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

1 Introduction

The unexpected, rapid, and worldwide spread of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the disease it causes (COVID-19), and the associated consequences has spawned an explosion of research and research initiatives. Many of those initiatives aim to study whether people or their acquaintances were infected with COVID-19. Such studies require special sampling because, in every country of the world, there is a very low probability that a randomly drawn person will either have been infected with COVID-19 and an even lower probability that a person will know or even be acquainted with someone who died while infected with COVID-19.

Table 1 makes that point quite clearly. It lists the total number of confirmed COVID-19 cases and COVID-19 deaths reported on the Johns Hopkins University website as of May 18, 2020. The second and fourth columns list the total count of infections and deaths for the eleven countries we will study. The third and fifth columns report the number of infections and deaths one would expect to observe if one followed 100,000 people randomly drawn from the population before SARS-CoV-2 appeared.

Across all countries, COVID-19 infections and deaths are

rare. While there is still great uncertainty about the measurement error in these data, it is well-known that rates of infection and mortality rise rapidly for people older than 65 and who have other morbidities. But for our purpose, the probabilities shown in Table 1, convinced us that, in the general population samples of the household panel study data we will use, few people will have had direct and indirect experiences with COVID-19.

To take advantage of our data, we aim to analyze economic and social well-being effects not of the COVID-19 itself but of the policies and information environment that COVID-19 spawned. To do so, we will also collect, compile, and harmonize data on diagnosed cases of and deaths attributed to COVID-19, local and federal government policies that restricted economic and social life, and media coverage of COVID-19 infections and deaths and media coverage of mitigation policies. Our aim is to compile the COVID-induced mitigation policies both as they were implemented and rescinded. Our outcome data will come from rich administrative data and ten of the world's leading longitudinal panel studies. Because our data follow respondents over time, often over decades, we and other researchers will be able to use them to observe, describe, and model changes in an individual's outcome or behavior relative to a baseline level or long-running trend. In addition to their regular content, each survey has already fielded or will soon field questions that specifically ask about COVID-19-related experiences.

Our initiative resulted because of the collaboration be-

Contact information: Dean R. Lillard, 1787 Neil Avenue, 235B Campbell Hall, Columbus, OH 43210, USA (E-mail: lillard.13@osu.edu)

Table 1
Observations of COVID infections and death is a low-probability event

	Infections		Deaths	
Country	Count	per 100,000	Count	per 100,000
Australia	6,251	26.0	104	0.4
China	83,090	6.0	3,328	0.2
Germany	177,289	220.8	8,123	10.1
Italy	225,886	362.6	32,007	51.4
Japan	15,615	12.4	749	0.6
Russia	290,678	204.8	2,722	1.9
South Korea	11,055	21.4	264	0.5
Sweden	30,377	296.9	3,698	36.1
Switzerland	30,597	368.6	1,886	22.7
UK	246,408	376.8	34,794	53.2
US	1,533,039	461.9	87,217	26.3

Note: 2018 population estimates used to compute infections and deaths
per 100,000 residents in each country.

tween the household-based panel studies to create internationally comparable variables using data from the following household-based surveys: Household Income and Labour Dynamics in Australia (HILDA), German Socio-economic Panel (SOEP), Italian Lives (ITA.LI), Japan Household Panel Study (JHPS), Korea Labor and Income Panel Study (KLIPS), Russia Longitudinal Monitoring Study-Higher School of Economics (RLMS-HSE), Swiss Household Panel (SHP), "Understanding Society" (UK Household Longitudinal Study) (UKHLS), and the US Panel Study of Income Dynamics (PSID). Researchers associated with each of these studies collaborate as part of the Cross-National Equivalent file project (CNEF) (*https* : *//cnef.ehe.osu.edu/*) to create a file that is a subset of the original parent survey data (Burkhauser, Butrica, Daly, & Lillard, 2001; Burkhauser & Lillard, 2005, 4; Frick, Jenkins, Lillard, Lipps, & Wooden, 2007). Note that a file exists for Canada that includes data from the 1992 to 2009 Survey of Labor and Income Dynamics. The ITA.LI is a new survey that will contribute its first file this year.

The initiative discussed here resulted directly from discussions and collaboration among key researchers associated with each panel study. The current initiative also exemplifies one of the principal ways that CNEF aims to add harmonized data. Rather than harmonizing data for the sake of creating new variables, CNEF develops and expands the equivalized variable set to pursue research questions. The current initiative was born from discussions among CNEF researchers at each of the parent panel studies to first find out whether the directors of each study planned to field COVID-related questions to their samples - either in special supplemental surveys or as part of their regular survey administration. That discussion very quickly expanded to include the Principal Investi-

gators of the China Family Panel Studies (CFPS). As part of this initiative we plan to develop a CFPS-CNEF file.

One of the goals of the early discussion was to try to develop COVID survey questions that were, in the parlance of data harmonization, comparable *ex ante*. Data harmonization practitioners often regard *ex ante* harmonization as a touchstone because, in theory, one asks the same question regarding the same concept using the same wording and offering the same response options (in principal to a sample of respondents selected in the same way). The resulting data are, by definition, comparable. While it is indeed a goal, our discussion quickly confronted the reality of trying to harmonize survey questions across studies funded by national organizations. In practice, a funding agency in one country will often have objectives that necessitate specific questions or questions with a specific focus that is shared only in various degrees with the objectives of a funding agency in a different country. As a result, one can only harmonize questions across independent studies *ex ante* when each interested party shares a common goal. However, even when goals differ, there is almost always a core content that gets measured in ways that allow researchers to harmonize the resulting data *ex post*. In our early discussions, we confronted these issues and are continuing to confront them.

To try to inform the development of the COVID survey instruments, the CNEF associated studies shared existing COVID survey instruments with each other. The SOEP group was the leader and first in the field with a supplemental survey. Their instrument became a reference point for the other studies that were developing instruments. The ITA.LI group followed, adopting some but not all of the SOEP questions and response categories. The UKHLS instrument followed next and deviated more but still has quite a bit of over-

lap. Subsequently, the Swiss, Koreans, and Chinese developed instruments that share content and/or wording in varying degrees with the other three instruments that have been developed to date. The instruments for Australia, Japan, Russia, and the United States have not been finalized or fielded. Eventually, each survey will distribute the COVID survey data they collect. CNEF will make available all COVID data that can be *ex ante* or *ex post* harmonized.

Because the initiative will focus on understanding the economic and social consequences of the COVID-induced nonpharmacological interventions, our project will also compile and harmonize data on what was done in each country and sub region. Many research groups have been compiling those data in various forms. What may differentiate our project is that, as with CNEF harmonization efforts, we will harmonize the mitigation policy data with a particular research aim in mind. That means, for example, that when we study effects of COVID mitigation policies on employment, we will characterize the policies in terms of their likely or theoretical impact on decisions to work or decisions to fire or retain workers.

2 Data

The ten household panel studies share common features. All ten panel surveys collect information on household composition, income, employment, housing, marital status, subjective well-being, satisfaction with life and other factors, time spent giving or receiving help, and demographic characteristics. The surveys also differ, asking questions on different topics at different times to sometimes different samples.

Table 2 summarizes the basic features of the ten household surveys. It also lists the domains of the COVID-related data that have been or is planned to be collected. Other articles in this issue discuss the COVID data for the SOEP, UKHLS, and supplemental files of the PSID (Burton, Lynn, & Benzeval, 2020, 2; Kühne, Kroh, Liebig, & Zinn, 2020, 2; Sastry, McGonagle, & Fomby, 2020, 2). Note that, in addition to adding COVID content to the main annual survey, the UKHLS has funding and is fielding a special COVID supplement. That survey began in April 2020. The end date of that survey is still to be determined. Each panel's homepage provides details on their main survey structure and data.

All of the surveys started in a base year with a set of households drawn from a probability sample that typically oversample groups hard to locate. All people residing in the household in the base year get labeled as "original" sample members. Most of the surveys attempt to interview all household members above a given age (usually fifteen or sixteen but fourteen in the SHP and nine in the CFPS). Most of the surveys gather data on younger children by proxy. A significant advantage of the surveys is their following rules. Almost every survey uses a so-called eternal following rule (JHPS is an exception). In practice, the surveys not only re-

interview members in the original household in subsequent years but also follows every person who lived in the base-year household (original sample members) when they leave to form their own independent households. The surveys also interview members of the new household. Any child born to an original sample member gets designated as an original sample member, followed and reinterviewed in future years. Some of the surveys designate people who marry into a household as an original sample member. The upshot of these following rules is that, as time passes, the resulting sample includes respondents across multi-generational families living in different independent households. Almost every one of the surveys has added at least one refreshment sample to deal with panel attrition and changing population representativeness as the composition of national populations change with emigration and immigration.

We will also use administrative data from Sweden to construct a sample that mirrors the structure of the other household surveys. The Swedish data will combine data from various administrative agencies that include employment, health, public benefits, and pension records. Like the household panels, the Swedish sample will include members from multigenerational families living in the same and different households. Once constructed, we will negotiate with Statistics Sweden to make it possible for researchers to access those data or use the code that compiles them to recreate the sample.

3 COVID-19 mitigation policies, infections, and deaths

In our initiative, we will measure and harmonize across countries data on non-pharmacological interventions implemented both at the country level and in regions within countries, rates of COVID-19 infections and deaths. Dozens of research groups around the world are engaged in collecting these data. We will harmonize these data with an eye to the outcomes we will study. The goal is to use cross-country and within-country differences in the timing and implementation of various policies to explain the economic and social outcomes people experienced.

4 Conclusion

Our inititative will exploit a set of household-based panel surveys that have created a set of comparably defined variables that measure social and economic outcomes. Working as a group, the researchers associated with each panel study have or soon will administer special supplemental COVID-related questions. The prevalence rates of COVID-19 infections and deaths make it unlikely that there will be large numbers of survey respondents who have directly experienced COVID. However, we will study how COVID-induced non-pharmacological interventions affected the social and economic outcomes experienced by survey respondents. The

Table 2

Corona-related content associated with CNEF Studies

	Australia HILDA	China CFPS	Germany SOEP	Japan JHPS	Korea KLIPS	Italy ITA.LI	Russia RLMS-HSE	Switzerland SHP	United Kingdom UKHLS[b]	United States PSID
Country										
Study acronym										
Start year	2001	2010	1984	2004	1998	2019	1994	1999	2009	1968
Recent/current wave started	Aug-19	Jun-18	Feb-20	Jan-20	Apr-19	Jun-19	Oct-19	Sep-19	Jan-20	Mar-19
Interviews (individuals)	17,462	30,000	4,229	6,000	23,000	6,090	15,000	8,900	40,000	26,395
Interviews (households)	9,6665	12,000	4,229	4,000	11,700	3,466	7,000	5,700	22,400	9,614
Recent wave (expected) end	Feb-20	Feb-19	Aug-20	Apr-20	Dec-19	May-20	Feb-20	Mar-20	May-22	Dec-19
Survey-mode[a]	4, 5	4, 5	3, 4, 7, 8	1, 6	4, 5	4, 5, 7	3, 4	4, 5, 7	4-6	4, 5
Next data collection start	Aug-20	Jun-20	Apr-20	May-20	Apr-20	Apr-20	Oct-20	May-20	Apr-20[b]	Mar-21
Next data collection end	Mar-21	Feb-21	May-20	Jun-20	Dec-20	Dec-20	Feb-21	Jun-20	tbd[b]	Dec-21
Expected interviews	17,000	24,000	7,000	6,000	23,000	9,500	15,000	5,400	18,000	9,900
COVID survey-mode[a]	5, 6	4, 5	5	2, 6	4, 5, 8	7	3, 4	3, 6	6	4-6
Duration of survey (mins)[a]	8	Core	20-25	20-25	10-15	15-20	tbd	15	15-20	Core
Number of questions	tbd	20	69	55	40	40	tbd	65	150	tbd
COVID Topical modules on										
Corona prevalence in household	no	yes	yes	no	no	yes	yes	yes	yes	Core
Self-rated individual risk perception	no[c]	yes	yes	yes	yes	yes	yes	no	tbd	Core
Expected risk of corona	yes	no	yes	yes	yes	yes	yes	yes	tbd	Core
Health behavior	yes	yes	yes	yes	yes	yes	yes	yes	tbd	Core
Current employment status/changes	yes	yes	yes	yes	yes	yes	yes	yes	yes	Core
Separate module for self-employed	minimal	yes	yes	partly	yes	partly	yes	yes	yes	Core
Current child care	yes	Core	yes	yes	yes	yes	yes	yes	tbd	Core
Perceived child care burden	no[c]	no	yes	yes	yes	yes	yes	yes	tbd	Core
Schooling	yes	yes	yes	yes	yes	yes	yes	yes	tbd	Core
Current time use	yes	yes	yes	yes	yes	no	yes	yes	tbd	Core
Current satisfaction (domains)	yes (8)	yes (6)	yes (6)	yes (6)	yes (6)	yes (6)	yes (6)	yes (6)	tbd	tbd
Satisfaction with corona management	no	no	yes	yes	yes	no	yes	yes	tbd	tbd
Caring (self/others)	no	Core	yes	yes	yes	yes	yes	yes	yes	Core
Affective well-being	no	no	yes	yes	yes	yes	yes	yes	yes	tbd
Resilience	maybe	no	yes	yes	yes	no	yes	yes	tbd	tbd
Social cohesion	no	no	yes	yes	no	no	yes	yes	tbd	tbd
Anxiety/worry/fear	maybe	Core	yes	yes	yes	partly	yes	yes	yes	Core
Generalized trust	no[c]	Core	yes	yes	no	no	yes	yes	tbd	tbd
Current life satisfaction	yes	Core	yes	yes	yes	yes	yes	yes	yes	Core

tbd = "to be determined"

[a] Survey mode coding: 1) In person; 2) Mail; 3) PAPI; 4) CAPI; 5) CATI; 6) Web; 7) CAWI; 8) SCQ.

[b] UKHLS: Two waves in field simultaneously. COVID questions being asked on main interview and special new monthly survey. Information here and content information below refers to the special monthly COVID supplemental survey. End date is still to be determined.   [c] HILDA: Collected in other years.

eleven countries have the significant advantage that they represent countries that implemented widely different policies at different times and experienced widely different levels of infections and deaths. We will exploit those differences to identify how the policies affected the lives of individuals in the eleven countries.

Each of the household panels will make the COVID-related data available for researchers. The CNEF project will also make available the harmonized data from both the core and COVID-related household panel data and the harmonized COVID-induced mitigation policies for each country. CNEF will make the data available on a timetable that is partly a function of the data release schedule of each household panel listed above and whether and when CNEF gets the resources it needs to harmonize the panel and COVID-related data. As a general principle, CNEF will release the data as soon as possible. For the timetable and status of the project, researchers should periodically check at https://cnef.ehe.osu.edu/

Acknowledgement

References

Burkhauser, R., Butrica, B., Daly, M., & Lillard, D. (2001). The Cross-National Equivalent File: A product of cross-national research. In I. Becker, N. Ott, & G. Rolf (Eds.), *Soziale Sicherung in einer dynamischen Gesellschaft [Social insurance in a dynamic society].* Campus Verlag.

Burkhauser, R., & Lillard, D. R. (2005). The contribution and potential of data harmonization for cross-national comparative research. *Journal of Comparative Policy Analysis*, *7*, 313–330.

Burton, J., Lynn, P., & Benzeval, M. (2020). How Understanding Society: The UK household longitudinal study adapted to the COVID-19 pandemic. *Survey Research Methods*, *14*, 235–239.

Frick, J. R., Jenkins, S. P., Lillard, D. R., Lipps, O., & Wooden, M. (2007). The Cross-National Equivalent File (CNEF) and its member country household panel studies. *Journal of Contextual Economics [Schmollers Jahrbuch]*, *127*, 627–654.

Kühne, S., Kroh, M., Liebig, S., & Zinn, S. (2020). The need for household panel surveys in times of crisis: The case of SOEP-CoV. *Survey Research Methods*, *14*, 195–203.

Sastry, N., McGonagle, K., & Fomby, P. (2020). Effects of the COVID-19 crisis on survey fieldwork experience and lessons from two major supplements to the U.S. Panel Study of Income Dynamics. *Survey Research Methods*, *14*, 241–245.

Schnell, R., & Smid, M. (2020). Methodological problems and solutions for sampling in epidemiological SARS-CoV-2 research. *Survey Research Methods*, *14*, 123–129.

Commentary

The paper we were invited to review describes an initiative to collect and harmonize survey data allowing to analyze the social and economic consequences of Non-Pharmaceutical Interventions (NPIs) implemented as a response to the COVID-19 pandemic. The initiative covers household-based panel surveys in ten countries, which already collaborate within the Cross-National Equivalent File project (CNEF). Rather than focusing on direct and indirect experiences with COVID-19, these surveys will allow examining implications of the pandemic on households' occupational and financial situations, family life, happiness, child care, schooling, health behaviours, perceptions of the management of the coronavirus pandemic, among other topics. These data will enable longitudinal analyses together with the regular editions of the respective panel surveys, while the goal is also to coordinate the content of the surveys across countries to facilitate ex post harmonization and cross-national comparisons. The second main product will be the collection of harmonized data on NPIs at the country- and regional level. Since the primary aim of the initiative is to provide resources for cross-national research, in our commentary, we focus primarily on issues of data comparability.

The initiative will provide a welcome and rich data source, enabling social scientists and public administrators to study the consequences of the pandemic and the subsequent policy interventions for a wide range of social and economic topics. A clear advantage of this initiative is that it builds on and connects to prior work done by the CNEF. Furthermore, the project covers a diverse set of high and upper-middle-income countries with very different policy records in handling the crisis. The surveys will be administered in Western and East Asian democracies, as well as in Russia and China. Aside from the clear benefits of such broad collaboration, this diversity may have implications for the comparability of the resulting data due to the politicization of COVID-19 and limited access to transparent reporting in some of the countries.

The survey part of the initiative promises to provide harmonized survey data on the social and economic well-being of populations affected by the COVID-19 pandemic, and naturally faces several challenges related to ensuring data comparability. On the one hand, the panel survey projects have a history of collaboration as part of the CNEF and are well

aware of the advantages of ex ante harmonization. On the other hand, specific interests of the funders make it impossible to agree on a standard set of survey items for all country questionnaires—a practical obstacle of many multi-country survey projects. Moreover, questions need to preserve longitudinal comparability within the panels. Still, there will likely be some overlap in questions across all or most countries, and standardizing at least some elements of the questionnaire design and the overall survey process will facilitate future comparative work. The author emphasizes that the data will be collected with a particular research aim in mind (even though the details of the research questions are not discussed yet). Having an idea of the (type of) analyses that the data will be used for facilitates any harmonization effort, as it helps to reach decisions that are optimal in the light of concrete goals (given constraints in time and resources), rather than trying to meet expectations of imaginary users.

Deciding on survey items and—more broadly—issues of measurement is just one aspect of comparability. In the case of the initiative in question, sampling is determined by the survey projects, while the timing of the COVID-19 module is likely tied to the issues of funding and the funders' expectations. Given the different developments of the pandemic and the diverse strategies countries adopted in response, timing is an essential factor in the comparability of the survey data. Other aspects include possible systematic non-response caused by the pandemic situation itself, especially given that in some countries data collection seems to be planned via face-to-face surveys, but in a panel project, this should be straightforward to detect and possibly mitigate. Relatedly, there might be survey mode effects as some panels switch from interviewer-administered surveys to self-administered modes.

A significant strength of the project is the harmonization of data on NPIs as there is a lot of temporal and geographical variation of these measures even within countries. Yet, the article is not specific about what information will be collected, how the data will be compiled or how they will be linked to the survey data. One challenge is related to the diversity of administrative layers involved: NPIs can be implemented at the state, the regional and at the municipality level. A second challenge is the diversity of measures that can be adopted at the same time in one geographical unit: lockdown, social distancing, compulsory wearing of face-masks, closure of parks, shops, churches, university, schools, kindergartens, etc. Ideally, detailed information on NPIs would be collected at multiple administrative levels on a daily basis (as geographical units can move back and forth with regard to the implemented measures). Linking the reception of the NPIs by the media to geographic units below the national level is another challenge the initiative faces.

Furthermore, the project wants to put together a harmonized dataset of incidence and death rates of COVID-19.

This effort will need to take into account, among others, the differences in reporting cases across countries—not only due to official guidelines but also to deviations when applying those guidelines—as well as differences in access to testing (see Schnell & Smid, 2020, 2, in this issue)[1].

In summary, this initiative has the potential to provide a valuable data infrastructure for studying social and economic consequences of the COVID-19 pandemic in a diverse set of countries. Yet, the extent to which the datasets will allow for valid comparisons among countries will depend on the success of the harmonization efforts of the survey data and the quality of the collected data on NPIs. A follow-up publication could provide more details about the harmonization strategies the initiative will adopt, with a more in-depth discussion of the analytical opportunities created by the new data as well as its limitations.

Pablo Christmann
GESIS – Leibniz Institute for the Social Sciences
Marta Kołczyńska
Institute of Political Studies of the Polish Academy of Sciences

_____

[1]References are listed among the references of the main paper