

Employment Data Evaluation

Addison Larson

September 25, 2018

1 OVERVIEW

This paper evaluates differences in the 2013 vintage of two purchased employment data sources—the National Establishment Time-Series (NETS) and InfoGroup—for Conshohocken, Montgomery County, PA. Specifically, these two datasets are evaluated based on summations of employment over census blocks, block groups, and tracts; the overall composition of establishments by number employed; establishment-level differences in employment; differences in large employers; and overall data quality. See Table 1, *General Properties of NETS and InfoGroup Data for Conshohocken*, for a brief overview of the two datasets.

The Delaware Valley Regional Planning Commission (DVRPC) purchases NETS data when it becomes available. This occurred most recently in 2013. However, NETS data requires extensive manual cleanup for DVRPC’s nine-county region. If InfoGroup data requires substantially less correction, then it may be worth the additional cost to purchase it instead of NETS. This study assumes NETS data is already clean and consistent.

There are critical differences between NETS and InfoGroup data sources, as well as between these sources and LEHD LEHD (Longitudinal Employer-Household Dynamics) data.

1. **NETS** is a mystery.
2. **InfoGroup** is also a mystery. Doesn’t this have sole proprietors?
3. **LEHD** is publicly-available employment and origin-destination data from the Census Bureau. LEHD is tabulated at the block level and is synthesized from multiple imputation of tax and employment records. There can be noise in the data at the block level for privacy reasons. Even though LEHD data is derived from tax records, it is likely geocoded to the block level using the Census Geocoder, which is less spatially accurate than other geocoding options ***HALP! Is this true? Or is it that something like Google can better handle idiosyncracies in address quality?***.

CONFLICTING SIGNALS.

	NETS	InfoGroup
Total Number of Establishments	637	304
No. Establishments > 0 Employees	631	286
No. Establishments > 50 Employees	24	19
Proposed Sample Size (90% CI, MOE \pm 5%)	143	—

Table 1: General Properties of NETS and InfoGroup Data for Conshohocken.

2 SUMMATIONS OF EMPLOYMENT COUNTS BY GEOGRAPHIC LEVEL

When InfoGroup and NETS point-level employment data are aggregated to the block, block group, and census tract level, the counts differ greatly within each spatial unit. This section uses 2015 LEHD Workplace Area Characteristics as an employment comparison for NETS and InfoGroup.

BLOCK-LEVEL DIFFERENCES. Table 2, *Differences in Employment, Block Level*, includes the counts by block and three additional columns: NETS x IG, NETS x LEHD, and IG x LEHD. A value of “Yes” in any of these columns indicates absolute percentage difference¹ between the counts in the datasets in excess of 100%. The results of the absolute percentage difference calculations indicate that InfoGroup and LEHD counts are the most similar at the block level, with 17 of 46 blocks differing over 100% in job counts.

BLOCK GROUP- AND TRACT-LEVEL DIFFERENCES. It is suspicious that InfoGroup has 655 employees recorded for Block Group 420912041021 and Tract 42091204102, given that NETS and InfoGroup have nearly zero employees recorded for the same area. In addition, it is suspicious that LEHD has 180 employees recorded for Block Group 420912059062 and Tract 42091205906, when neither NETS nor InfoGroup have similar counts. The latter discrepancy may be attributable to the different source years of the data, since LEHD data comes from 2015. The discrepancies may also point to the need to manually correct both InfoGroup *and* LEHD data.

OVERALL COUNT DIFFERENCES. In general, NETS has little in common with either InfoGroup or LEHD. LEHD has the fewest job counts overall, at 4,311. NETS has nearly 1,500 more jobs than LEHD for the neighborhood of Conshohocken—this hints that one of these datasets must be far off the mark.

¹The absolute percentage difference is calculated as:

$$\frac{|v_1 - v_2|}{\frac{v_1 + v_2}{2}} \times 100$$

Block GEOID	NETS	InfoGroup	LEHD	NETS x IG	NETS x LEHD	IG x LEHD
420912041021003	0	655	16	Yes	Yes	Yes
420912042001006	72	0	0	Yes	Yes	No
420912042001008	430	1071	195	No	No	Yes
420912042001009	880	427	584	No	No	No
420912042001011	1056	551	105	No	Yes	Yes
420912042001012	858	24	34	Yes	Yes	No
420912042001014	33	34	118	No	Yes	Yes
420912042001015	1	0	0	Yes	Yes	No
420912042001016	41	44	34	No	No	No
420912042001018	110	24	23	Yes	Yes	No
420912042001019	2	0	1	Yes	No	Yes
420912042001021	60	57	106	No	No	No
420912042001022	44	44	9	No	Yes	Yes
420912042001034	63	176	0	No	Yes	Yes
420912042001035	93	493	542	Yes	Yes	No
420912042001037	4	5	1481	No	Yes	Yes
420912042001038	8	0	3	Yes	No	Yes
420912042001039	3	0	1	Yes	No	Yes
420912042001040	25	17	2	No	Yes	Yes
420912042001041	1	0	0	Yes	Yes	No
420912042001048	126	64	36	No	Yes	No
420912042001054	1	0	0	Yes	Yes	No
420912042001056	2	0	0	Yes	Yes	No
420912042002000	317	287	126	No	No	No
420912042002002	86	27	0	Yes	Yes	Yes
420912042002005	16	2	8	Yes	No	Yes
420912042002006	2	5	0	No	Yes	Yes
420912042002008	284	3	1	Yes	Yes	No
420912042002009	3	0	0	Yes	Yes	No
420912042002010	2	0	0	Yes	Yes	No
420912042002012	746	547	468	No	No	No
420912042002013	54	0	6	Yes	Yes	Yes
420912042002014	257	175	142	No	No	No
420912042002015	6	0	0	Yes	Yes	No
420912042002016	6	2	2	No	No	No
420912042002017	36	36	19	No	No	No
420912042002018	2	0	0	Yes	Yes	No
420912042002019	10	7	0	No	Yes	Yes
420912042002020	2	0	0	Yes	Yes	No
420912042002021	3	0	0	Yes	Yes	No
420912042002023	3	0	0	Yes	Yes	No
420912042002025	2	0	0	Yes	Yes	No
420912042002027	1	19	31	Yes	Yes	No
420912042002028	19	22	36	No	No	No
420912042002029	10	4	2	No	Yes	No
420912059062045	0	48	180	Yes	Yes	Yes
TOTAL	5780	4870	4311	26	32	17

Table 2: Differences in Employment, Block Level.

Block Group GEOID	NETS	InfoGroup	LEHD
420912041021	0	655	16
420912042001	3913	3031	3274
420912042002	1867	1136	841
420912059062	0	48	180
TOTAL	5780	4870	4311

Table 3: Differences in Employment, Block Group Level.

Tract GEOID	NETS	InfoGroup	LEHD
42091204102	0	655	16
42091204200	5780	4167	4115
42091205906	0	48	180
TOTAL	5780	4870	4311

Table 4: Differences in Employment, Tract Level.

3 COMPOSITION OF ESTABLISHMENTS BY DATA SOURCE AND NUMBER EMPLOYED

For the neighborhood of Conshohocken, NETS has 637 observations and InfoGroup has 304. Because of the discrepancy in the number of overall records, Figure 1, *Percentage of Establishments at Six Employment Levels*, shows the percentage of establishments at each employment level. The results indicate that InfoGroup has a higher percentage of large employers. NETS has more employees in Conshohocken partially because it has a higher number of records overall.

However, this information gives rise to the question: does InfoGroup systematically report more employees than NETS for the same business establishments?

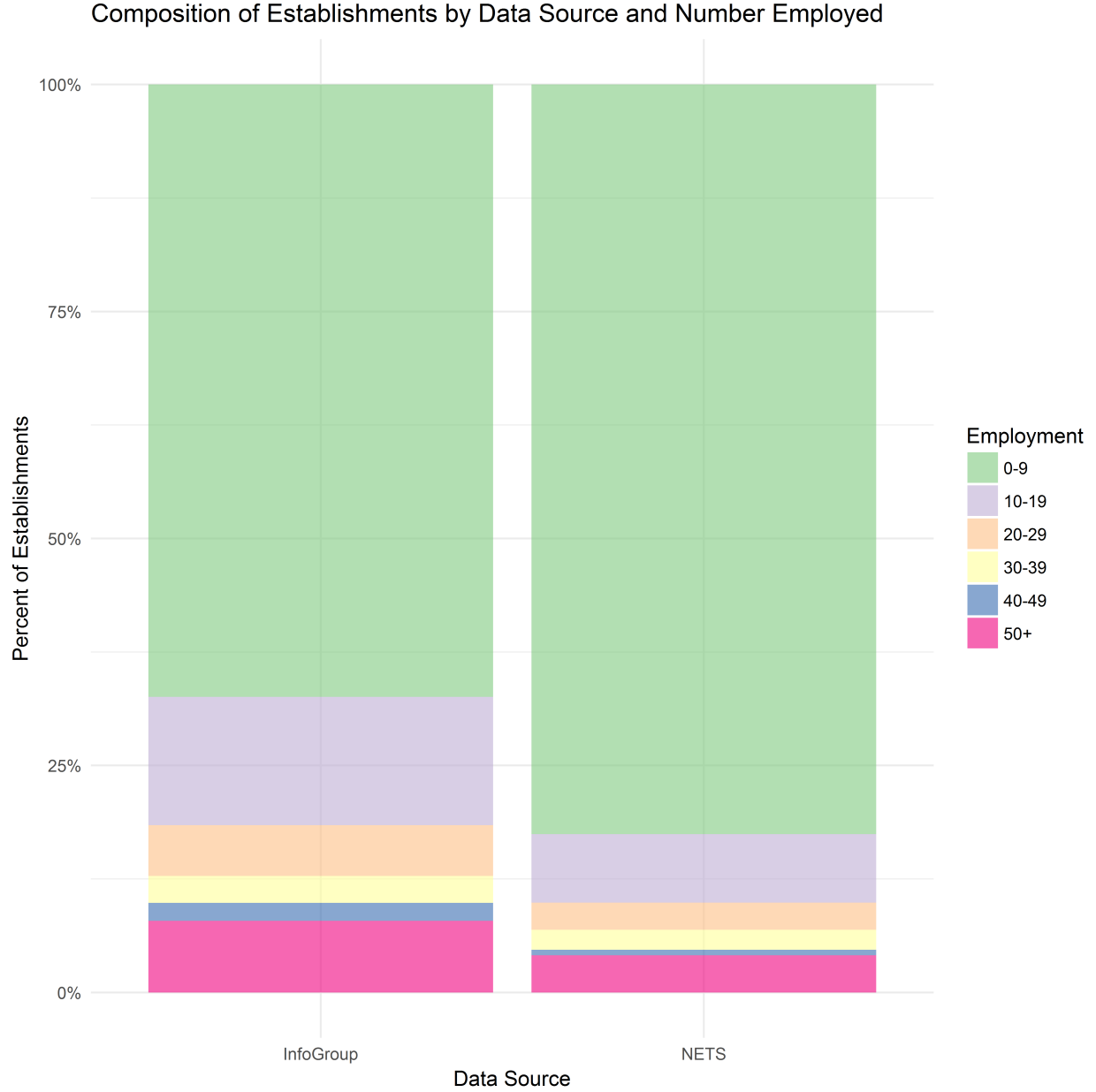


Figure 1: Percentage of Establishments at Six Employment Levels.

4 ONE-TO-ONE COMPARISON OF INFOGROUP SAMPLE TO NETS DATA

For the random sample of 143 InfoGroup records, we manually matched the InfoGroup firms to their counterparts in the NETS dataset. We found 72 matches in total. Of these, there were:

- 44 instances where InfoGroup reported more employees than NETS for the same establishment;
- 18 instances where NETS reported more employees than InfoGroup for the same establishment; and

- 10 instances where employment was the same in both NETS and InfoGroup.

The results indicate that, while InfoGroup has fewer overall records, it tends to report more employees than NETS for identical establishments. The average absolute percentage difference (see Footnote 2) between NETS and InfoGroup employment was 79.32%, and the median was 66.67%.

5 LARGE EMPLOYERS

It is important to compare large employers in Conshohocken for at least three reasons: 1) They are the most visible employers in the area and therefore the easiest to verify; 2) They skew employment totals for a single geographic unit, making it important that they are correctly placed; and 3) Small percentage differences between data sources for these establishments can translate into large aggregate differences in employment for the area. Figure 2, *Distribution of Establishments by Employment and Source*, shows the InfoGroup and NETS records with over 50 employees. Accounting for the overall difference in the number of records between NETS and InfoGroup, NETS reports establishments with roughly 50 employees twice as often as InfoGroup. This may drive the overall difference in employment tallies between NETS and InfoGroup for the study area.

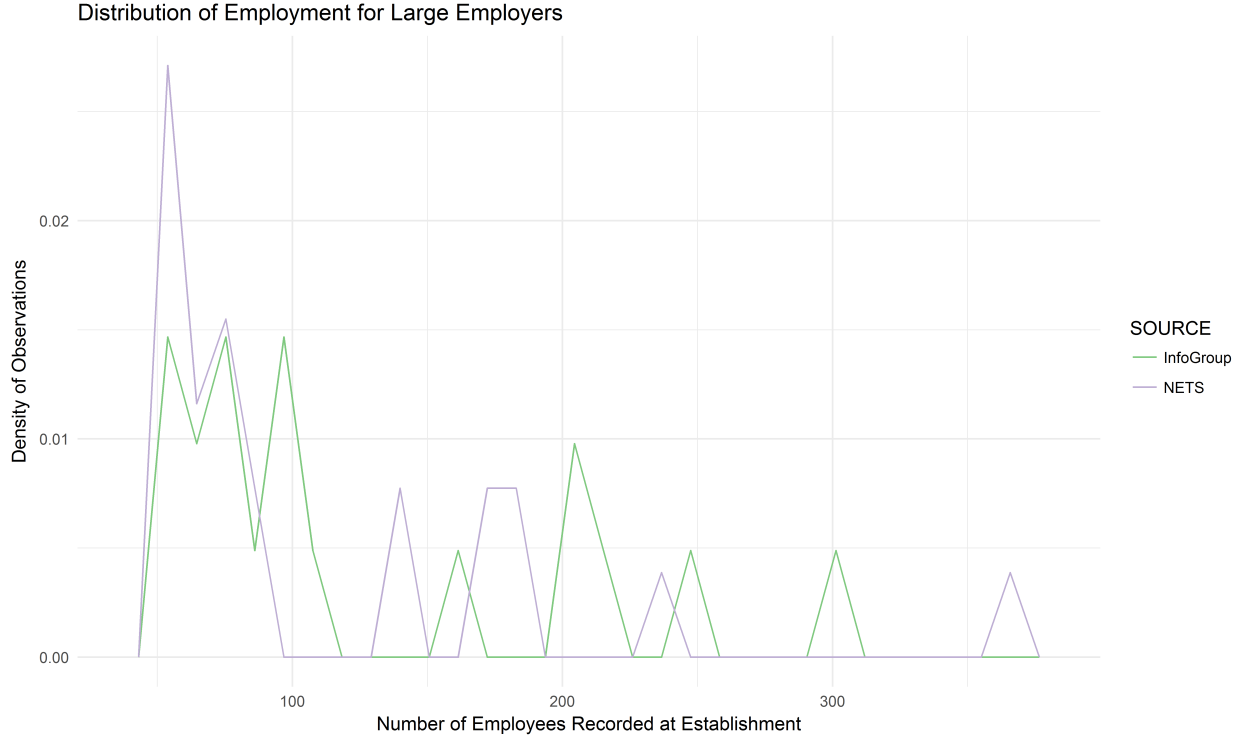


Figure 2: Distribution of Establishments by Employment and Source.

24 establishments in NETS and 19 in InfoGroup have a total reported number of employees exceeding 50. Of these, 7 large employers are present in both datasets. View Table 5, *Differences in Employment for Seven Large Employers* for employment counts by employer and data source. In four instances, InfoGroup reported more employees than NETS for the same establishment. However, it is worth highlighting two employers: Employer D, where NETS reported 113 more employees than InfoGroup; and Employer E, where InfoGroup reported 118 more employees than

NETS. Employer D occupies only one floor of a modest-sized office building: it is likely that the InfoGroup record is closer to the true number of employed in this instance. Furthermore, an online search shows that Employer E self-reports as having over 250 employees at this particular location, pointing again to the InfoGroup record as the more accurate record. Further investigation of the veracity of large employer records might indicate which of these data sources is more reliable in general.

Employer	NETS	InfoGroup
A	56	75
B	56	64
C	171	165
D	363	250
E	182	300
F	138	110
G	60	75

Table 5: Differences in Employment for Seven Large Employers.

6 OVERALL DATA QUALITY

This section, as with the rest of the analysis, assumes NETS data is already clean and consistent. To evaluate the overall quality of InfoGroup data, we took a sample ($n = 143$) of the InfoGroup data for Conshohocken and manually cleaned it: