Social Science Research In A Pandemic

Kosali Simon
Coady Wing

Indiana University
Workshop in Methods
February 26, 2021
Lots of Interdisciplinary Covid-19 Research at IU.

Rapid release working papers: NBER, SSRN, Arxiv, etc.

Slower publication cycle

Focus on Behavioral Responses to Epidemiological Conditions, Changing Social Distancing Policies.
  - Mobility and Social Distancing
  - Health Status and Health Care Utilization
  - Labor Market Outcomes
Recent Work From Our Group


Initial U.S. state closures had small effects on state economies compared to national downturn, but local effects increase as closures continue

Bento et al, PNAS
Lozano Rojas et al., NBER#27127 – Worse than the Problem?
Gupta et al., NBER#27280 – Effects of Social Distancing
Gupta et al (Brookings Papers on Econ Activity) Review of the literature

Large disparities in labor market effects and likely COVID-19 exposure exacerbated by job patterns, rehiring, and churn

Montenovo et al., NBER#27132 – Disparities in Job Losses
Cheng et al., NBER#27419 – Back to Business
Andersen et al NBER WP

Many disruptions in Non-COVID-19 health care use, state policy impacts

Ziedan et al – using electronic health care records in US
Nguyen et al – using pharmacy records from >90% pharmacies in US
The Plan

Theory and Conceptual Frameworks (5-10 minutes)

Obtaining Data and Measuring New Concepts

Research Designs and Statistical Methods
  Measuring Prevalence
  Event Studies
  Heterogeneity by Geography and Sub-Population
    (Racial Disparities, Gender Disparities, Age Disparities)

Labor Disparities and International Comparisons (5-7 minutes)
Theory and Conceptual Frameworks That Guide Research During the Epidemic
Compartmental Models: Dynamics of an Epidemic

SEIR and SIR are core models in epidemiology:

Susceptible $\rightarrow$ Infected $\rightarrow$ Recovered

Susceptible $\rightarrow$ Exposed $\rightarrow$ Infected $\rightarrow$ Recovered

Emphasis To:
- Patterns of contact in the population
- Viral transmission rates
- Summary Parameters like $R_0$, which is the number of new infections that each case is expected to produce.

Implication:
- Policies that reduce contact or transmission can help control the epidemic.
Simple economic models of choices under constraint.

Advantage over basic SIR is that economic models allow people to react to changing epidemiological conditions.

Weakness is that choice models are typically not embedded in a full model of how choice leads to endogenous changes in epidemic.

Both frameworks are important for thinking about public policy and the effects of the epidemic on economic activity, social welfare, labor market outcomes, etc.
Home Production During An Epidemic

Consumer Preferences:

\[ u = u(\text{consumption, occupation, health}) \]

Commodity production:

\[ \text{consumption} = (x_c, e_c, d_c) \]

\[ \text{occupation} = (x_o, e_o, d_o) \]

Health Production Function

\[ \text{health} = (x_h, e_h, \rho D) \]

\[ D = \sum_{j \in \{c, o, h\}} d_j \text{ is total physical interaction across all activities.} \]

\( \rho \) is an infection risk parameter normalized so that \( \rho = 1 \) during ”normal” times

\[ \frac{\partial \text{health}}{\partial \rho D} < 0 \text{ so that exposure reduces health.} \]
Private demand for "social distance"

**Derived Demand** for physical interaction with other people:

\[ d_j = d_j(p_{xj} w, M, \rho) \]

\[ \frac{\partial d_j}{\partial \rho} < 0 \rightarrow \text{Infection risk reduces demand for interaction. Voluntary/Private Response.} \]

But distancing may confer health benefits on other people.

Positive externalities case for government policy to induce even larger reductions in physical interaction.
What kinds of policies?

**Direct Policies – focus on physical interaction**
- Pigouvian taxes on physical interaction or subsidies for distancing.
- Regulate the size of group events.
- Information/Encouragement to reduce physical interaction.

**Indirect Strategies – focus on complements to physical interaction**
- Levy a tax on complements to physical interaction.
- Close businesses that sell complements to interaction.
Examples: public transit, parks and beaches, restaurant meals, schools.

**Transmission Strategies** – focus on altering the infection risk parameter, $\rho$.
- Mask mandates, subsidies, encouragement.
- Handwashing and cleaning encouragement and requirements.
- Vaccine development and distribution.
The COVID-19 crisis spread across the US quickly...

By the end of March, nearly everyone in the US was living in a state that had

--at least one confirmed case
--declared an emergency
--closed restaurants
--issued stay-at-home mandates
Timeline of public sector responses

Note: Figure shows for each state, the timeline of their policy and information events shown in the legend; these are all the data presented in Figures 1, 2.1, and Appendix Table A1.
Obtaining Data and Measuring/Proxying Key Constructs
Cross-Cutting Themes Related To Data

--New Data Sources: representativeness, measurement error, publishability

--How to measure fast-changing policy environment?

--How to measure newly important concepts like daily mobility, consumption, work, etc.

--Disruptions to existing data collection infrastructure

--Ability to measure racial and geographic disparities: repeated concerns about ecological vs ego-centric data.

--Local vs National vs International Data
COVID-19 Pandemic & related crises

COVID-19 Health (deaths, cases, tests, vaccines)

Non C-19 Health (Acute, mental health, opioid crisis, child health/domestic violence)

Economic activity: Jobs (unemployment), mobility data, maternal labor supply

Miscellaneous

Education sector (school closures, online)
Overview By level of Data Engagement

Admin or survey sources

- Raw data
- Dashboards

Restricted

- Individual license/training
- Organization contracted

Download (unrestricted or minimal agreements)

Analysis, New evidence & insights

Analyze in your own computing environment
OR
Analyze in safety-enhanced different environments
Some global coverage examples

• [https://data.world/resources/coronavirus/](https://data.world/resources/coronavirus/)

Many resources at https://www.worldbank.org/

Also IMF and other organizations
Many others to be reviewed when Placing our results in context of global studies
Measuring Social Distancing

Data collected from cell phones and other smart devices.

- Safegraph – time at home
- Place IQ -- mixing
- Apple Mobility
- Google Community Mobility Reports

Data contain records from devices that include an app that contributes to the data vendor’s resource.

Each device is assigned a “home address”.

Today I will use two measures as an example. See working papers for analysis of a larger set of outcomes.
Measuring Social Distancing: Time Spent At Home

Median of hours that the device spent at its home location for each state x day. (SafeGraph)

Thin Lines = individual states.
Thick Line = cross-state average

The lines turn red when states issue “stay at home” mandates.

Much of the increase in time spent at home occurs before the mandate.

Note: Each grey line represents a state, and shows the mean number of hours a device spent in total in the house during the day. Red lines represent states with SAH laws, for the period after the law is in effect. The thick blue line represents a “smoothed” national local average (a generalized additive model (GAM)) of the states; there is a rise of 42.02% from March 1 (9.98 hours) to April 14 (14.18 hours).
Measuring Social Distancing: Mixing Index

Mixing Index (PlaceIQ)
-- collect the set of locations that a device visited during the day.
-- compute the number of other devices that visited each of those locations at some point during the day.
-- take the average of the other device counts across locations visited.

Thin Lines = individual states.
Thick Line = cross-state average

The lines turn red when states issue “stay at home” mandates.

Much of the decline in mixing happens before the stay at home mandates.
Basic Monthly CPS data for April 2020 was published on May 13th 2020, includes demographic and employment information at the micro level.

O*Net (Occupational Information Network): reports measures of occupation tasks. We build new: Face-to-Face and Remote Work Indices for each occupation

We classify industries as essential using Blau et al. (2020)’s definition of essential.

The New York Times collects information on the cumulative number of confirmed COVID-19 cases in a state. We used data for the CPS Survey April focal week.

The CDC China reports data based on deaths in China as of February 11, 2020. We use this to build a COVID-19 risk (as perceived in early phase) index by age and gender.
Employment levels fell very suddenly

Note: The figure presents the seasonally adjusted series for All Employees in non-farm jobs (millions) between January 2000 and May 2020. The shaded areas represent the 2001 Recession (March 2001 to November 2001), the Great Recession (December 2007 to June 2009), and the COVID-19 Recession. The figure implies that jobs lost during April and May 2020 exceed jobs lost in either of the two previous recessions.
The COVID-19 Recession led to more job losses in the first few months than the entire Great Recession.

Bigger declines for younger workers, Hispanic workers.

Smaller declines for people with advanced degrees.
Epidemic affected labor market outcomes across multiple household types
How did we build measures of Face-to-Face and Remote Work compatible jobs using the O*NET Indices?

<table>
<thead>
<tr>
<th>Index</th>
<th>O*Net Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>Face to Face</td>
<td>How often do you have face-to-face discussions with individuals or teams in this job?</td>
</tr>
<tr>
<td></td>
<td>To what extent does this job require the worker to perform job tasks in close physical proximity to other people?</td>
</tr>
<tr>
<td>Remote Work</td>
<td>How often do you use electronic mail in this job?</td>
</tr>
<tr>
<td></td>
<td>How often does the job require written letters and memos?</td>
</tr>
<tr>
<td></td>
<td>How often do you have telephone conversations in this job?</td>
</tr>
</tbody>
</table>

Note: The O*Net “Work Context” module (2019 version: available [www.onetcenter.org](http://www.onetcenter.org)) reports summary measures from worker surveys of the tasks involved in 968 occupations using the Standard Occupation Code, 2010 version. The questions use a 1-5 scale, where 1 indicates rare/not important. We developed three indices: (1) Face-to-Face interactions, (2) the potential for Remote Work, and (3) the extent to which work occurs Outside the Home using these variables. The value of each index for an occupation is a simple average O*Net questions listed in the table.
How did we build a measure of Essential Industries?

1. The DHS guidance outlines 14 categories that are defined as essential critical infrastructure sectors.

2. We follow Blau et al. (2020)’s definition of essential industries: they match the DHS text descriptions to the NAICS 2017 four-digit industry classification from the U.S. Census Bureau and to the CPS Industry Classification and to the CPS industry classification system.
Remote work capacity offered protection for some workers.

Note: Sample consists of April CPS 2020 respondents age 21 and above who are in the labor force. We compute the average percent recent unemployed in each occupation and plot that against the occupation's index value. Each occupational index has been standardized to have mean 0 and standard deviation 1. Each bubble represents a Census Occupation, and its dimension is proportional to the size of the workforce that holds that occupation in our sample. We include a line plotting the prediction from a linear regression of recently unemployed on each occupation index. To improve readability of the graphs, we excluded from the sample the 52 occupations whose recent unemployment rate is above the 90th percentile (i.e. greater than 36.17%).
Research Designs and Statistical Methods

Methods That Have Been Important In Our Work
Event Study Designs for Studying the Effects of Social Distancing Policies

\[ Y_{st} = \sum_{a=-21}^{-2} \alpha_a \times 1(TSE_{st} = a) + \sum_{b=0}^{21} \beta_b \times 1(TSE_{st} = b) + \theta_s + \gamma_t + \epsilon_{st} \]

\( A_s \rightarrow \text{Date when state } s \text{ adopts the policy} \)

\( TSE_{st} = t - A_s \rightarrow \text{time to adoption date for state } s \text{ in period } t \)

\( \alpha_a \) capture differential pre-trends.

\( \beta_b \) represent the day x day effects of state policy change. (policy effects)

The \( \gamma_t \) is the nationwide “time trend” that represents the likely path of physical mobility in the absence of state level policy changes. (non-policy effects)
Policy Event Studies: Time Spent At Home

Early events seem to matter the most.

Suggests that providing information played a big role.

Caution: These events happen very close in time.

Similar results for a large collection of alternative measures of mobility.
Time spent at home rose from 9.1 hours to 13.9 hours between the last week of February and the second week in April.

Estimates from the model suggest that without the emergency declaration policies, time at home would only have rose to 11.3 hours.

State policy “explains” about 55% of the rise in social distancing.
Exploiting Pre-Scheduled Events To Assess The Effects of Large Gatherings on Transmission

Since NHL, NBA, NCAA schedules were set up long before the epidemic, some cities were “quasi-randomly” assigned to have more games during the early stage of the epidemic.

Cities that had more games in January-February tended to have accumulated more cases and deaths by mid-May.

Hosting one additional NBA/NHL game results in an additional 428 COVID-19 cases and 45 COVID-19 deaths in the county where the game was played.

That’s 783 cases and 52 deaths for the MSA as a whole.

There were about 22 NHL and NBA games played in the average MSA yielding more than 17,000 cases and nearly 1,160 deaths per MSA.
Oaxacca-Blinder Decompositions To Study Disparities In Labor Market Outcomes Across Groups

Descriptive analysis suggests that disparities in job losses may be rooted partly in pre-epidemic sorting across occupations and industries.

How much of the racial disparities in job losses come from this type of pre-epidemic sorting vs differential layoffs?
Decomposing Aggregate Gaps into components due to “pre-epidemic sorting” vs “differential job loss”

Labor Market Outcomes For Sub-Groups A and B:

\[ Y_A = \alpha_A + x_A \beta_A + \epsilon_A \]
\[ Y_B = \alpha_B + x_B \beta_B + \epsilon_B \]

Aggregate Gap in Outcomes Can Be Decomposed:

\[ \bar{Y}_A - \bar{Y}_B = (\bar{X}_A - \bar{X}_B)\beta_A + \bar{X}_B(\beta_A - \beta_B) + (\alpha_A - \alpha_B) \]

Endowment Effect  Coefficient Effect
What Share of the Aggregate Gap Comes From Pre-Epidemic Sorting?

\[
\bar{Y}^A - \bar{Y}^B = (X^A - X^B)\beta^A + \bar{X}^B(\beta^A - \beta^B) + (\alpha^A - \alpha^B)
\]

\[
E = \left(\frac{X^A - X^B}{\bar{Y}^A - \bar{Y}^B}\right)\beta^A \times 100\%
\]

Overall pre-epidemic share can be decomposed further to determine the share due to sorting into remote work occupations, age and gender structure, education strata, etc.
Decomposition of April Job Losses For Several Sub-Groups
Simple Regressions To Study Job Loss Across People With Different Backgrounds

Main Regression

\[ y_{ij} = \text{Face}_j \beta_1 + \text{Remote}_j \beta_2 + \text{Essential}_j \beta_3 \]
\[ + \text{Mortality}_i \beta_4 + \text{Female}_i \beta_5 + \text{Child}_i \beta_6 + (\text{Child}_i \times \text{Female}_i) \beta_7 + \text{C19}_s \beta_8 \]
\[ + X_i \delta + \epsilon_{ij} \]

In alternative specifications we:

1. include the mortality risk variable and logged state COVID cases
2. Interact mortality risk with job task indices
3. Interact state’s COVID rate and job characteristics
4. Add occupation and industry fixed effects
People working in "essential industries" had far fewer job losses.

These patterns are much weaker in March, which is before the main covid shock.
Unprecedented labor market impacts, but weakly linked to local cases and policies

Some related findings from our other papers
More evidence of weak local responses
Hiring and Separation Rates Across Genders

Hiring Rates for Females and Males
1966 - 2020

Separation Rates for Females and Males
1966 - 2020
Hiring and Separation Rates by Firm Size

Hiring Rates for Small and Mid/Large-Sized Firms
1986 - 2020

Separation Rates for Small and Mid/Large-Sized Firms
1986 - 2020
Estimating the Population Prevalence of SARS-CoV-2
Only a tiny fraction of the population is tested each week.

Most testing is symptomatic, which creates selection bias: people who are tested are more likely to be infected than people who are not.

One solution: biometric surveys that test a random sample of the population. (Menachemi et al)

Another approach: combine clinical testing data with identifying assumptions to reduce uncertainty.
Useful Quasi-Facts

1. People who are admitted to the hospital for any condition are tested for Covid at much higher rates than the general population.

2. Many people are admitted to the hospital for things that are apt to have nothing to do with risk of Covid exposure.

-- Car accident
-- Heart Attack or Stroke
-- Labor/Delivery
-- Etc
Consider Three Core Assumptions

**Test Monotonicity:** People who are tested do not have lower prevalence than people who are not tested.

**Non-Covid Hospitalization Monotonicity:** People who are hospitalized for certain non-Covid health conditions do not have lower prevalence than the general population.

**Non-Covid Hospitalization Independence:** People who are hospitalized for certain non-Covid conditions have the same prevalence at the general population.
Research Strategy

Build a linked Hospital-Covid Test data set

Combine the data with alternative assumptions

Derive upper and lower bounds on true prevalence given the data and the maintained assumptions
Weekly Covid Test Rates In Indiana By Sub-Group
Bounds on Prevalence

Blue region is the upper and lower bound on prevalence using only the population testing data and the test monotonicity assumption.

Red region is the upper and lower bound on prevalence using test monotonicity & hospital independence.
International and Broader Literature
High frequency labor surveys show that the pandemic is having particularly adverse effects on younger workers, women and people that are more vulnerable.
**Figure 19. Asia: Share of Employment by Gender (All Industries) (Percent)**

Source: ILO, IMF staff calculations
Notes: Asia refers to Australia, Hong Kong, Indonesia, Japan, Korea, Malaysia, New Zealand, Philippines, Singapore, Thailand and Vietnam. Data refers to 2018.

**Figure 20. Asia: Change in Labor Force Participation Rate (by Gender) (Percentage points)**

Source: Haver Analytics, IMF staff calculations
Notes: COVID-19 = coronavirus disease; GFC = global financial crisis. Asia refers to Australia, Japan, Korea, Hong Kong, Thailand, and The Philippines. Data are seasonally adjusted. For COVID-19, data are up to June 2020.
Figure 23 Asia: Change in Employment Rates by Education Level (Percentage points)
Kim et al (2020)

• Singapore Life Panel (SLP) (monthly, individual level- information on consumption spending and labor market outcomes, mostly among age 50-70)

• Low-wealth workers suffered more from COVID-19 labor market shocks

• Surprisingly, employment rate decreased in May among healthier workers (-3.2 pp) but not among sicker workers

Kim, Seonghoon, Kanghyock Koh, Xuan Zhang. “Short-Term Impact of COVID-19 on Consumption and Labor Market Outcomes.” IZA DP No. 13354:
Aum et al (2020)

- Korean data from Labor Force Survey at Establishments (LFSE), a monthly survey of 40,000 sampled employers by the Ministry of Employment and Labor

- Economically Active Population Survey (EAPS), a monthly survey of 35,000 households collected around the 15th of each month by Statistics Korea, which includes worker characteristics (education, age, gender) and jobs (occupation and employment type)
Table 3: COVID-19 effect on employment by worker characteristics

<table>
<thead>
<tr>
<th></th>
<th>$\beta_2$</th>
<th>$\gamma$</th>
<th>Hourly wage (Aug 2019)</th>
<th>Share (percent)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>By education</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Middle school</td>
<td>0.39</td>
<td>-5.00***</td>
<td>11.7</td>
<td>(13.6)</td>
</tr>
<tr>
<td>High school</td>
<td>-1.82***</td>
<td>-3.79***</td>
<td>15.2</td>
<td>(38.5)</td>
</tr>
<tr>
<td>College</td>
<td>-0.68***</td>
<td>-2.37***</td>
<td>24.0</td>
<td>(47.9)</td>
</tr>
<tr>
<td><strong>By gender</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>-0.07</td>
<td>-2.82***</td>
<td>21.8</td>
<td>(57.2)</td>
</tr>
<tr>
<td>Female</td>
<td>-1.40***</td>
<td>-2.19***</td>
<td>16.1</td>
<td>(42.8)</td>
</tr>
<tr>
<td><strong>By age</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15-19</td>
<td>-21.05***</td>
<td>-13.08***</td>
<td>9.6</td>
<td>(0.8)</td>
</tr>
<tr>
<td>20-29</td>
<td>-3.34***</td>
<td>-6.80***</td>
<td>14.5</td>
<td>(14.0)</td>
</tr>
<tr>
<td>30-39</td>
<td>-1.60***</td>
<td>-0.23</td>
<td>20.6</td>
<td>(20.6)</td>
</tr>
<tr>
<td>40-49</td>
<td>-0.27***</td>
<td>-3.62***</td>
<td>22.5</td>
<td>(24.1)</td>
</tr>
<tr>
<td>50-59</td>
<td>-0.70***</td>
<td>-1.33***</td>
<td>21.7</td>
<td>(23.8)</td>
</tr>
<tr>
<td>60+</td>
<td>2.64***</td>
<td>-3.58***</td>
<td>14.0</td>
<td>(16.8)</td>
</tr>
<tr>
<td><strong>By employment type</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regular worker</td>
<td>-0.46***</td>
<td>-0.50***</td>
<td>22.2</td>
<td>(54.4)</td>
</tr>
<tr>
<td>Temporary worker</td>
<td>-4.43***</td>
<td>-7.20***</td>
<td>12.6</td>
<td>(21.6)</td>
</tr>
<tr>
<td>Employer</td>
<td>-3.50***</td>
<td>-1.97***</td>
<td>-</td>
<td>(5.4)</td>
</tr>
<tr>
<td>Self-employed</td>
<td>3.01***</td>
<td>-2.43***</td>
<td>-</td>
<td>(15.0)</td>
</tr>
<tr>
<td>Unpaid family worker</td>
<td>7.79***</td>
<td>-13.87***</td>
<td>-</td>
<td>(3.6)</td>
</tr>
</tbody>
</table>

Notes: Dependent variable is employment change in percent, except in the top panel, where it is percentage point change in unemployment and non-participation rates. $\beta_2$: coefficient on time dummy, $D_t(Mar)$. $\gamma$: coefficient on interaction term $D_t(DG)D_t(Mar)$. Robust standard errors in parentheses. *, **, *** represent significance at 10, 5, 1 percent. Hourly wage in thousand KRW (approximately 0.82 USD). Shares of the categories are from January 2020.
Fig. 3: Effect on employment by worker characteristic: within and between industries

Notes: The dark bars represent the estimates from Table 3. The light bars represent the implied coefficients if for each each demographic group, employment changes were solely due to industrial effects only, estimated in Table 1.
Dang and Nguyen (2020)

- individual-level data from the quarterly Labor Force Surveys in Vietnam
- lockdown increases the unemployment rate, the temporary layoff rate, and decreases the quality of employment. It also reduces workers’ numbers of working hours and their monthly incomes and wages
- unskilled workers are least affected, while skilled workers and workers in the service and trade sectors are most affected by the pandemic

India, South Africa


Monitoring trends in Europe—research examples
They have fielded a monthly survey related to COVID since April 2020.
lower earnings groups more than twice as likely to experience economic hardship relative to top quintile

Among pre-COVID employed individuals, men had a higher probability of being furloughed or dismissed from work, as well as whites in middle-income jobs.

Gaps attributable largely to structural gender earnings inequalities within occupations and because women and racial-ethnic minorities employed in essential occupations.
A Cross-National Design to Estimate Effects of COVID-Induced Non-Pharmacological Interventions

Dean R. Lillard
Ohio State University
German Institute for Economic Research (DIW Berlin), and
National Bureau of Economic Research (NBER)

We describe a research initiative that will explore the economic and social effects not of the COVID-19 itself but of the policies and information environment that COVID-19 spawned. We will exploit the substantial intra and inter-country temporal and geographic variation in non-pharmacological intervention policies induced by the COVID-19 disease. We will use data from ongoing household-based panel studies from 10 countries and rich administrative data from an eleventh. Six of the ten household panels have already fielded or will shortly field COVID-related questions to their main samples. A seventh, the PSID, has fielded questions to samples of the Child Development Supplement and Transition into Adulthood Supplement. The PSID and the other three panels will include COVID related questions in their next regular survey. All of them will be completed in 2021.

Keywords: COVID-19, COVID-Induced Non-Pharmacological Interventions; employment; earnings; income; subjective well-being; risk perceptions; social outcomes
Next steps & Q&A

• Understanding policies for recovery stages

• Educational equity & longlasting impacts

• Health & infrastructure for new crises