interaction at Work

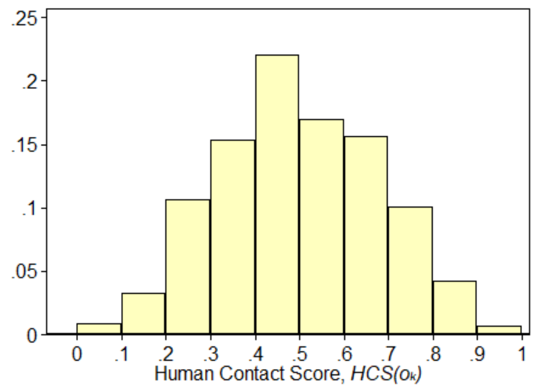
In this section, we provide direct evidence on how flu incidence differs by the extent of human contact interaction at work. We construct a measure of occupation-industry-specific human contact interaction at work in Section 4.1. We assess how flu incidence varies with human contact at work in Section 4.2.

4.1 Measuring Human Contact by Occupation and Industry

To measure occupation-industry-specific human contact interaction at work, we construct a score that summarizes human interaction at work using information on work requirements for detailed

[35] See more details of the decomposition exercise in Appendix C.

Figure 5: Human Contact Score Distribution, O'NET



Notes: The human contact score is computed by occupation as the first principal component resulting from the eigenvalue decomposition of the covariance of all 19 O'NET descriptors in Table B1. This table shows the empirical distribution of the human contact score resulting from 702 occupations (5-digit codes).

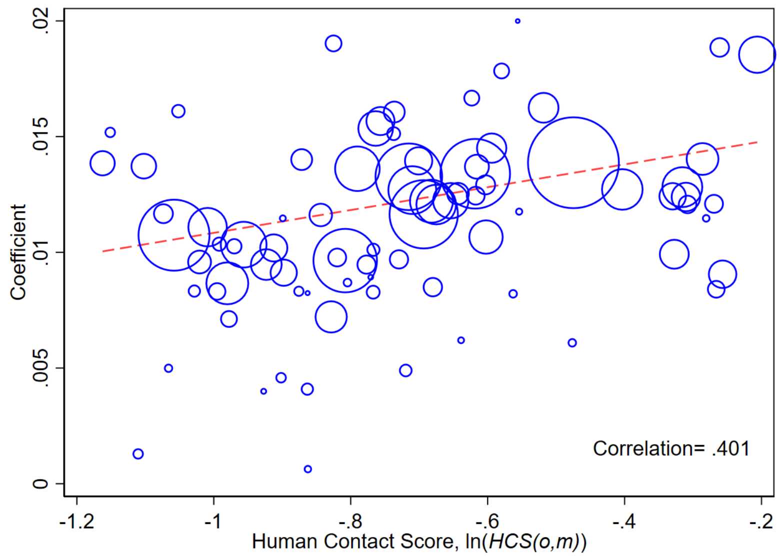
occupations from the O'NET. We use the O'NET 24.2 Database of worker and job characteristics in the United States, which includes information for up to 774 occupations (o_k) based on the Standard Occupation Classification of 2010 (SOC 2010).[36]

The O'NET database provides detailed information about work activities that capture interaction with other human beings at work as well as work context requirements. In particular, we use the set of 17 descriptors on "work activities: interaction with others" that capture activities such as, among others, "Assisting and Caring for Others", "Training and Teaching Others" and "Performing for or Working Directly with the Public," see details in Table B1 in Appendix B. In addition, we use two "work context" indicators that capture "Physical Proximity" as in Mongey et al. (2020) and "Contact with Others." In total, we have a baseline set of 19 O'NET descriptors.[37] In order to take into account these specific forms of human interaction at work, we conduct a principal component analysis through an eigenvalue decomposition of the covariance of all 19 descriptors across all occupations. The principal component analysis summarizes *all* these descriptors related to human contact interaction in one score index. We re-scale this first principal component between zero and one, i.e. $HCS(o_k) \in [0, 1]$, where 1 denotes the maximum human interaction and 0 denotes the minimum human interaction. We show the histogram of the human contact score in Figure 5. The median occupation shows a human contact score of

[36] More information on the O'NET survey questionnaires can be found in Appendix B.

[37] Details about the 19 descriptors in O'NET Work Activities and Work Contexts are shown in Table B1 in Appendix B. The occupational codes in the O'NET are based on the 6-digit SOC classification but include a 7th digit in addition. We work with the 6-digit codes by taking averages over the corresponding 7-digit occupations.

Figure 6: Occupation-Industry-Specific Flu Incidence and Human Contact at Work



Notes: The horizontal axis shows the log of the occupation-industry-specific human contact score, $HCS(o,m)$, as constructed in equation (6). The vertical axis shows the occupation-industry-specific flu incidence estimated as the occupation-industry fixed effects in equation (5). The circle size of each data point captures the associated occupation-industry employment share. We also show a linearly fitted line on the scatterplot with associated correlation of 0.401. Data from the MEPS 2002-2016, O'NET and EOS used as described in the text.

0.49 and mean of the distribution is 0.5 with a standard deviation of 0.181.[38]

Then, we construct an occupation-industry-specific human contact score using the human contact score that we built for detailed 702 O'NET occupations together with a set of occupation-industry employment shares from the "National industry-specific and by ownership" data in the Occupational Employment Statistics (OES) that are available for occupations based on the 6-digit SOC2010 (o_k)[39] and for industries on 2-digit NAICS (m_j). Since we are interested in studying how contagion correlates with human contact at work, we further aggregate the detailed O'NET-EOS occupations (o_k) up to eight occupations ($o = \{1,..,8\}$) available in MEPS and the detailed EOS industries (m_j) up to thirteen industries ($m = \{1,...,13\}$) available in MEPS (see our Section 3.2). In this manner, our occupation-industry-specific human contact score is,

$$HCS(o,m) = \sum_{o_k \in o} \sum_{m_j \in m} \psi_{o_k,m_j} HSC(o_k), \tag{6}$$

[38]The lowest one percent of the distribution shows a score of 0.13 and the highest one percent of the distribution shows a score of 0.88. The distribution is positively skewed with a coefficient of 0.08. The kurtosis coefficient is of 2.53. Shapiro-Wilk normality tests reject a normality of the distribution of our human contact score.

[39]In Appendix B we describe how we cross-walk occupations in O'NET and OES.

where we compute the detailed occupation-industry employment shares, ψ_{o_k,m_j}, using the OES.[40] Here, note that our human contact score $HCS(o,m)$ captures potential heterogeneity within occupations across industries. For example, we find that the "sales and related" occupation has a higher human contact score in the "information" industry than in the "natural resources" industry. This difference is due to the fact that a finer-level occupation with a high human contact score such as "first-line supervisors of non-retail sales workers" (which belongs to a broader "sales" occupation) has a higher employment share in "information" than in "natural resources".

As a first assessment on how contagion correlates with human contact at work, we scatterplot the estimated occupation-industry specific flu incidence from equation (5) against the occupation-industry-specific human contact score constructed in equation (6); see Figure 6. We find a positive and significant correlation (of 0.401) between human contact at work and the odds of infection.

4.2 Does Contagion Vary with Human Contact at Work?

To study whether contagion rates differ by the extent of human contact interaction at work, we use the following specification that replaces the occupation-industry fixed effects in equation (5) for the (log of) the occupation-industry-specific human contact score:

$$flu_{it} = cons + \gamma \ln(HCS) + \sum_t \omega_t \mathbf{1}_t + \sum_x \beta x_{it} + u_{it}, \tag{7}$$

where our parameter of interest, γ, captures the effects of human contact at work on the probability of flu incidence. We include period controls and our benchmark set of individual controls. Our results are in Table 5. We restrict this analysis to the employed sample—wage earners.[41]

Our main finding is that, for all specifications, the higher the human contact score, the higher is the contagion. In column (1) of Table 5, we show that, unconditionally, a one percent increase in the human contact score is associated with an increase in the probability of infection by 0.638 p.p. This result is robust to the introduction of our battery of individual controls; see column (2) of Table 5. In this case, the human contact score increases the probability of infection by 0.542 p.p. The effects of the individual observable characteristics are similar to those obtained when studying the effects of employment on the probability of infection; see Section 3.1.

[40] More details on how we construct our human contact score by using industry $\times$ occupation employment shares from OES are described in Appendix B.

[41] We focus on wage earners because the self-employed can be easily characterized by more than one occupation—e.g., a small bakery business owner is likely to be a manager that deals with providers and sales while at the same time bake. This makes the human contact score for the self-employed potentially noisy. However, at the same time, we find that our results including the self-employed deliver similar insights, see Appendix D.

Table 5: Occupation-Industry-Specific Human Contact at Work and Flu Incidence

VARIABLES	(1) Year/Month FE	(2) Benchmark	(3) High Incidence	(4) Low Incidence	(5) + N. Employees	(6) More Jobs	(7) Hours Worked
$\ln(HCS)$	0.00638***	0.00542***	0.00679***	0.00412***	0.00556***	0.00537***	0.00548***
	(0.00124)	(0.00128)	(0.00200)	(0.00157)	(0.00128)	(0.00128)	(0.00129)
N. Employees					0.000756		
					(0.000678)		
N. Emp.^{2}					-0.000235*		
					(0.000131)		
Multiple Jobs						0.00123	
						(0.00111)	
Hours							1.51e-05
							(3.20e-05)
Age		0.000226	0.000688**	-0.000189	0.000251	0.000223	0.000199
		(0.000182)	(0.000286)	(0.000223)	(0.000182)	(0.000182)	(0.000185)
Age^2		-3.79e-06*	-9.56e-06***	1.29e-06	-4.05e-06*	-3.76e-06*	-3.49e-06
		(2.16e-06)	(3.44e-06)	(2.60e-06)	(2.16e-06)	(2.16e-06)	(2.20e-06)
Female		0.000568	0.00158*	-0.000352	0.000554	0.000560	0.000656
		(0.000588)	(0.000906)	(0.000734)	(0.000588)	(0.000588)	(0.000611)
Never Married		-0.000758	-3.37e-05	-0.00142	-0.000760	-0.000775	-0.000803
		(0.000769)	(0.00123)	(0.000913)	(0.000769)	(0.000768)	(0.000773)
HH Size		-0.000195**	-0.000267*	-0.000137	-0.000199**	-0.000193**	-0.000198**
		(9.87e-05)	(0.000157)	(0.000122)	(9.87e-05)	(9.86e-05)	(9.96e-05)
N. Other Conditions		0.00168***	0.00118***	0.00219***	0.00169***	0.00168***	0.00171***
		(0.000259)	(0.000400)	(0.000322)	(0.000259)	(0.000259)	(0.000261)
Student Status		0.00203	0.00360	0.000494	0.00199	0.00200	0.00222
		(0.00168)	(0.00270)	(0.00199)	(0.00167)	(0.00168)	(0.00171)
Constant	0.0492***	0.0447***	0.0396***	0.0256***	0.0445***	0.0446***	0.0447***
	(0.00283)	(0.00479)	(0.00679)	(0.00493)	(0.00480)	(0.00479)	(0.00489)
Observations	1,838,330	1,838,330	923,710	914,620	1,838,330	1,838,330	1,820,581
R-squared	0.005	0.005	0.005	0.003	0.005	0.005	0.005
Year/Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pseudo R2	0.00480	0.00512	0.00484	0.00284	0.00516	0.00512	0.00513

Notes: This table shows results from a linear probability model for flu incidence on MEPS data 2002-2016. Robust standard errors clustered at the individual level are in parentheses *** p<0.01, ** p<0.05, * p<0.1. We focus on wage earners.

We conduct some robustness exercises. First, we split the sample into years of high aggregate flu incidence (above median) and low aggregate flu incidence (below median); see, respectively, column (3) and (4) in Table 5. A higher human contact score is associated with a higher probability of infection, 0.679 p.p., in years of high aggregate incidence, whereas this figure is reduced to 0.412 p.p. in years of low aggregate incidence. Second, our results are robust to firm size. A potentially relevant aspect for contagion at work is the size of the firm by the number of employees. Our reasoning is that larger firms can facilitate more human contact interaction compared with smaller firms. This would imply that the virus can spread more easily within large firms than within small firms. When we control for the number of employees we find that higher human contact at work is associated with a higher probability of infection of 0.556 p.p.; see column (5) of Table 5. Further, the number of employees increases the probability of infection in a concave fashion. Third, we control for individuals holding more than one job and for weekly hours worked conditional on working. Our main finding remains: a higher human contact at work is associated with a higher probability of infection of 0.537 p.p. and 0.548 p.p. when controlling for multiple jobs and hours worked, respectively; see columns (6) and (7) of Table 5. The coefficients on holding more jobs and hours worked are positive, yet not significant.

5 Conclusion

In this paper we exploit a rare opportunity to link occupations and industries to flu incidence at the individual level using a nationally representative panel survey. With these data, we study how contagion differs by employment status. We find that the employed are on average 35.3% more likely to be infected with the flu virus. Further, within the employed, we find that contagion differs by occupation, industry and occupation-industry mix. Importantly, differences in the within-industry occupation structure only partially explain cross-industry differences in flu incidence which makes the interaction between occupations and industries relevant to understand contagion. Finally, we construct a measure of occupation-industry-specific exposure to human contact interaction at work and find that higher human contact at work is positively associated with higher contagion rates. We hope that our results are relevant for the understanding of the spread of the flu as well as of other infectious diseases that are transmitted via droplets, fomites or close human contact interaction (e.g., SARS and Covid-19). The fact that contagion risk is heterogeneous across occupations and industries opens the door for the assessment of optimal nonpharmaceutical policies against a pandemic (e.g. stay-home policies, economic lockdowns or mandated sick leaves) that are potentially shaped by occupation-industry-specific contagion risk.

References

Adams-Prassl, A., Boneva, T., Golin, M., and Rauh, C. (2020). Inequality in the Impact of the Coronavirus Shock: Evidence from Real Time Surveys. *Journal of Public Economics*, 189:104245.

Adda, J. (2016). Economic Activity and the Spread of Viral Diseases: Evidence from High Frequency Data. *Quarterly Journal of Economics*, 131(2):891–941.

Aleman, C., Busch, C., Ludwig, A., and Santaeulàlia-Llopis, R. (2020a). Evaluating the Effectiveness of Policies Against a Pandemic. Working Paper 1209, Barcelona GSE.

Aleman, C., Iorio, D., and Santaeulàlia-Llopis, R. (2020b). A Quantitative Theory of the HIV Epidemic: Education, Risky Sex and Asymmetric Learning. Working paper, Barcelona GSE.

Almagro, M. and Orane-Hutchinson, A. (2020). The Determinants of the Differential Exposure to COVID-19 in New York City and Their Evolution over Time. *Journal of Urban Economics*, page 103293.

Alvarez, F. E., Argente, D., and Lippi, F. (2020). A Simple Planning Problem for COVID-19 Lockdown. NBER Working Papers 26981, National Bureau of Economic Research, Inc.

Anderson, N. J., Bonauto, D. K., Fan, Z. J., and Spector, J. T. (2012). Distribution of Influenza-Like Illness (ILI) by Occupation in Washington State, September 2009-August 2010. *PLoS ONE*, 7(11).

Aspri, A., Beretta, E., Gandolfi, A., and Wasmer, E. (2021). Mortality Containment vs. Economics Opening: Optimal Policies in a SEIARD Model. *Journal of Mathematical Economics*.

Atkeson, A. (2020). What Will Be the Economic Impact of COVID-19 in the US? Rough Estimates of Disease Scenarios. Working Paper 26867, National Bureau of Economic Research.

Azzimonti, M., Fogli, A., Perri, F., and Ponder, M. (2020). Pandemic Control in ECON-EPI Networks. Working Paper 27741, National Bureau of Economic Research.

Barro, R. J., Ursúa, J. F., and Weng, J. (2020). The Coronavirus and the Great Influenza Pandemic: Lessons from the "Spanish Flu" for the Coronavirus's Potential Effects on Mortality and Economic Activity. Working Paper 26866, National Bureau of Economic Research.

Belongia, E. A., Simpson, M. D., King, J. P., Sundaram, M. E., Kelley, N. S., Osterholm, M. T., and McLean, H. Q. (2016). Variable Influenza Vaccine Effectiveness by Subtype: A Systematic Review and Meta-analysis of Test-Negative Design Studies. *Lancet*, 8:942–951.

Berger, D., Herkenhoff, K., and Mongey, S. (2020). An SEIR Infectious Disease Model with Testing and Conditional Quarantine. Working paper, BFI.

Bognanni, M., Hanley, D., Kolliner, D., and Mitman, K. (2020). Economics and Epidemics: Evidence from an Estimated Spatial Econ-SIR Model. IZA Working Paper No. 13797.

Cajner, T., Crane, L. D., Decker, R. A., Grigsby, J., Hamins-Puertolas, A., Hurst, E., Kurz, C., and Yildirmaz, A. (2020). The US Labor Market during the Beginning of the Pandemic Recession. Technical report, National Bureau of Economic Research.

Carrillo-Tudela, C. and Visschers, L. (2020). Unemployment and Endogenous Reallocation over the Business Cycle. IZA Working Paper No. 13307.

Casares, M., Gomme, P., and Khan, H. (2020). COVID-19 Pandemic and Economic Scenarios for Ontario. Working paper, Forthcoming, Canadian Journal of Economics.

Chetty, R., Friedman, J. N., Hendren, N., and Stepner, M. (2020). The Economic Impacts of COVID-19: Evidence from a New Public Database Built Using Private Sector Data. NBER Working Paper No. 27431.

Darougheh, S. (2019). Occupation-industry Mismatch in the Cross-section and the Aggregate. Danmarks Nationalbank Working Papers 147, Copenhagen.

Edwards, C. H., Tomba, G. S., and De Blasio, B. F. (2016). Influenza in Workplaces: Transmission, Workers' Adherence to Sick Leave Advice and European Sick Leave Recommendations. *European Journal of Public Health*, 26(3):478–485.

Eichenbaum, M. S., Rebelo, S., and Trabandt, M. (2020). The Macroeconomics of Epidemics. NBER Working Papers 26882, National Bureau of Economic Research, Inc.

Eyméoud, J.-B., Petrosky-Nadeau, N., Santaeulàlia-Llopis, R., and Wasmer, E. (2021). Contrasting US and EU Labor Markets during COVID-19. *FRBSF Economic Letter*.

Fajgelbaum, P. D., Khandelwal, A., Kim, W., Mantovani, C., and Schaal, E. (2020). Optimal Lockdown in a Commuting Network. Working paper, Barcelona GSE.

Farboodi, M., Jarosch, G., and Shimer, R. (2020). Internal and External Effects of Social Distancing in a Pandemic. Working Paper 27059, National Bureau of Economic Research.

Ferguson, N. M., Cummings, D. A., Cauchemez, S., Fraser, C., Riley, S., Meeyai, A., Iamsirithaworn, S., and Burke, D. S. (2005). Strategies for Containing an Emerging Influenza Pandemic in Southeast Asia. *Nature*, 437(209–214).

Garibaldi, P., Moen, E. R., and Pissarides, C. A. (2020). Modelling Contacts and Transitions in the SIR Epidemics Model. *Covid Economics: Vetted and Real-Time Papers*, 5:1–20.

Glover, A., Heathcote, J., Krueger, D., and Ríos-Rull, J.-V. (2020). Health versus Wealth: On the Distributional Effects of Controlling a Pandemic. Working Paper 27046, National Bureau of Economic Research.

Halloran, M. E., Ferguson, N. M., Eubank, S., Longini, I. M., Cummings, D. A., Lewis, B., Xu, S., Fraser, C., Vullikanti, A., Germann, T. C., Wagener, D., Beckman, R., Kadau, K., Barrett, C., Macken, C. A., Burke, D. S., and Cooley, P. (2008). Modeling Targeted Layered Containment of an Influenza Pandemic in the United States. *Proc Natl Acad Sci U S A*, 105(12):4639–4644.

Iorio, D. and Santaeulàlia-Llopis, R. (2016). Education, HIV Status, and Risky Sexual Behavior: How Much Does the Stage of the HIV Epidemic Matter? Working paper, Barcelona GSE.

Kaplan, G., Moll, B., and Violante, G. (2020). The Great Lockdown and the Big Stimulus: Tracing the Pandemic Possibility Frontier for the U.S. Working paper, National Bureau of Economic Research.

Klepac, P., Kucharski, A. J., Conlan, A. J., Kissler, S., Tang, M. L., Fry, H., and Gog, J. R. (2020). Contacts in Context: Large-Scale Setting-Specific Social Mixing Matrices from the BBC Pandemic Project. *medRxiv*.

Luckhaupt, S. E., Calvert, G. M., Li, J., Sweeney, M., Santibanez, T. A., and Centers for Disease Control and Prevention (CDC) (2014). Prevalence of Influenza-like Illness and Seasonal and Pandemic H1N1 Influenza Vaccination Coverage among Workers–United States, 2009-10 Influenza Season. *MMWR. Morbidity and mortality weekly report*, 63(10):217–21.

Mongey, S., Pilossoph, L., and Weinberg, A. (2020). Which Workers Bear the Burden of Social Distancing Policies? Working Paper 27085, National Bureau of Economic Research.

Mossong, J., Hens, N., Jit, M., Beutels, P., Auranen, K., Mikolajczyk, R., Massari, M., Salmaso, S., Tomba, G. S., Wallinga, J., Heijne, J., Sadkowska-Todys, M., Rosinska, M., and Edmunds, W. J. (2008). Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases. *PLOS Medicine*, 5(3):1–1.

Obiols-Homs, F. (2020). Precaution, Social Distancing and Tests in a Model of Epidemic Disease. Working paper, Barcelona GSE.

Østergaard, L., Mortensen, R. N., Kragholm, K., Dalager-Pedersen, M., Koch, K., Køber, L., Torp-Pedersen, C., and Fosbøl, E. (2021). Work Exposure and Associated Risk of Hospitalisation with Pneumonia and Influenza: A Nationwide Study. *Scandinavian Journal of Public Health*, 49(1):57–63.

Pichler, S., Wen, K., and Ziebarth, N. R. (2020). COVID-19 Emergency Sick Leave Has Helped Flatten The Curve In The United States. *Health Affairs*, 39(12):2197–2204. PMID: 33058691.

Pichler, S. and Ziebarth, N. R. (2017). The Pros and Cons of Sick Pay Schemes: Testing for Contagious Presenteeism and Noncontagious Absenteeism Behavior. *Journal of Public Economics*, 156:14–33.

Piguillem, F. and Shi, L. (2020). Optimal COVID-19 Quarantine and Testing Policies. EIEF Working Papers Series.

Pollan, M., Perez-Gomez, B., Pastor-Barriuso, R., Oteo, J., Hernan, M. A., Perez-Olmeda, M., Sanmartin, J. L., Fernandez-Garcia, A., Cruz, I., de Larrea, N. F., Molina, M., Rodriguez-Cabrera, F., Martin, M., Merino-Amador, P., Paniagua, J. L., Munoz-Montalvo, J. F., Blanco, F., and Yotti, R. (2020). Prevalence of SARS-CoV-2 in Spain (ENE-COVID): a Nationwide, Population-based Seroepidemiological Study. *The Lancet*, (10250).

Potter, G. E., Smieszek, T., and Sailer, K. (2015). Modeling Workplace Contact Networks: The Effects of Organizational Structure, Architecture, and Reporting Errors on Epidemic Predictions. *Network Science*, 3(3):298–325.

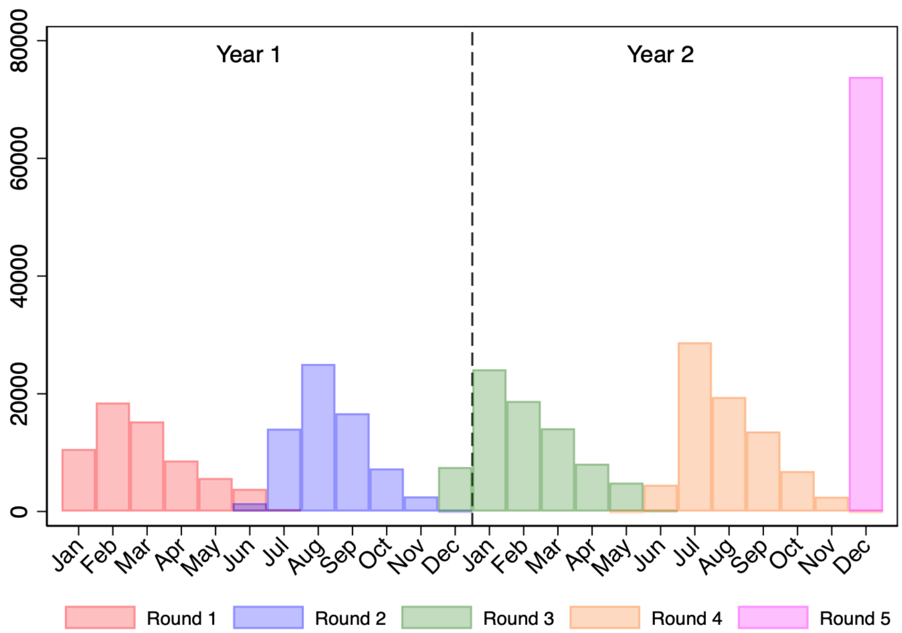
Vahtera, J., Virtanen, P., Kivimäki, M., and Pentti, J. (1999). Workplace as an Origin of Health Inequalities. *Journal of Epidemiology and Community Health*, 53(7):399–407.

Zhai, Y., Santibanez, T. A., Kahn, K. E., Black, C. L., and de Perio, M. A. (2018). Paid Sick Leave Benefits, Influenza Vaccination, and Taking Sick Days due to Influenza-like Illness among U.S. Workers. *Vaccine*, 19(36(48)):7316–7323.

A The MEPS Panel Rounds

Here we describe the MEPS panel round design by interview month. The MEPS consists of a set of unbalanced interview rounds that collect individual panel information (for a total of 5 rounds) from January of a year t to December of a year $t+1$. That is, the MEPS provides information for each interviewed individual for a total of two consecutive calendar years.

Figure A1: MEPS Frequency of Monthly Interviews: By Panel Rounds, All Years



Notes: Frequency of monthly interviews per round for our entire MEPS Sample, 1996-2016.

The month of the interview differs across individuals within each round.[42] In Figure A1, we show the actual histogram of interview months (per round) in our MEPS sample. In the first interview round, individuals are asked about their labor market and health status covering the period from January in year t to the date of the first round interview. Note that the month of the interview in this round 1 ranges (across individuals) from January of year t to July of year t. We find that the median month of interview of the first round is March of year t. In the second interview round, individuals are asked about the period between the interview month of the first round and the interview month of the second round. The interview month of second round ranges from June of year t to December of year t and the median interview month in this round 2 is August. Note that given the heterogeneity across individuals in the month of the interview of the first and second rounds, the period covered in round 2—that goes from the interview month of round 1 to the interview month of round 2—differs by individual. The same occurs for all other rounds. In round 3, individuals are asked about the period between the interview month of round 2 and the interview month of round 3. The interview month of round 3 ranges from December

[42]Unfortunately, we do not have information about re-infections. If in a given year τ an individual positively reports the flu in a given round $\bar{r}$, then this individual is not asked about the flu in all future rounds $r > \bar{r}$ within the same year τ.

of year t to June of year $t+1$ and the median interview month is February of year $t+1$. In round 4, individuals are asked about the period between the interview month of round 3 and the interview month of round 4. The interview month of round 4 ranges from June of year $t+1$ to December of year $t+1$ and the median interview month of the fourth round is August of year $t+1$. Last, in round 5, most of interviews occur in December of year $t+1$ covering the entire period between the interview month of the fourth round and that December in $t+1$.

B O'NET and OES

B.1 Human Contact Score from O'NET

The O'NET questionnaires include questions about the importance of different abilities required at specific occupations. We are particularly interested in the O'NET database that regards with "Work Activities" and includes several indicators for the importance of "Interacting with Others", summarized in Table B1, as well as two of the "Work Context" indicators, "Physical Proximity" and "Contact with Others". Table B1 shows a total of 19 different types of human interaction conducted at work we use in our analysis.[43] The answers to these questions are recorded from each respondent choosing a number between (and including) 1 and 5 (1 - Not important at all, 2 - Somewhat important, 3 - Important, 4 - Very important, 5 - Extremely important). These answers are more specific for the 'Work Context" indicators, "Contact with Others" (1 - No contact with others, 2 - Occasional contact with others, 3 - Contact with others about half the time, 4 - Contact with others most of the time, 5 - Constant contact with others) and "Physical Proximity" (1 - I don't work near other people (beyond 100 ft.), 2 - I work with others but not closely (e.g., private office), 3 - Slightly close (e.g., shared office), 4 - Moderately close (at arm's length), 5 - Very close (near touching)).

We show the distribution of each of the specific forms of human interaction in Figure B2. Each distribution is ordered with an importance scale that ranges from 1 (lowest) to 5 (highest). We normalize these categories to be between zero and one by subtracting one from the importance scale and dividing by four. We find substantial differences across forms of human interaction. The median normalized importance scale is lowest for work activities related to "Staffing Organizational Units" (0.24), "Selling or Influencing Others" (0.33), "Monitoring and Controlling Resources" (0.38) and "Performing Administrative Activities" (0.43). The median normalized importance scale is highest for "Contact with Others" (0.86), "Communicating with Supervisors, Peers or Subordinates" (0.75), "Establishing and Maintaining Interpersonal Relationships" (0.66) and "Contact with Others" (0.59).[44] Finally, in Figure B3 we show the correlation of each of these descriptors with the occupation-specific

[43] There are up to 774 occupations in O'NET, but not all of them have values for all the descriptors that we use. This leaves our sample of occupations in a total of 702.

[44] These differences also arise for inequality statistics such as the standard deviation that is lowest for "Contact with Others" (0.115), " Communicating with Supervisors, Peers, or Subordinates" (0.1152), "Establishing and Maintaining Interpersonal Relationships" (0.135) and "Coordinating the Work and Activities of Others" (0.131) and highest for "Selling or Influencing Others" (0.174) "Communicating with Persons Outside Organization" (0.189) "Assisting and Caring for Others" (0.204) "Performing for or Working Directly with Public" (0.239). These differences also arise in terms of the extent of skewness. The specific forms of human contact are typically positively skewed, except for four clearly negatively skewed activities "Communicating with Supervisors, Peers, or Subordinates" (-0.95), "Contact with Others" (-0.82), "Communicating with Persons Outside Organization" (-0.266) and "Establishing and Maintaining Interpersonal Relationships" (-.149). The degree of positive skewness varies from relatively small coefficient values for "Performing for or Working Directly with Public" (0.010) to relatively high values for "Staffing Organizational Units" (0.95). We further conducted some Shapiro-Wilk normality

human contact score. We find these correlations to be positive and signficant for each and all the descriptors.

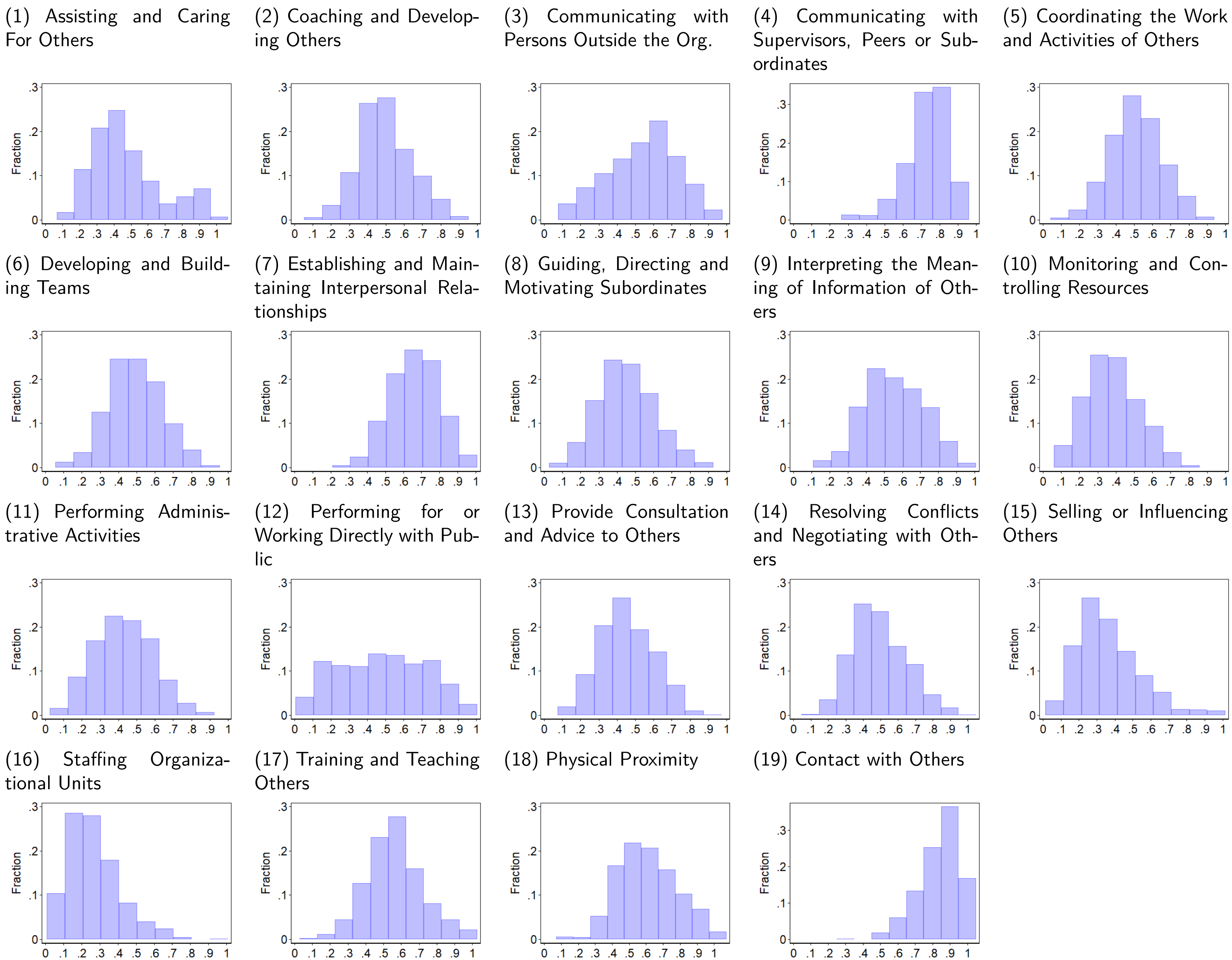
tests. We can reject normality for all our specific forms of human interaction except for "Training and Teaching Others", "Developing and Building Teams" and "Performing Administrative Activities."

Table B1: Work Activities: Interaction with Others, and Work Context

Name	
Work Activities: Interaction with Others	
1. Assisting and Caring for Others	Providing personal assistance, medical attention, emotional support, or other personal care to others such as coworkers, customers, or patients.
2. Coaching and Developing Others	Identifying the developmental needs of others and coaching, mentoring, or otherwise helping others to improve their knowledge or skills.
3. Communicating with Persons Outside Organization	Communicating with people outside the organization, representing the organization to customers, the public, government, and other external sources. This information can be exchanged in person, in writing, or by telephone or e-mail.
4. Communicating with Supervisors, Peers, or Subordinates	Providing information to supervisors, co-workers, and subordinates by telephone, in written form, e-mail, or in person.
5. Coordinating the Work and Activities of Others	Getting members of a group to work together to accomplish tasks.
6. Developing and Building Teams	Encouraging and building mutual trust, respect, and cooperation among team members.
7. Establishing and Maintaining Interpersonal Relationships	Developing constructive and cooperative working relationships with others, and maintaining them over time.
8. Guiding, Directing, and Motivating Subordinates	Providing guidance and direction to subordinates, including setting performance standards and monitoring performance.
9. Interpreting the Meaning of Information for Others	Translating or explaining what information means and how it can be used.
10. Monitoring and Controlling Resources	Monitoring and controlling resources and overseeing the spending of money.
11. Performing Administrative Activities	Performing day-to-day administrative tasks such as maintaining information files and processing paperwork.
12. Performing for or Working Directly with the Public	Performing for people or dealing directly with the public. This includes serving customers in restaurants and stores, and receiving clients or guests.
13. Provide Consultation and Advice to Others	Providing guidance and expert advice to management or other groups on technical, systems-, or process-related topics.
14. Resolving Conflicts and Negotiating with Others	Handling complaints, settling disputes, and resolving grievances and conflicts, or otherwise negotiating with others.
15. Selling or Influencing Others	Convincing others to buy merchandise/goods or to otherwise change their minds or actions.
16. Staffing Organizational Units	Recruiting, interviewing, selecting, hiring, and promoting employees in an organization.
17. Training and Teaching Others	Identifying the educational needs of others, developing formal educational or training programs or classes, and teaching or instructing others.
Work Context	
1. Physical Proximity	To what extent does this job require the worker to perform job tasks in close physical proximity to other people?
2. Contact with Others	How much does this job require the worker to be in contact with others (face-to-face, by telephone, or otherwise) in order to perform it?

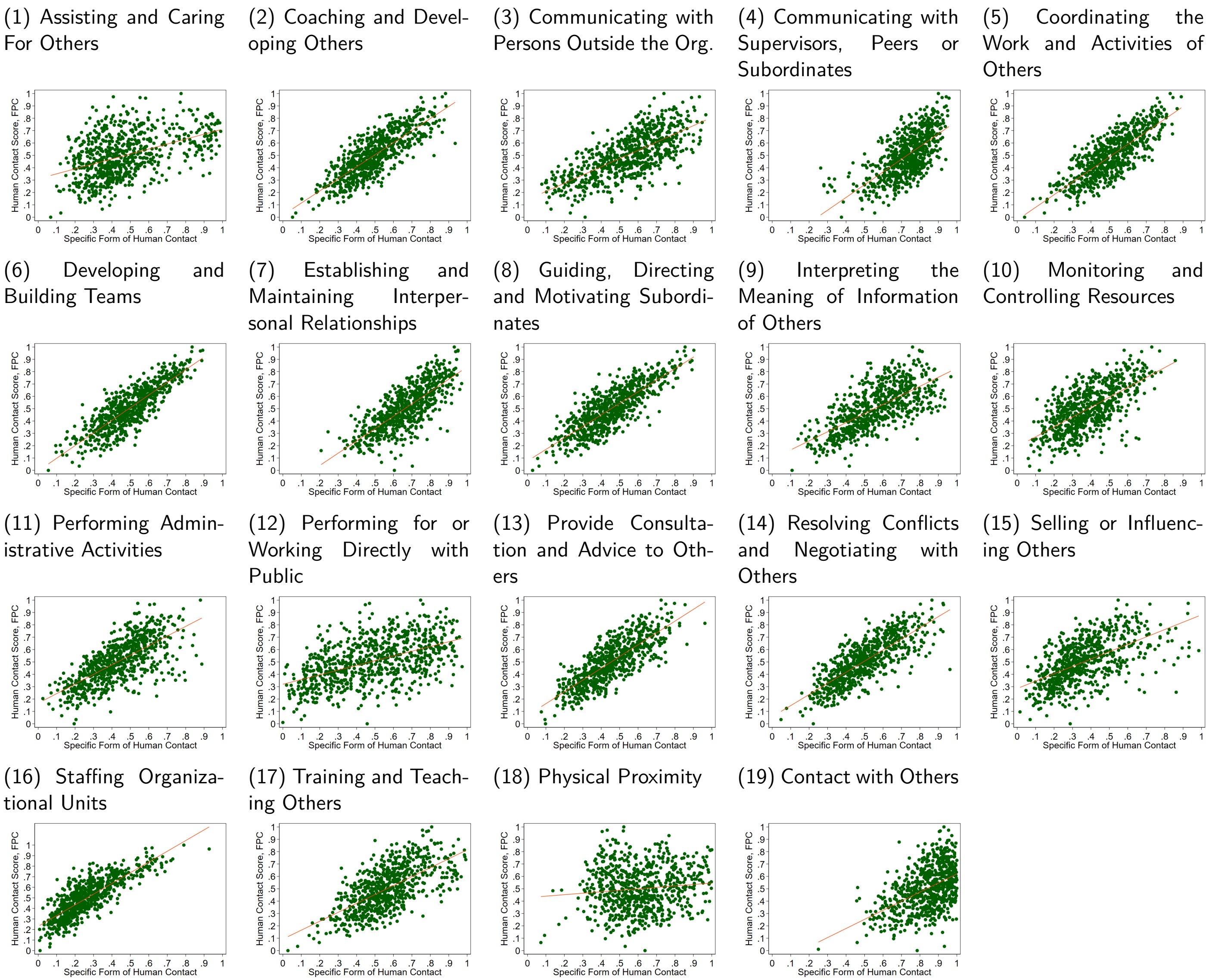
Note: This table shows the descriptors included in the O'NET on Work Activities: Interacting with Others and on Work Context.

Figure B2: Density (Fraction) of Specific Forms of Human Contact



Notes: This figure shows the distribution(fraction) of the importance of each of the specific forms of human interaction by occupation in the O'NET, based on the SOC 5-digit codes.

Figure B3: Relation between Human Contact Score and Specific Forms of Human Contact



Notes: This figure shows the correlation between each of the specific forms of human contact and the human contact score (first principal component) based on the O'NET data.

B.2 Merging Human Contact Scores of O'NET to MEPS occupations

To create Human Contact (HC) scores for occupation-industry cells in the MEPS from the Human Contact score across detailed occupations in the O'NET, we use data on the employment shares of detailed occupations within industries from the Occupational Employment Statistics (OES) between 2004 and 2018. Our final occupation-industry-specific HC score is a weighted average of the HC scored of the detailed occupations from O'NET, where the weights are the employment shares in the corresponding MEPS occupation-industry cell.

This involves the following steps:

1. **Crosswalk between the NAICS industry codes in the OES and the MEPS industry codes based on the census classification:** The industry codes in the OES are based on the 2-digit NAICS classification while those in the MEPS are wide industry groups based on the 4-digit census codes.[45] There is a direct correspondence between the two classifications such that the 20 NAICS codes can be aggregated to the 13 MEPS industry codes.
2. **Merge our HC scores from O'NET to the OES 2010-2018 detailed occupations based on the SOC 2010 codes:** Turning to the occupations, the OES occupations data is based on the SOC 2000 before 2010 and SOC 2010 after. Since O'NET is also based on SOC 2010 codes, we first merge our O'NET HC score to OES data between 2010 and 2018. There are some occupational categories in the OES with no corresponding code in the O'NET: these are denoted as 'All other' within more specific occupational groups (an example is the OES SOC 2010 code 47-4099 for "Construction and Related Workers, All Other"). Since total employment is large for several of these 'All Other' categories, we attach to these the weighted average of the HC scores of the other occupations within the same wide occupation group where the weights are total employment in each of those occupations (in this example this would be the weighted average of the HC scores for occupations within the wide occupation group 47-4000 "Other Construction and Related Workers").
3. **Harmonize 2004-2010 and 2010-2018 OES data by using a crosswalk between SOC 2000 and SOC 2010 codes:** Next, we harmonize occupational data between the 2004-2010 and 2010-2018 OES data.[46] To create a unique occupational classification for the OES data we use the SOC2000-SOC2010 crosswalk available from the BLS (https://www.bls.gov/soc/soccrosswalks.htm). For those SOC 2000 which become a unique category in 2010, we simply aggregate total employment to the 2010 SOC category. However, it occurs more frequently that a single 2000 SOC category is split into several 2010 categories. In this case we condense the SOC 2010 category to the SOC 2000 category and its HC score is the weighted average of the HC scores of the SOC 2010 categories using the occupational employment of each of the SOC2010 categories as weights.
4. **Crosswalk between SOC 2010 and the occupational groups in the MEPS based on the census classification:** Having a HC score for OES data between 2004 and 2018, we then use a crosswalk between the SOC codes in the OES and the census codes together with the condensing rules of the MEPS wide

[45] The condensing rules for industry codes in the MEPS can be found at https://meps.ahrq.gov/data_stats/download_data/pufs/ind_occ/ind3.shtml

[46] This is especially important for having information on the occupation structure of the industry "Public Administration", which is only included in the OES in years 2004, 2005 and 2007

occupations[47].

5. **Create an HC score at the MEPS occupation$\times$industry level:** Finally, we construct our occupation-industry-specific HC scores for the MEPS as a weighted average of the HC scores of detailed occupations within the MEPS wide occupation-industry cells, weighing by the employment shares of detailed occupations in the corresponding occupation-industry cell from the OES.

[47] The condensing rules can be found at https://meps.ahrq.gov/data_stats/download_data/pufs/ind_occ/occ3.shtml.

C How Much Cross-Industry Flu Contagion is Explained by Occupations? A Decomposition Exercise

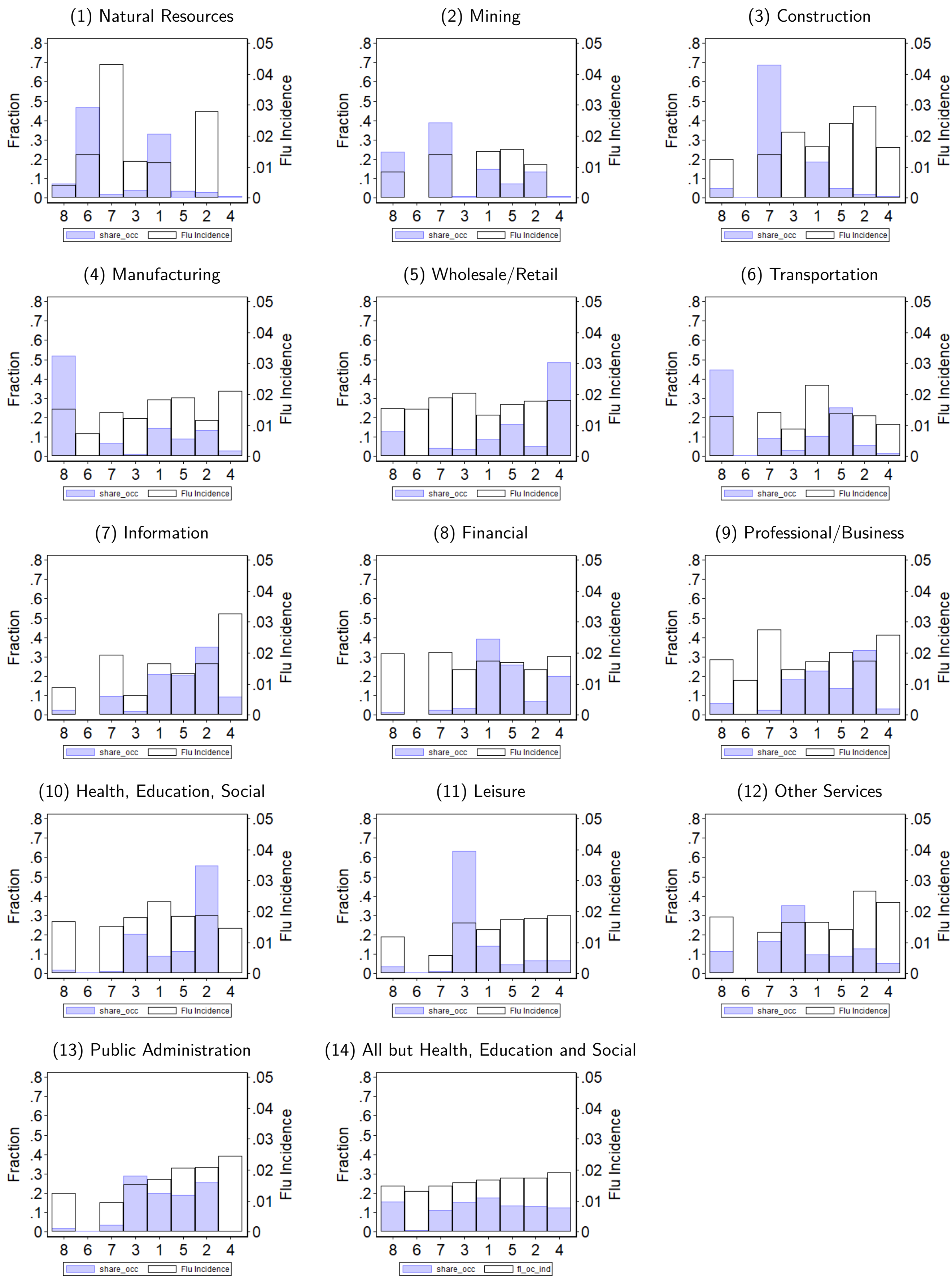
A potential reason for the industry differentials in flu incidence is the within-industry occupation structure. In Figure C4, we show the fraction of employed individuals by occupation separately within each industry. There are substantial differences across industries in both: (1) the occupational structure and (2) in the probability of being infected with the flu by occupation.

Comparing the industry with the largest flu incidence—"education, health, and social services"—and the rest of the economy we find that "education, health, and social services" have a disproportionately larger fraction of their employed in occupations that are relatively more subject to flu contagion risk. Precisely, risky occupations such as those related to management, professionals, and office support account for, respectively, 8.9%, 55.8%, and 20.5% of the total employment in "education, health, and social services," whereas these figures are 17.5%, 13.3% and 13.6% for the rest of the economy. Indeed, in the rest of the economy less risky occupations are more predominant, except for sales, the riskiest occupation, that accounts for 12.6% of total employment in the rest of the economy and for 0.3% in "education, health, and social services." The probability of infection by occupation also differs by industries. In the industry with the largest flu incidence, "education, health, and social services," we find that the probability of infection is larger in all occupations except in sales compared with the rest of the economy.[48]

In this context, we conduct a simple decomposition exercise to assess the sources of flu incidence differences across industries. We focus on a comparison between the industry with the largest flu incidence—"education, health, and social services" (EHSS)—and the rest of the economy (RoE) with, respectively, incidence levels of $flu_{EHSS} = 1.89\%$ and $flu_{RoE} = 1.66\%$. The differential incidence can be driven by the fact that "education, health, and social services" has a different occupation structure than the rest of the economy, or by the fact that the probability of being infected with the flu by occupation in "education, health, and social services" differs from the rest of the economy. To address this question, we first impose the occupation structure of "education, health, and social services" onto the other industries, keeping everything else constant. We find that occupation structure explains $\frac{(flu_{RoE}|occ.\ structure_{EHSS})-flu_{RoE}}{flu_{EHSS}-flu_{RoE}} =$ 20.3% of the difference in flu incidence between "education, health, and social services" and the rest of the economy. Second, we impose the probably of being infected with the flu by occupation in "education, health, and social services" onto the other industries, keeping everything else constant. We find that the fact that occupations in "education, health, and social services" show a different probability of being infected with the flu explains $\frac{(flu_{RoE}|flu\ by\ occ_{EHSS})-flu_{RoE}}{flu_{EHSS}-flu_{RoE}} =$ 62.6% of the differences in flu incidence between "education, health, and social services" and the rest of the economy. Splitting in two the joint effect of the occupation structure and the flu incidence by occupation, we find that differences in occupation structure and the differences in the probability of being infected with the flu by occupation explain, respectively, 28.9% and 71.1% of the total difference in flu incidence between "education, health, and social services" and the rest of the economy.

[48] We cannot make a comparison of flu incidence for the farming occupation since it is not present in the "education, health, and social services" industry.

Figure C4: Flu and Occupations within Industries



Notes: The horizontal axis in all panels displays the occupations with the following labels: (1) management, business, and financial operations; (2) professional and related occupations; (3) service occupations; (4) sales and related occupations; (5) office and administrative support; (6) farming, fishing and forestry; (7) construction, extraction, and maintenance; and (8) production, transportation, and material moving operations. The left vertical axis shows the fraction of individuals by occupation in a given industry. The right vertical axis shows the flu incidence by occupation in a given industry.

D Occupation-Industry-Specific Human Contact at Work and Flu Incidence, Including Self-Employed

Table D1: Occupation-Industry-Specific Human Contact at Work and Flu Incidence

VARIABLES	(1) Full Sample	(2) + Controls	(3) High Incidence	(4) Low Incidence	(5) + N. Employees	(6) More Jobs	(7) Hours Worked
$\ln(HCS)$	0.00524***	0.00448***	0.00598***	0.00305**	0.00458***	0.00445***	0.00548***
	(0.00113)	(0.00116)	(0.00181)	(0.00143)	(0.00117)	(0.00116)	(0.00129)
N. Employees					0.00104		
					(0.000641)		
N. Emp.^{2}					-0.000276**		
					(0.000126)		
Multiple Jobs						0.00102	
						(0.00100)	
Hours							1.51e-05
							(3.20e-05)
Age		0.000322*	0.000806***	-0.000114	0.000339*	0.000320*	0.000199
		(0.000172)	(0.000265)	(0.000219)	(0.000173)	(0.000172)	(0.000185)
Age^2		-5.08e-06**	-1.10e-05***	1.34e-07	-5.26e-06***	-5.05e-06**	-3.49e-06
		(2.02e-06)	(3.16e-06)	(2.53e-06)	(2.03e-06)	(2.02e-06)	(2.20e-06)
Female		0.000684	0.00148*	-3.10e-05	0.000681	0.000677	0.000656
		(0.000543)	(0.000836)	(0.000678)	(0.000544)	(0.000543)	(0.000611)
Never Married		-0.000455	0.000520	-0.00134	-0.000452	-0.000470	-0.000803
		(0.000730)	(0.00116)	(0.000879)	(0.000729)	(0.000729)	(0.000773)
HH Size		-0.000187**	-0.000267*	-0.000122	-0.000191**	-0.000185**	-0.000198*
		(9.24e-05)	(0.000148)	(0.000113)	(9.25e-05)	(9.24e-05)	(9.96e-05)
N. Other Conditions		0.00177***	0.00114***	0.00240***	0.00178***	0.00177***	0.00171***
		(0.000243)	(0.000370)	(0.000307)	(0.000243)	(0.000243)	(0.000261)
Student Status		0.00214	0.00399	0.000354	0.00212	0.00212	0.00222
		(0.00165)	(0.00264)	(0.00197)	(0.00165)	(0.00165)	(0.00171)
Constant	0.0471***	0.0409***	0.0356***	0.0231***	0.0406***	0.0408***	0.0447***
	(0.00260)	(0.00450)	(0.00630)	(0.00480)	(0.00450)	(0.00450)	(0.00489)
Observations	2,078,384	2,078,384	1,046,949	1,031,435	2,078,384	2,078,384	1,820,581
R-squared	0.005	0.005	0.005	0.003	0.005	0.005	0.005
Year/Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pseudo R2	0.00467	0.00505	0.00478	0.00294	0.00509	0.00505	0.00513

Notes: This table shows results from a linear probability model for flu incidence on MEPS data 2002-2016. Robust standard errors clustered at the individual level are in parentheses *** p<0.01, ** p<0.05, * p<0.1.