

Data Shop

Data Shop, a department of Cityscape, presents short articles or notes on the uses of data in housing and urban research. Through this department, the Office of Policy Development and Research introduces readers to new and overlooked data sources and to improved techniques in using well-known data. The emphasis is on sources and methods that analysts can use in their own work. Researchers often run into knotty data problems involving data interpretation or manipulation that must be solved before a project can proceed, but they seldom get to focus in detail on the solutions to such problems. If you have an idea for an applied, data-centric note of no more than 3,000 words, please send a one-paragraph abstract to david.a.vandenbroucke@hud.gov for consideration.

Introducing the Ohio New Establishment Dynamics Data

Joel A. Elvery
Ellen Cyran
Cleveland State University

Abstract

The Ohio New Establishment Dynamics (O-NED) data set tracks the number of establishments that first started employing people between the second quarter of 1997 and the first quarter of 2008 and measures the employment and payroll data for these new establishments. O-NED enables researchers to measure the growth trends of cohorts of new establishments for up to 5 years after the cohort's birth. These data are the first publicly available data that document the growth rates of new establishments at the substate level. The finest unit of geography O-NED measures is a county. This article describes how O-NED is constructed and defines the variables included in the data. It closes with two examples of how researchers can use the data.

Introduction

Publicly available data are insufficient to answer a number of questions about the growth patterns of new establishments. The Ohio New Establishment Dynamics (O-NED) data set fills some of those gaps for Ohio. An establishment is a single physical location of a firm; firms can have one or many establishments. O-NED focuses on establishments during their first 5 years, a crucial shakeout period for new establishments. New establishments are grouped based on their year of birth and these groupings are called *birth cohorts*. O-NED provides annual tabulations of the number of establishments, employment, and wages for each birth cohort. The tabulations cover the period from April 1, 1997, through March 31, 2008. The data include 11 cohorts—7 for 5 full years and 4 for less than 5 years. O-NED provides separate tabulations for births unaffiliated with preexisting firms (entrepreneurial births), births affiliated with preexisting firms (other births), and fast-growing births (gazelles) of each type. O-NED is an outgrowth of the work of Knaup (2005), Knaup and Piazza (2007), Talan and Hiles (2007), and the U.S. Bureau of Labor Statistics (BLS) Entrepreneurship Team. This article describes the construction and structure of the data and provides two examples of research that the data enable.

Few data measuring business dynamics are publicly available. In 2004, BLS introduced the Business Employment Dynamics (BDM) data series. The quarterly BDM data series enables data users to measure the job creation and destruction and the establishment birth and death numbers that underlie the employment totals published in the Quarterly Census of Employment and Wages (QCEW). The BDM quarterly update is released 7 months after the quarter it covers. BLS continues to improve the BDM data by adding new features. One limitation of the BDM data is that the finest level of geographic detail is the state and the finest level of industry detail is the major sector. Furthermore, it contains only measures of establishment births and deaths and does not shed light on the growth patterns and survival rates of new establishments over time.

In December 2008, the U.S. Census Bureau began releasing Business Dynamics Statistics (BDS), which tabulates annual job creation and destruction statistics by firm age and either firm size or initial firm size. The tabulations are available for the United States as a whole or by sector or state. BDS and O-NED are similar because both measure the employment and number of establishments in businesses aged 1 to 5 years and enable analysts to track how employment and the number of establishments change over time. Several important differences exist, however. O-NED is designed for tracking the growth of new establishments in Ohio for their first 5 years, whereas BDS provides a more comprehensive set of job creation and destruction statistics for the nation as a whole. O-NED uses establishment age and measures firm age only for entrepreneurial establishments, for which firm age equals establishment age. BDS uses firm age, not establishment age, and includes more firm age categories. O-NED provides more geographic detail than BDS, but BDS includes tabulations by firm size and decomposes changes in employment and the number of establishments into the portions due to new entrants, continuing establishments, and exiting establishments. BDS covers the United States as a whole from 1977 through 2009, whereas O-NED covers only Ohio from 1997 through 2007.

Construction of the O-NED Data

The microdata we used to create O-NED is a combination of the longitudinally linked QCEW microdata from BLS and the edited ES202 data housed at the Maxine Goodman Levin College of Urban Affairs (Levin College) at Cleveland State University.¹ Both data sets cover only Ohio and are provided through a special partnership between Cleveland State University and the Bureau of Labor Market Information of the Ohio Department of Jobs and Family Services (ODJFS). BLS provided the longitudinally linked QCEW microdata to ODJFS for this project.

We combined the microdata sources to take advantage of edits that researchers at Levin College made over a period of years. In particular, O-NED took industry and geography codes from the edited ES202 data. One challenge in creating tabulations for cohorts of establishments is that the industry and geography codes of establishments can change over time. New establishments are especially likely to have code changes because some enter the data set with incomplete information, and BLS and ODJFS assign those establishments codes after they have gathered more information. To minimize the effect of these code changes on our tabulations, we applied the last valid codes we had for establishments to the data for all quarters.²

The O-NED data cover establishments born from April 1, 1997, through March 31, 2008. The sample is restricted to private establishments that did not experience any identifiable splits or consolidations during their first 5 years. This restriction greatly reduced the volatility in the data, because most splits are not truly new establishments but are continuing establishments that changed how they report their data.

Most splits are identified using relevant comment codes, but some are identified based on substantial changes in the number of establishments affiliated with a single employer identification number (EIN). EIN is used to determine which establishments belong to the same firm. Based on careful exploration of the data, we developed a set of rules to identify these splits by finding EINs that simultaneously have increases in their number of establishments and unusually large decreases in the average size of their establishments. Most cases treated as uncoded splits are those for which the EIN had employment of more than 50 people in the birth quarter, the number of establishments grew from the quarter before the birth to the birth quarter, and the average employment of establishments affiliated with the EIN fell by 80 percent from the quarter before the birth to the birth quarter. It is harder to identify splits for EINs with few units or little employment, and, based on exploring the data, we developed a conservative formula to identify these small splits.³

To further reduce the problem of false births, we examined the data for about 250 large EINs in which it was unclear whether the EINs experienced splits or had an unusually large number of

¹ The ES202 is Ohio's version of the QCEW microdata and is based on establishment data collected as part of the unemployment insurance system.

² If an establishment has an invalid code, such as a county code of 999, we use a previous, valid code when possible. We assign cases that have only invalid codes that change a single invalid code for all quarters. Based on our work verifying a subset of the code changes, we believe that no more than 25 percent of the code changes lost by pushing back codes were valid changes.

³ See Elvery and Cyran (2010) for more details on this and other topics.

births. We examined the data to see if the new establishments affiliated with the EINs had predecessors or if ownership changed, which would suggest they are false births. For most cases, we did not find conclusive evidence that they were false births and treated them as births. Even with the careful use of the data and hand checking of large births suspected to be false, it is likely that some false births remain in the data. Uncoded splits and ownership changes, which can appear to be births, are more prevalent for new establishments affiliated with preexisting EINs than for other new establishments. Therefore, we believe that entrepreneurial births are less likely to be false births than are nonentrepreneurial births (Elvery and Cyran, 2010).

One goal of O-NED is to demonstrate what can be created with existing BLS microdata. As such, we use definitions that are consistent with those BLS uses for BDM and those proposed by the Organisation for Economic Co-operation and Development (Ahmad, 2006). Exhibits 1 and 2 provide the precise definitions. A birth is any new establishment in the state, regardless of its ownership. An entrepreneurial birth is the birth of a new firm, not an additional establishment of an existing

Exhibit 1

Definitions of Terms

Term	Definition
Establishment	A work site in the QCEW/ES202 data.
Year t	The last three quarters of calendar year t combined with the first quarter of the following calendar year.
Quarter 1 of year t	First quarter of the calendar year following calendar year t .
Birth	The addition of an establishment with positive employment to the QCEW/ES202 data. Establishments that are combinations of previous establishments will not be considered births. Establishments that split off from existing establishments will be births.
Entrepreneurial birth	An establishment birth wherein the establishment has a UI account number and EIN that is unique in Ohio. The establishment must also not be a split off from an existing establishment.
Birth cohort j	Establishments born in year j .
Entrepreneurial birth cohort j	Establishments that had an entrepreneurial birth in year j .
Establishment counts	The number of establishments in the first quarter of year t .
Employment	Average reported monthly employment in quarter 1 of year t . Employment at birth is the reported employment for the quarter the establishment is born. Employment is restricted to employment covered by the Ohio UI system.
Wages	Reported wages for quarter 1 of year t or, in the case of wages at birth, during the quarter the establishment is born.
Total wages	Reported wages for all of year t .
Gazelle in year t	An establishment with an average annual employment growth rate of at least 20%, averaging across years t , $t-1$, and $t-2$, and that has employment of at least 10 in year $t-2$.

EIN = employer identification number. ES202 = establishment data from the Ohio unemployment insurance system. QCEW = Quality Census of Employment and Wages. UI = unemployment insurance.

Exhibit 2

Definitions of Data Elements Included in the O-NED Data

Data Element	Value	Definition
EMP	502	Sum of average monthly employment during quarter 1 of 2002 for the 1997 cohort.
QTRWAGE	3382199	Sum of quarterly wages during quarter 1 of 2002 for the 1997 cohort.
ANNLWAGE	14215225	Sum of quarterly wages from quarter 2 of 2001 through quarter 1 of 2002 for the 1997 cohort.
UNITS	49	Number of establishments from the 1997 cohort active in quarter 1 of 2001.
EMPVSYR1	125	EMP as a percentage of the cohort's employment at the end of year 1.
UNITSVSYR1	73	UNITS as a percentage of the number of establishments in the cohort at the end of year 1.
QWVSYR1	144	QTRWAGE as a percentage of the cohort's employment at the end of year 1.
AWVSYR1	323	ANNLWAGE as a percentage of the cohort's employment at the end of year 1.
EMPVSPY	81	EMP as a percentage of the cohort's employment in quarter 1 of 2001.
UNITSVSPY	88	UNITS as a percentage of the number of establishments in the cohort in quarter 1 of 2001.
QWVSPY	83	QTRWAGE as a percentage of the cohort's employment in quarter 1 of 2001.
AWVSPY	91	ANNLWAGE as a percentage of the cohort's employment in quarter 1 of 2001.

O-NED = Ohio New Establishment Dynamics.

Notes: O-NED tabulates the variables for each cell. A cell is a combination of a unit of geography, a unit of industry, a type of birth or gazelle, a cohort, and an age. The values come from the cell of Akron CBSA, manufacturing sector, entrepreneurial births, 1997 cohort age 4.

firm.⁴ A nonentrepreneurial birth is a new establishment affiliated with an existing firm. Year of birth is the year the establishment first entered the data. Survival to age t is defined as having positive payroll and employment for at least one quarter of the year that is t years after birth year. Employment and wages are those covered by the unemployment insurance system of Ohio.

Although the QCEW data are updated quarterly, we annualize the data. We define a year as a set of four quarters, starting with the second quarter of a calendar year and ending with the first quarter of the following year. For example, establishments that first enter the data from the second quarter of 1998 through the first quarter of 1999 would be counted as part of the cohort born in 1998. This unit of time is chosen because the QCEW microdata register a disproportionate share of establishment births in the first quarter of the year. A portion of these establishments were likely actually born earlier, so keeping them with those born in the previous three quarters groups establishments by cohort more effectively than using calendar years would. Using four quarters enables more geographic and industry detail by increasing the number of establishments per data cell. Focusing on annual data also keeps the data tractable for a broad group of potential users.

⁴ An EIN is treated like a firm. An EIN of all zeros is sometimes given to new establishments until they report their permanent EIN. Therefore, we treat a birth with an EIN of all zeros as an entrepreneurial birth.

Data users can conduct many different and rich analyses with O-NED, but the utility of the data depends on how much industry and geographic detail is feasible, given the simultaneous need for confidentiality and reliability. We experimented with a variety of ways to balance the tradeoffs of detail, usability, and sample size. In the end, O-NED tabulations are made for the following units of geography: Ohio; Metropolitan Status (central county of metropolitan area, other metropolitan, and nonmetropolitan); Economic Development Region;⁵ Metropolitan/Nonmetropolitan status by Economic Development Region; Core Based Statistical Area (CBSA); Central/Noncentral county status by CBSA; and County. The industry detail varies by geography. A cross-sector total is available for all units of geography. Major sector-level tabulations are available for all units of geography except County. Three-digit North American Industry Classification System-level tabulations are available only for Ohio as a whole.

Description of Data Files

This section describes the file structure and variables included in the publicly released O-NED tabulations.⁶ The data are released in two formats, a SAS data file and a set of Microsoft Excel® spreadsheets.

Each observation in the data represents a unique combination of a unit of geography, a unit of industry, a type of birth or gazelle, a cohort, and an age. For each observation, the data include the number of establishments, employment, first quarter wages, and annual wages. Survival rates of these variables, both year-to-year and compared with the end of the birth year (age = 1), are also included. Each observation also includes variables to identify the observation and navigate the data, including indicators of geographic and industry detail, geography codes, sector codes, cohort, age, year of data, an indicator for type of birth and whether it is a set of births or a set of gazelles, and a suppression flag.

Exhibit 2 shows the core data elements for one observation of data. These data cover the year 2001 for entrepreneurial births from the 1997 cohort that are located in the Akron, OH CBSA and are in the manufacturing sector. The exhibit includes the values and a description of what the numbers mean for this observation. The 49 establishments remaining in this cohort after 4 years had 502 employees and paid them a total of \$14.2 million from the second quarter of 2001 through the first quarter of 2002. The cohort had 25 percent more employment and 27 percent fewer establishments than in its birth year, and both employment and the number of remaining establishments declined between years 3 and 4.

⁵ Ohio's Economic Development Regions are contiguous combinations of counties that have similar or intimately linked economic functions. They are defined by the Ohio Department of Development.

⁶ The data and documentation are available online at <http://urban.csuohio.edu/economicdevelopment/ONED.html>.

Although the structure of the data is space efficient and easy to navigate, data users will find that they need to combine data across observations to calculate some statistics. For example, calculating the percentage of all entrepreneurial births from the 2003 cohort that were gazelles at age 5 requires combining the number of establishments in the cohort at age 0 and the number of gazelles in the cohort at age 5.

Example of Use

The key feature of O-NED is that it measures the growth trends of cohorts of new establishments for the crucial first 5 years of existence.⁷ Using survival rates—the cohort’s percentage of the initial number of establishments surviving or the percentage of year-1 employment retained—is a convenient way to measure these trends. Exhibit 3 graphs establishment and employment survival rates for groups of cohorts with 5 years of complete data. Exhibit 3a shows establishment survival rates for grouped cohorts. Establishment survival rates among cohorts were similar for all years and were unaffected by whether the business was started before or during the 2001-to-2002 recession. About 85 percent of all establishments survived to the second year. At the end of year 5, slightly more than one-half (53 percent) of all establishments were still in existence for all cohorts.

Business establishments that started toward the end of the expansionary years (1997 and 1998 cohorts) saw steep declines in employment after their third year (exhibit 3b). These establishments initially grew very rapidly but experienced a big hit during the recessionary years. For establishments that started just as the recession began (1999 and 2000 cohorts), employment also decreased with the onset of the recession. The decline in employment was at a lesser rate than the previous cohorts, however, because these establishments did not have the chance to grow as much as companies that were started in previous years. Establishments that started during the recession are grouped in the 2001 to 2003 cohorts. As seen on the graph, these establishments were not severely affected; they declined in employment at a much slower rate.

Exhibit 4 provides the size and survival trends of the average birth cohort for each of the eight largest CBSAs in Ohio. This exhibit covers only entrepreneurial births, not those affiliated with an existing firm. Entrepreneurial businesses in the Columbus, OH CBSA had more employment growth than those in other CBSAs, with an employment survival rate of 111.0 percent.⁸ Entrepreneurial establishments represented a higher-than-typical share of total employment in the Columbus, OH CBSA, where the average cohort had 1.53 percent of total employment in its fifth year. The Cleveland, OH CBSA had the largest number of new births and the largest amount of employment, however. In five of the CBSAs, the average cohort of entrepreneurial births had less employment at the end of the fifth year than at the end of the first year.

⁷ We draw these examples from Yamoah, Austrian, and Elvery (2009).

⁸ We create the survival rates by averaging across each cohort’s survival rate and, therefore, the rates do not match what one would calculate based on the cohort sizes in years 1 and 5.

Exhibit 3

Establishment (a) and Employment (b) Survival Rates for Grouped Cohorts

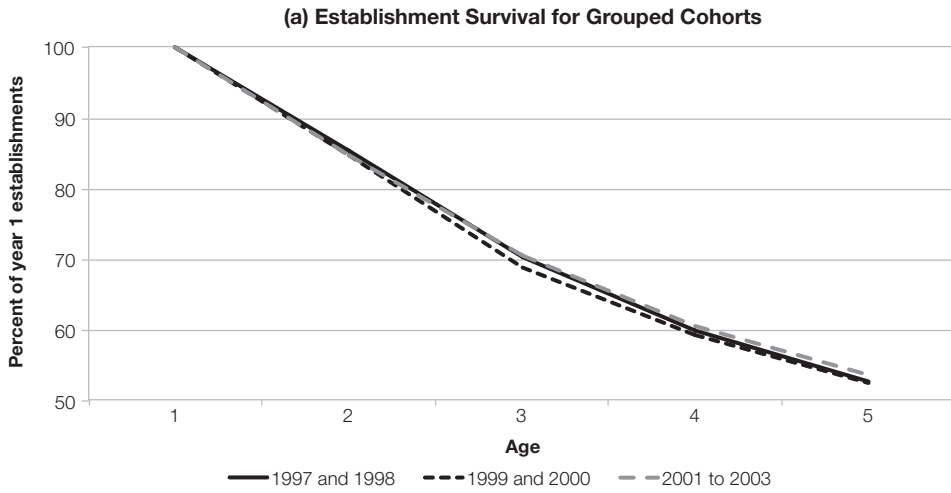


Exhibit 4**Size and Survival of Average Entrepreneurial Birth Cohort, by CBSA**

CBSA	Year 1		Year 5		Survival Rate		Employment Share	
	Employment	Units	Employment	Units	Employment	Units	Year 1	Year 5
Cleveland	11,433	3,224	12,243	1,888	101.2%	56.7%	1.26%	1.38%
Columbus	9,795	2,670	11,233	1,441	111.0%	53.5%	1.35%	1.53%
Cincinnati	8,859	2,349	9,728	1,290	104.7%	53.4%	1.28%	1.41%
Dayton	3,760	963	3,949	547	99.3%	54.3%	1.11%	1.20%
Akron	3,625	987	3,790	588	98.6%	57.3%	1.35%	1.40%
Toledo	3,293	832	3,330	478	96.3%	54.9%	1.22%	1.25%
Youngstown	2,331	557	2,131	325	84.8%	56.1%	1.42%	1.33%
Canton	1,733	490	1,765	285	97.3%	56.4%	1.15%	1.20%

CBSA = Core Based Statistical Area.

Conclusion

O-NED data described in this article are now available to anyone who wants to use them through the website of the Center for Economic Development at the Levin College at Cleveland State University. These data provide the first set of publicly available tabulations of establishment and employment survival for multiple cohorts from longitudinally linked QCEW microdata. O-NED includes tabulations down to the county level, providing more geographic detail than any comparable data. The value of these data will be determined by how analysts and researchers use it. We encourage people to dive in and use it and hope that similar data can be made available for the nation as a whole.

Acknowledgments

This research was made possible by a data improvement grant from the Ewing Marion Kauffman Foundation and a grant from the President's Initiative Fund of Cleveland State University. The Bureau of Labor Market Information of the Ohio Department of Jobs and Family Services provided the necessary data, with the cooperation of the U.S. Bureau of Labor Statistics. The authors thank all these parties.

Authors

Joel A. Elvery is an assistant professor in the Department of Urban Studies at Cleveland State University.

Ellen Cyran is a senior programmer and analyst in the Center for Economic Development at Cleveland State University.

References

- Ahmad, Nadim. 2006. A Proposed Framework for Business Demography Statistics. Working paper (June). Paris, France: Organisation for Economic Co-operation and Development.
- Elvery, Joel A., and Ellen Cyran. 2010. Methodology and Documentation for the Ohio New Establishment Dynamics Data. Unpublished manuscript. Available at http://urban.csuohio.edu/economicdevelopment/ONED/ONED_Method_Doc.pdf.
- Knaup, Amy E. 2005. "Survival and Longevity in the Business Employment Dynamics Data," *Monthly Labor Review* 128 (5): 50–56.
- Knaup, Amy E., and Merissa C. Piazza. 2007. "Business Employment Dynamics Data: Survival and Longevity, II," *Monthly Labor Review* 130 (9): 3–10.
- Talan, David, and David R. Hiles. 2007. "Quarterly Census of Employment and Wages/Business Employment Dynamics: Data Overview." Presentation prepared for the 2007 Kauffman Symposium on Entrepreneurship and Innovation Data, Washington, DC, November 23.
- Yamoah, Afia, Ziona Austrian, and Joel Elvery. 2009. New Establishment Dynamics: Business Formation and Survival Trends in Ohio. Unpublished manuscript. Available at http://urban.csuohio.edu/economicdevelopment/ONED/ONED_Ohio_Trends.pdf.

Additional Reading

- Sadeghi, Akbar, James R. Spletzer, and David M. Talan. 2009. "Business Employment Dynamics: Annual Tabulations," *Monthly Labor Review* 132 (5): 45–56.