

Two memos

Memo #1: Time in each setting – ATUS check

I am going through the process of implementing our risk-of-infection procedure and wanted to go over a quick check for one of the assumptions we make when we divide infections. I'll go through the example I used to think through it.

Let's assume total infections were as follows in month 1 and 2:

	Job A	Job B
Covid	40	80
No covid	160	120

If we assume that in both months, time distribution across settings remained the same, e.g., this distribution in both months (assuming everyone is risk averse):

	Job A	Job B
Home	40%	50%
Work	40%	30%
Leisure	20%	20%

Then our current risk model + IPF procedure would yield the correct number of infections per setting in both months. But let's imagine the time distribution changes from month 1 to month 2.

In month 1, we get this distribution:

	Job A	Job B
Home	30%	70%
Work	45%	15%
Leisure	25%	15%

And in month 2:

	Job A	Job B
Home	50%	30%
Work	35%	45%
Leisure	15%	25%

Maintaining the total number of infections (40 and 80), setting infections were as follows:

		Job A	Job B
Month 1	Home	12	56
	Work	18	12
	Leisure	10	12
	No covid	160	120
Month 2	Home	20	24
	Work	14	36
	Leisure	6	20
	No covid	160	120

Based on our current method, we'd average time in each setting across all months in the data, obtaining the following distribution:

	Job A	Job B
Home	40%	50%
Work	40%	30%
Leisure	20%	20%

Which would yield the following predicted probabilities:

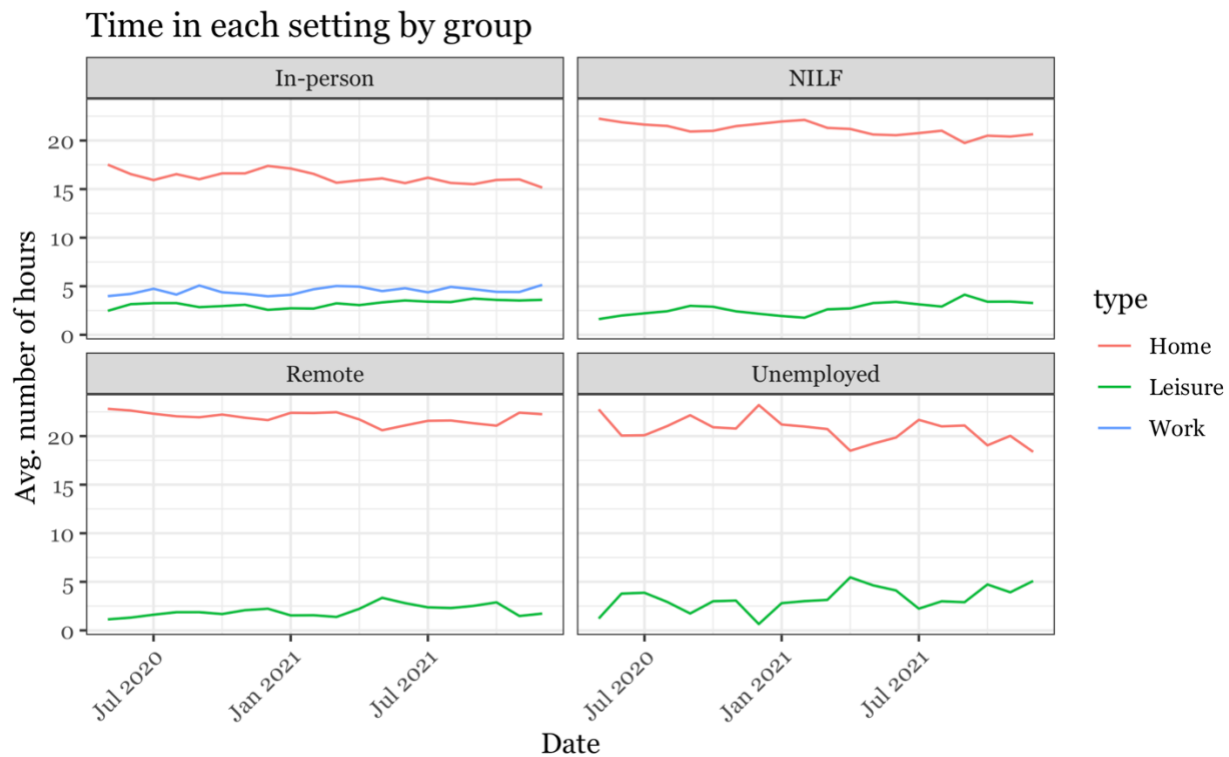
		Job A	Job B
Month 1	Home	0.08	0.2
	Work	0.08	0.12
	Leisure	0.04	0.08
	No covid	0.8	0.6

And fitting it to known totals from month 1, we'd obtain:

		Job A	Job B
Month 1	Home	20	48
	Work	10	12
	Leisure	10	20
	No covid	160	120

In other words, because we are averaging time in each setting from month 1 and month 2, we end up over or underestimating infections per setting in month 1. The upshot: for our procedure to be accurate, we would want the percent of time spent in each setting to remain consistent. With this

in mind, I looked at ATUS data from April 2020 to December 2021 to look at time spent in each setting per month. Below a graph showing the average number of hours in each setting by group:



Visually, it seemed to me that the “Unemployed” category was the one we might have to worry about, but the rest seem relatively stable to me. Looking at the variance of these monthly averages:

top_category	type	variance
In-person	Home	0.3956702
	Leisure	0.1326322
	Work	0.1390636
NILF	Home	0.4343926
	Leisure	0.4399344
Remote	Home	0.3390999
	Leisure	0.3422310
Unemployed	Home	1.6475068
	Leisure	1.5330101

I'm not sure how you feel about these results, I'd love to know your thoughts. Alternative, we can also avoid the assumption that the time distribution remains consistent and instead do a two-step process where:

- 1) First, we distribute covid infections across jobs but not across settings. Using this last example, we would get a model that would correctly predict 40 and 80 infections for job A and job B, respectively.
- 2) Then, we'd fit a new model using these new totals and the correct time distribution for each setting. E.g., for month 1, we'd have 40 sets of these rows for job A and 80 sets of these rows for job B:

Job	Weight	Home	Work	Leisure
A	1/3	0.9	0	0
A	1/3	0	1.35	0
A	1/3	0	0	0.75

Job	Weight	Home	Work	Leisure
B	1/3	2.1	0	0
B	1/3	0	0.45	0
B	1/3	0	0	0.45

- 3) Taking the predicted probabilities of the new model and fitting it to known totals, we would correctly obtain:

		Job A	Job B
Month 1	Home	12	56
	Work	18	12
	Leisure	10	12

In sum, if I understand it correctly, we don't need to assume time in each setting remains consistent across months, although the procedure might be more elegant/succinct if we do. Furthermore, based on ATUS data, it seems time in each setting has remained somewhat consistent, making the assumption defensible. I'd be happy with any decision! We could always run a robustness check where we don't make the assumption and compare results.

Memo #2 – Fuzzy categorization of in-person and remote workers

The second thing I wanted to flag regards the distinction between in-person and remote workers. The nature of the Pulse survey is such that our "in-person" category likely captures a lot of hybrid workers. As a reminder, this is the question we use to distinguish workers:

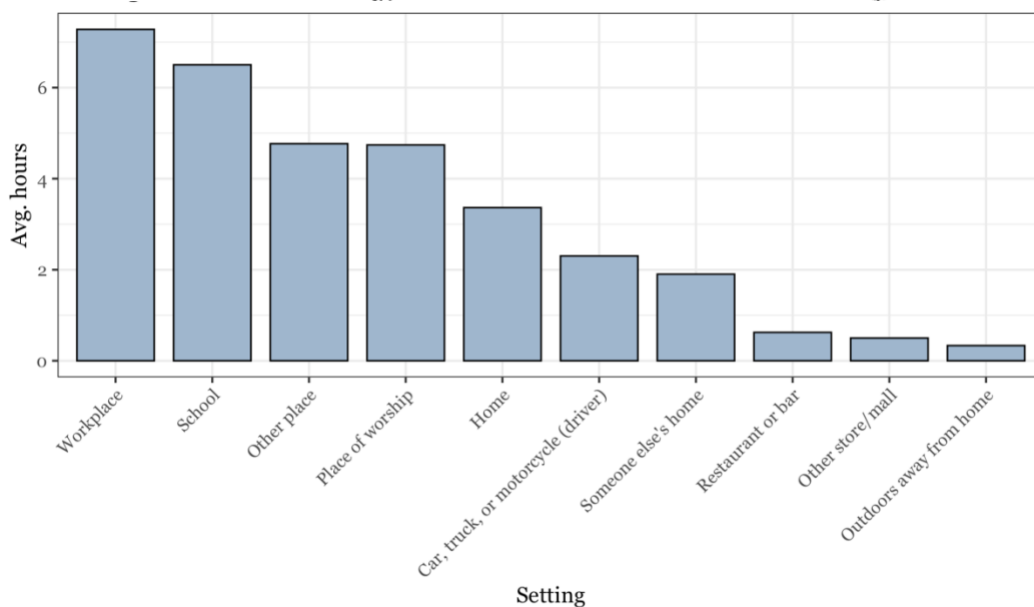
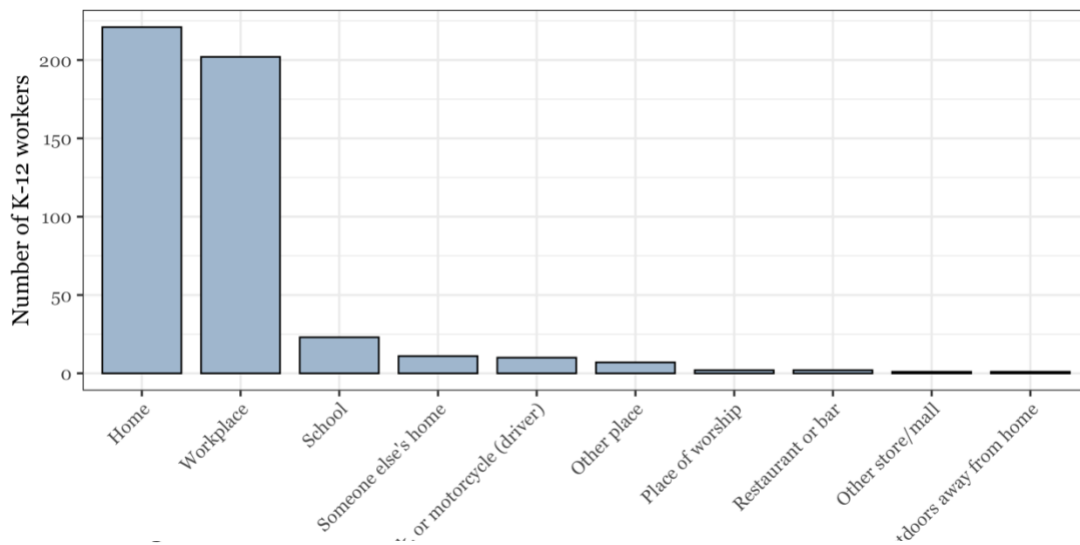
“Since January 1, 2021 [or “In the last 7 days” for later surveys], have you worked or volunteered outside your home?”

Thus, a hybrid worker, who might spend most days at home but goes to the workplace once or twice a week, would be categorized as “in-person”. I kept this in mind when I categorized ATUS respondents by following this procedure:

- 1) First, use the matched CPS responses to determine whether they’re employed, unemployed, or NILF
- 2) Then, from those employed, determine the setting of where they did activities related to their main job.
 - a. What I’d done in the past is categorize anyone who spend over 50% of work time away from home as in-person. However, based on the Pulse survey, this threshold is too high since even a few hours in the workplace would warrant a “Yes” response to the question above. Yet I was also worried about making the bar “too low” and categorizing anyone who spent *any* time doing work outside the home as in-person. What if, for instance, the person was working remotely at a coffeeshop or at a friend or family member’s house? The process I settled on was to categorize respondents as in-person if they spend time working outside the home, unless those outside-the-home places are settings where remote employees might spend time working (thinking of these as kinds of “remote workplaces”):
 - i. 3 = Someone else's home
 - ii. 4 = Restaurant or bar
 - iii. 7 = Other store/mall
 - iv. 9 = Outdoors away from home
 - v. 10 = Library
 - vi. 12-21 = Modes of transport (if person is commuting to one of such remote workplace)
 - b. One concern with this strategy is that there might be people doing actual in-person work in these settings (e.g., a librarian working at the library rather than a remote worker stationed there for the day). I assumed that in such a case, the librarian would mark the setting as “2=Workplace”.
 - i. As a sanity check, I looked at the distribution of work-time locations for K-12 school employees to see whether most said they did work in their workplace or at a school. Although the assumption proves to be somewhat right, the third most commonly selected was “school” and average hours for those who selected it was over 6 hours (figures on next page).

- c. To avoid this assumption, I thought we might distinguish those that might be working in the library because they're remote workers from those who are actually librarians by checking two things:
 - i. Their occupation/industry does not directly correspond with the setting
 - ii. Their occupation is labeled as “teleworkable” on the [Dingel and Neiman](#) paper we use to categorize CPS respondents.
- d. In summary, we categorize a worker as remote if they spend all work time inside the home or in one of six “remote workplaces”, so long as their occupation is “teleworkable” and does not correspond to one of the settings
 - i. Attaching a PDF here to show the comparison of population estimates coming from the ATUS vs the Pulse.

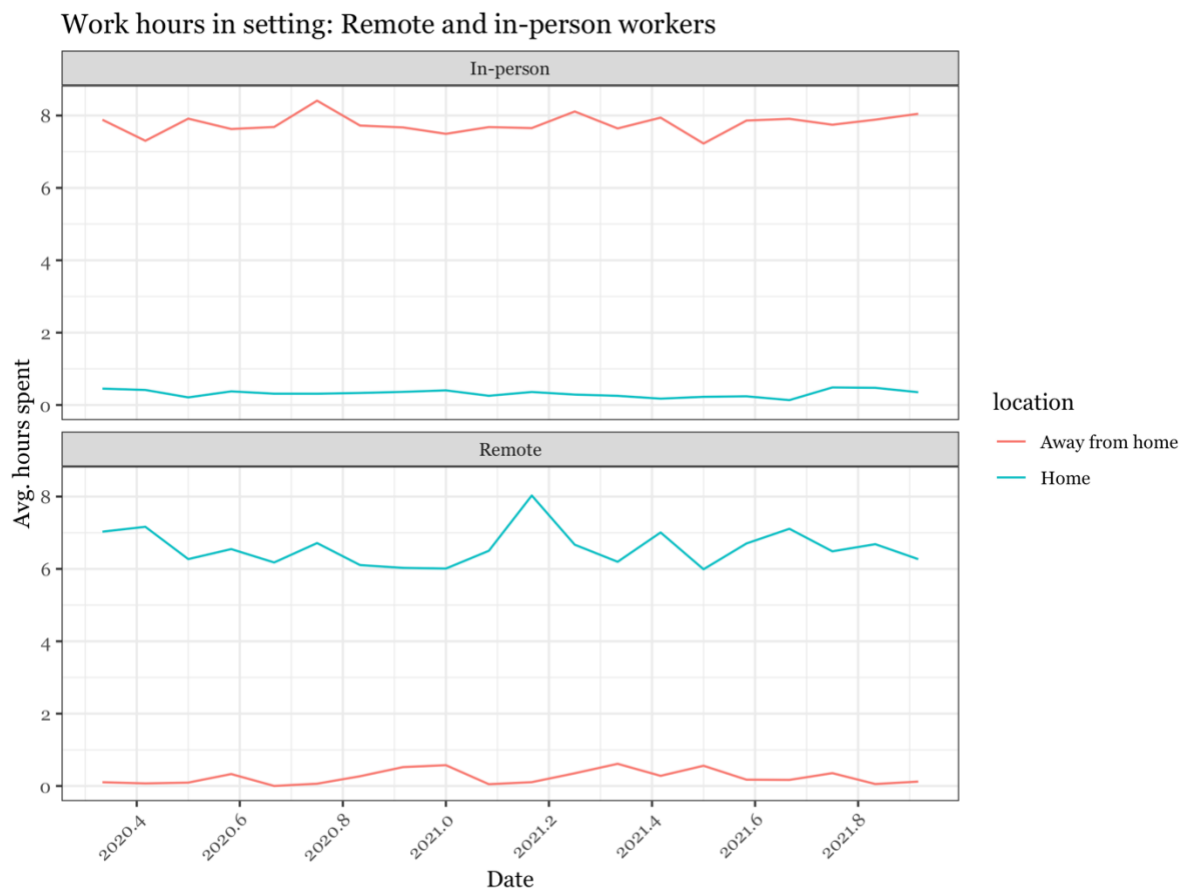
Setting of work hours for K-12 workers



I wanted to run this procedure by you to get your thoughts. I know the ATUS categorization is more of a second-order task but still one I wanted us to settle on. Most importantly though, I wanted to share because I realized we might be missing a small step in how we divide infections. The two key points:

- “In-person workers”, based on the Pulse survey, will also include hybrid workers. As such, some of their work time will be spent at home.
- “Remote workers” will not complete all of their work at home as some might spend some work hours in places like libraries, coffeeshops, outdoors, etc.

Below, is a graph that shows the average hours per setting where remote and in-person workers completed work activities:



For remote workers, our plan was to attribute some of the “home” infections to work based on how many hours at home were spent working. With atus data, we can also attribute a small number of infections happening outside the home (“leisure” infections) to work. Similarly, for in-person workers, we can attribute some “home” infections to work. However, this is a minor

fix given how little time in-person and remote workers are spending at home and away from home, respectively.