Sources for essential worker estimates

Original article

Gabriel Zucker[1]

This paper lays out the methodology behind our essential worker estimates. Estimates are based on 2018 1-year ACS data, available here. All analysis was performed using the national individual-level file, csv_pus.zip, using person-level frequency weights (*pwgtp*).

1. Defining <$15/hr workers

1.1 Defining hourly wages using ACS data

ACS does not report hourly wages. Rather, we calculate hourly wages based on:

- *wagp:* Earned income in last 12 months (only respondents with wagp!=0 were included in the analysis)
- *wkw:* Weeks worked in the last year
- *wkhp:* Hours typically worked per week, in the weeks worked

Wkw is coded categorically rather than continuously. We estimate the value at the midpoint of the range:

Table 1: Mapping of ACS weeks worked

Wkw value	Wkw label	Value assigned
1	50-52	51
2	48-49	48.5
3	40-47	44
4	27-39	34
5	14-26	19

Hourly wage is simply annual wages divided by annual hours (weeks times weekly hours).

Given the coding of wkw, uncertainty is of course greater among workers with fewer weeks worked in the last year. As a robustness check, we also run results including only workers with 48-52 weeks worked in the last year. The results below show that dropping <48-week workers does decrease the proportion of overall workers earning under $15/hr, but only by 2-3 percentage points in most cases. Some decrease is to be expected, since intermittent workers

[1] Gabriel Zucker, Public Interest Technology Fellow, zucker@newamerica.org

are liable to be earning less in general, so it seems likely the headline estimate is roughly correct.

We include all workers who are currently employed, whether or not they are currently at work.[2] 1.9% of the sample currently has a job but is not at work. Technically, some of this 1.9% should perhaps be excluded from the sample, depending on how long they plan to be out of work — but it is hard to tease this out from the available ACS data. Regardless, as that population represents only 1.9% of the sample, this correction would be unlikely to make much of a difference.

Arguably, some workers with apparent wages under $15/hr are not in practice being underpaid to the degree it appears in the aggregate. These include self-employed workers, whose implied hourly earnings may not be an entirely meaningful metric, and those working without pay in a family business or farm. As a robustness check, we also run results without these populations. The results below show that dropping these workers does not make a significant difference.

1.2 Consistency with published research

It is worth noting that, overall, ACS appears to show higher rates of underpaid workers than similar data at BLS.[3] BLS does not report a $15/hr cutoff, but BLS does report data on workers at or below the federal minimum of $7.25/hr, which they estimate at 2% *of the paid-hourly workforce* — a subpopulation which would appear to, if anything, skew disproportionately lower-income. Overall, meanwhile, 9.2% of all workers in the ACS sample earn at or under $7.25. This decreases slightly if we apply the restrictions described above (keeping only 48+ weeks of work, and dropping self-employed/uncompensated), but not by much, only to 7.2%. This appears to raise the possibility that ACS is somehow vastly overcounting very-low-paid workers. On a broader scale, however, the distributions of ACS and BLS hourly wage data do appear to match. Our workers show a mean hourly wage of $25.59 in 2018 compared to BLS's report of $27.76 in March 2019. (This downwards discrepancy is easily explicable as a result of top-coding high incomes in ACS.)

Our estimate that 39% of workers are paid under $15/hr, though, is very consistent with other previously reported research on the portion of workers under that benchmark. NELP's 2015 report estimates 42.4%, based on CPS data, and has been repeatedly cited. Oxfam reports 43.7%, using 2014 ACS data. Both of these numbers seem consistent with a 39% estimate based on data 4 years later, when inflation has eroded the worth of $15. (Oxfam and NELP's work is, again, consistent with Census and BLS telling slightly different stories about low-wage workers.)

[2] *esr*==1,2

[3] BLS does not make microdata publicly available enough to support most of the analysis in this paper.

2. Defining workers in essential industries

As discussed in the main paper, despite much public discussion of essential workers, no single authoritative definition has been issued for who these people are. We analyze three different proposals.

2.1. NYC - CEPR

The simplest analysis looks at the NYC Comptroller Office's list of a few frontline industries. Analysts at CEPR have helpfully posted their code mapping these frontline industries to Census industry codes, which are used to define workers in ACS data.[4] As CEPR uses ACS data and Census industry codes, there is no crosswalking required.

2.2. DHS - Brookings

DHS guidance qualitatively describes occupations within industries, but for analysis purposes, analysts at Brookings mapped the guidance only to industries. While technically not all workers in essential industries are in the listed essential occupations, it stands to reason that a majority of them are, especially among low-wage workers. (Section 4 accounts for any such discrepancy to a degree.)

The Brookings analysis uses 4-digit NAICS codes and further sorts industries into one of two categories: (1) entirely included, or (2) included on the essential list but "likely containing occupations that fall outside the DHS definitions." We interpret this two different ways: (1) Coding both categories 1 and 2 as entirely essential, or (2) Coding category 2 workers as having a 50% chance of being essential.

NAICS codes raise a further issue, because Census uses its own industry coding system, with imperfect crosswalking to NAICS industries. Census industry codes map imperfectly in two ways: (1) Census codes map at various levels of NAICS specificity, with some Census codes mapping to 6-digit NAICS codes, and some as few as 2, and (2) The Census codes may be mapped to multiple smaller NAICS partitions, or parts of several. These nuances are coded with letters P and M, such that, e.g., an ACS record with 21P is mapped to multiple partitions of NAICS 2-digit code 210000.

With regards to the first axis of variation, for simplicity, we defined "essentiality" scores for all three-, two- and one-digit NAICS codes based on the fraction of four-digit scores comprising them that are considered essential.[5] E.g., if a given three-digit code contains 5 four-digit

---

[4] The CEPR code appears to contain a few slightly different versions of included industries, (a) that listed in annotations, (b) that assigned to the variable *flind*, and (c) that assigned to the variable flind_d. We use (a), which is the most expansive, although all three lists are very similar.
[5] We merged the published list with NAICS's list of all four-digit industries, coding un-listed industries as 0s, so that averages represent the component of *all* industries in the category.

industries, of which 4 are essential, a worker in the three-digit industry is 80% likely to be essential.[6] If 2 were essential, 1 non-essential, and 2 partial, coded 0.5, the three-digit industry would be considered to have a 60% probability of being essential. This is of course highly imperfect, since four-digit industries vary greatly in size; in one glaring example, the transit industry is coded as only 83% essential due to the non-essentiality of charter bus operators, who in real life are unlikely to comprise a meaningful portion of transit workers. But, on balance, it's not clear how this would bias the overall estimates, even if it increases the variance.

The second axis of variation was disregarded; an ACS record with 21P was treated as if it were simply 21. This again, may introduce variance, but should not introduce bias.

As a robustness check, we calculate an alternate version of the estimates in which only NAICS codes with an essentiality score of 1 are included. As shown below, the change does not do much to overall percentages.

As a result, there are four versions of the DHS estimates, of which we present only the first in the main paper. The total number of workers implicated in each model of course varies somewhat drastically, which is to be expected; but the fractions under $15/hr, reassuringly, do not change very much. In addition to being the most plausible estimate, the first model presents a fairly middle of the road estimate relative to the other specifications along most indicators.

Table 2: Mapping of DHS industries

	Partial industries are included:	Workers are included:
DHS - 1	As 50% essential	With weights proportional to the industry's essentiality score
DHS - 2	As 50% essential	Only if the industry's essentiality score = 1
DHS - 3	As essential, equivalent to other industries	With weights proportional to the industry's essentiality score
DHS - 4	As essential, equivalent to other industries	Only if the industry's essentiality score = 1

2.3 Delaware essential industries

The State of Delaware issued a list in early April of industries allowed to remain open. Some non-essential industries come with several caveats with conditions under which the industries *are* indeed allowed to open; these were disregarded. The list, like the DHS-Brookings list,

[6] Two- and one-digit codes were defined on their component core 4-digit codes, not their component composite three- or two-digit scores.

contains 4-digit NAICS codes, which were mapped to Census industry codes via the same methodology described above. As before, we show two versions, only the first of which is used in the main paper. As above, the two models do not show very different results in terms of fractions of workers, although the magnitudes differ.

Table 3: Mapping of Delaware industries

	Workers are included:
DE - 1	With weights proportional to the industry's essentiality score
DE - 2	Only if the industry's essentiality score = 1

3. Other characteristics

We also report the following other characteristics of <$15/hr workers in essential industries:

- Percent female, based on *sex* variable
- Percent POC, based on *rac1p* variable!=1
- Health insurance coverage, based on *hicov*
- Disability status, based on *dis*

In the overall workforce, we estimate 47.6% of workers are women, which is consistent with the BLS estimate of 46.9%. We estimate 25.6% of workers are people of color, which is roughly consistent with the BLS estimate of 22.0%.

4. Correction from 2018 ACS to current crisis data

The resulting estimates reflect the reality of the labor market in 2018. Of course, the labor market has changed incredibly in the last several weeks, in ways that may impact these estimates, and detailed survey data is not yet available to track these changes. Most importantly, many of the workers who *were* earning $15/hr in essential industries in 2018 may not be commuting to work today, given both massive layoffs, and the increased prevalence of telework, neither of which can be measured with precision for this specific population. It seems likely that the prevalence of layoffs and teleworking is liable to be fairly low among low-wage workers in essential industries, but there is no way of knowing for sure. Working conservatively, we estimate that 25% of the workers who were working for <$15/hr in essential industries in 2018 are no longer commuting to work today.

5. Full results across models

Table 4: Numbers of workers and <$15/hr across all model specifications
Numbers in millions

	All workers		Dropping <48 weeks worked		Dropping self-employed and family business/farm	
	Total workers	% under $15/hr	Total workers	% under $15/hr	Total workers	% under $15/hr
All industries	139.8	39.0%	121.8	36.6%	133.4	39.2%
DE - 1	93.6	35.9%	83.8	33.9%	88.8	36.1%
DE - 2	88.1	34.6%	79.2	32.7%	83.5	34.7%
NYC	41.3	41.8%	36.3	39.9%	39.4	42.1%
DHS - 1	52.1	35.7%	47.0	33.9%	50.3	35.8%
DHS - 2	42.1	37.5%	37.9	35.7%	40.8	37.7%
DHS - 3	59.1	34.3%	53.6	32.5%	57.1	34.4%
DHS - 4	55.9	34.0%	50.8	32.2%	54.2	34.1%

Shaded cells are those included in the main paper. Final estimates are reduced by 25% as per Section 4.

Table 5: Characteristics of <$15/hr workers across model specifications
The below table shows only the 'all workers' model, not dropping those <48 weeks or the self-employed.

	Percent of <$15/hr that are women	Percent of <$15/hr that are POC	Percent of <$15/hr that lack health insurance	Percent of <$15/hr with a disability
DE - 1	48.4%	30.7%	19.1%	7.0%
DE - 2	47.9%	30.5%	18.8%	7.0%
NYC	54.5%	31.9%	19.7%	7.2%
DHS - 1	54.4%	32.8%	15.9%	7.3%
DHS - 2	57.7%	33.7%	15.0%	7.5%
DHS - 3	53.9%	32.4%	15.6%	7.2%
DHS - 4	56.0%	32.7%	14.4%	7.3%

Shaded rows are those included in the main paper. The DHS estimates are cited in the main figures.