MATHEMATICA
CENTER FOR
STUDYING DISABILITY POLICY

DRC Working PAPER

WORKING PAPER NUMBER: 2016-03

Accounting for Geographic Variation in DI and SSI Participation

March 2016

Jack Gettens*
Pei-Pei Lei
Alexis Henry

*Corresponding author's contact information:
Center for Health Policy and Research
University of Massachusetts Medical School
333 South Street
Shrewsbury, MA 01545
Email: jack.gettens@umassmed.edu
Reference Number: 40112.D-MP-15-11

The research reported herein was performed pursuant to a grant from the U.S. Social Security Administration (SSA) funded as part of the Disability Research Consortium. The opinions and conclusions expressed are solely those of the authors and do not represent the opinions or policy of SSA or any agency of the Federal Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of the contents of this report. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply endorsement, recommendation or favoring by the United States Government or any agency thereof.

ACKNOWLEDGEMENTS

The authors would like to thank Lynn Fisher and Nancy Early for providing data on the number of combined DI and SSI participants by county. The authors also thank David Stapleton; his thoughtful suggestions greatly improved the paper.

CONTENTS

TABLES

FIGURES

ABSTRACT

Background

There is wide geographic variation in DI/SSI participation among the working-age population. The reasons for the variation are not known. The geographic variation garners interest because the variation may possibly identify factors that affect DI/SSI participation that are not apparent from studies of individual-level data. The variation is comprised of two components, variation in disability prevalence (disability component) and variation in DI/SSI participation among persons with disabilities (participation component).

Objective

To account for the geographic variation in DI/SSI participation.

Methods

We use SSA administrative data and American Community Survey data to estimate the geographic variation in DI/SSI participation and the variation in the components. Descriptive statistics and thematic maps are used to describe the variation. We decompose the variance in DI/SSI participation into the two component variances. We use regression methods to examine the association between the participation component and area-level socioeconomic characteristics using exogenous predictors. Principal components analysis is used to decompose the variance of the participation component into the variance contributions of area-level characteristics.

Results

The variances of DI/SSI participation are greater than the sum of the component parts because of correlated components. At the state level, the disability component is 54% and the participation component is 21% of total variance in DI participation; and for SSI, the disability component is 35% and the participation component is 47% of total variance. The sub-state level results are consistent with the state-level results. Variance in the DI participation component is decomposed into the variance contributions of exogenous area-level characteristics as follows: demographics (17%), labor market (15%), disability types (7%), other (11%), and unaccounted (50%). Variance in the SSI participation component is decomposed as follows: demographics (20%), public assistance participation (10%), income (8%), labor market (9%), disability types (6%), other (9%), and unaccounted (38%).

Conclusions

Approximately 90% of the geographic variation in DI/SSI participation can be accounted for by the wide geographic variations in disability prevalence and socioeconomic characteristics. The accounting is different for DI compared to SSI. More of the variation in DI participation is accounted for by variation in disability prevalence and less by socioeconomic characteristics compared to SSI. Compared to DI, variation in the characteristics associated with economically disadvantaged areas accounts for more of the variation in SSI participation.

I. INTRODUCTION

There is wide geographic variation in Supplemental Security Income (SSI) and Social Security Disability Insurance (DI) participation when measured as a percentage of the working-age population. SSI participation ranges from 0.1% of the working-age population in Pitkin County, Colorado to 21% in Owsley County, Kentucky.[1] DI participation ranges from 0.4% of the working age population in Aleutians West County, Alaska, to 21% in Buchanan County, Virginia. The reasons for the variation are not known. The geographic variation garners interest because the variation is wide and it may identify factors that affect DI/SSI participation that are not apparent from studies of individual-level data. For example, it may be an indication of geographic differences in labor market opportunities, disability determination services, or access to employment support or financial assistance programs. Also, participation in DI has been growing rapidly, approximately doubling in the past 30 years (Daly, 2013). A substantial majority of the growth can be attributed to change in the size and age/sex composition of the labor force; however, the remainder of the increase is unexplained (Daly, 2013; Liebman, 2015). It is possible that explanations for the geographic variation in participation may provide insight into the increases in DI participation.

Participation in DI/SSI is contingent on severe disability. Hence, conceptually participants only exist within the subpopulation of persons with disabilities and not within the subpopulation of persons without disabilities. Thus, the geographic variation in DI (or SSI) participation is the composite of two sources of variation, the geographic variation in the prevalence of disability and the geographic variation in DI (or SSI) participation among persons with disabilities. This is

[1] Author's calculations. See Appendix for sources. Counties with fewer than 1000 persons aged 18 to 64 were not included.

described mathematically in the *Methods Section* below. We use SSA administrative data and American Community Survey data to estimate the following: (1) how much of the geographic variation in participation is a result of variation in disability prevalence and how much is a result of variation in DI/SSI participation among persons with disabilities, (2) the correlation between the geographic variation in disability prevalence and the variation in DI/SSI participation among persons with disabilities, and (3) how much of the variation in DI/SSI participation among persons with disabilities is associated with variation in socioeconomic characteristics.

The decomposition accounts for nearly all of the geographic variation in DI/SSI participation. To the best of our knowledge, this is first research that examines the geographic variation in DI/SSI participation using the insight that the variation in DI/SSI participation is the composite of variation in disability prevalence and variation in DI and SSI participation among persons with disabilities. This insight supports the decomposition of the variance into its component parts. The variance decomposition estimates are important because of their policy implications. The estimates quantify how much of the variation may potentially be addressed by policies that affect geographic variation in disability and how much may potentially be addressed by policies that affect geographic variation in participation among persons with disabilities. The decomposition also highlights the need for future research to explain the variation in disability prevalence.

This paper proceeds as a follows. The *Methods Section* describes the relationship between the variance in DI (and SSI) participation and the variances of disability prevalence and DI (and SSI) participation among persons with disabilities. As will be discussed, these variance relationships have implications for future research designed to explain the geographic variation in DI/SSI participation. The *Methods Section* also describes the variance decomposition measures and their sensitivity to bias. Persons with disabilities may be more likely to report their disability when participating in DI/SSI and this could bias the estimates. The *Result Section* includes statistics and thematic maps that describe the geographic variation in SSI and DI participation, disability prevalence, and SSI and DI participation among persons with disabilities. This section also includes the variance decomposition estimates and an assessment of possible bias in the decomposition estimates. The *Discussion, Limitations, and Conclusions Sections* describe the findings, their implications and the limitations of the study.

II. METHODS

In this section we describe the methods we used to decompose the variance in DI/SSI program participation into two components, the variance related to disability prevalence and the variance related to program participation among persons with disabilities. Variance decomposition is generally done for circumstances where the total variance is equal to the sum of the component variances. In the case of DI and SSI program participation, the total variance is the variance of a product (total = program participation x prevalence) and therefore, the total may not be equal the sum of the parts. In the paragraphs below, we define variance decomposition measures and describe their characteristics.

We use the term 'DI/SSI participation' to refer generally to either the DI program or the SSI program. We refer to the *combined* participation rate as the rate based on DI and/or SSI participants.

A. Basis of Variance Decomposition

DI/SSI participation in a geographic area is defined by equation 1.

$$pssa_g = \frac{nssa_g}{ntotal_g} \qquad (1)$$

In equation 1, g indexes geographic areas, $pssa_g$ is the DI/SSI participation rate, $nssa_g$ is the number of DI/SSI participants and $ntotal_g$ is the total number of working-age persons.

The DI/SSI participation rate among persons with disabilities is defined by equation 2.

$$pssad_g = \frac{nssa_g}{ndisability_g} \qquad (2)$$

In equation 2, $pssad_g$ is DI/SSI participation among persons with disabilities and $ndisability_g$ is the number of working-age persons with disabilities.

Disability prevalence is defined by equation 3.

$$pdisability_g = \frac{ndisability_g}{ntotal_g} \qquad (3)$$

In equation 3, $pdisability_g$ is disability prevalence.

By definition, $pssa_g$ is equal to the product of $pdisability_g$ and $pssad_g$ (see equation 4). The variance relationship is given by equation 5.

$$pssa_g = pdisability_g pssad_g \qquad (4)$$

$$Var(pssa) = Var(pdisabilitypssad) \qquad (5)$$

As equation 5 indicates, the variance of *pssa* is the variance of the product of *pdisability* and *pssad*. The variance of *pssa* is dependent on the variance of *pdisability*, the variance of *pssad*, and the correlation between *pdisability* and *pssad*. It is our objective to determine the relative contributions of *pdisability* and *pssad* to the variance of *pssa*. To facilitate this, we use the natural log transformation of Equation 4, which is additive (see Equation 6). The variance relationship is equation 7.

$$\ln(pssa) = \ln(pdisability) + \ln(pssad) \qquad (6)$$

$$\mathrm{Var}[\ln(pssa)] = \mathrm{Var}[\ln(pdisability)] + Var[\ln(pssad)] +$$
$$2Cov[\ln(pdisability), \ln(pssad)] \qquad (7)$$

We define two variance decomposition measures. The first, *Percent Variance Disability* ($PV_{disability}$), is the percent of variance in ln(*pssa*) that would exist conditional on the variance of ln(*pssad*) being zero (see equation 8). The second, *Percent Variance Participation* (PV_{ssad}), is the percent of variance in *pssa* that would exist conditional on the variance of *pdisability* being zero (see equation 9).

$$Percent\ Variance\ Disability\ (PV_{disability}) = \frac{Var[\ln(pdisability)]}{Var[ln(pssa)]}\ x\ 100 \ (8)$$

$$Percent\ Variance\ Participation\ (PV_{ssad}) = \frac{Var[\ln(pssad)]}{Var[ln(pssa)]}\ x\ 100 \quad (9)$$

B. Characteristics of Variance Decomposition Measures

In this section we describe the characteristics of the variance decomposition measures. The information in this section supports the interpretation of the variance decomposition estimates described in the *Results Section* below. Specifically, we describe how the correlation between *pdisability* and *pssad* and the relative dispersions of *pdisability* and *pssad* affect the variance decomposition estimates. As we describe below, the component with the greatest relative dispersion accounts for more of total variance.

We start with the case where *pdisability* and *pssad* are uncorrelated. The coefficient of variation (CV) is a measure of the standardized dispersion of a distribution and is defined as the ratio of the standard deviation to the mean. As a first-order approximation, Var[ln(pdisability)] is equal to the coefficient of variation squared of the untransformed *pdisability*. A comparable relationship exists for CV^2_{pssad}. Thus for the case where *pdisability* and *pssad* are uncorrelated, equation 7, 8, and 9 may be expressed in terms of CVs (see equations 10, 11 and 12).[2]

$$Var[\ln(pssa)] = CV^2_{pdisability} + CV^2_{pssad} \quad (10)$$

$$PV_{disability} = \frac{CV^2_{disability}}{CV^2_{disability}+CV^2_{pssad}} \quad (11)$$

$$PV_{participation} = \frac{CV^2_{pssad}}{CV^2_{disability}+CV^2_{pssad}} \quad (12)$$

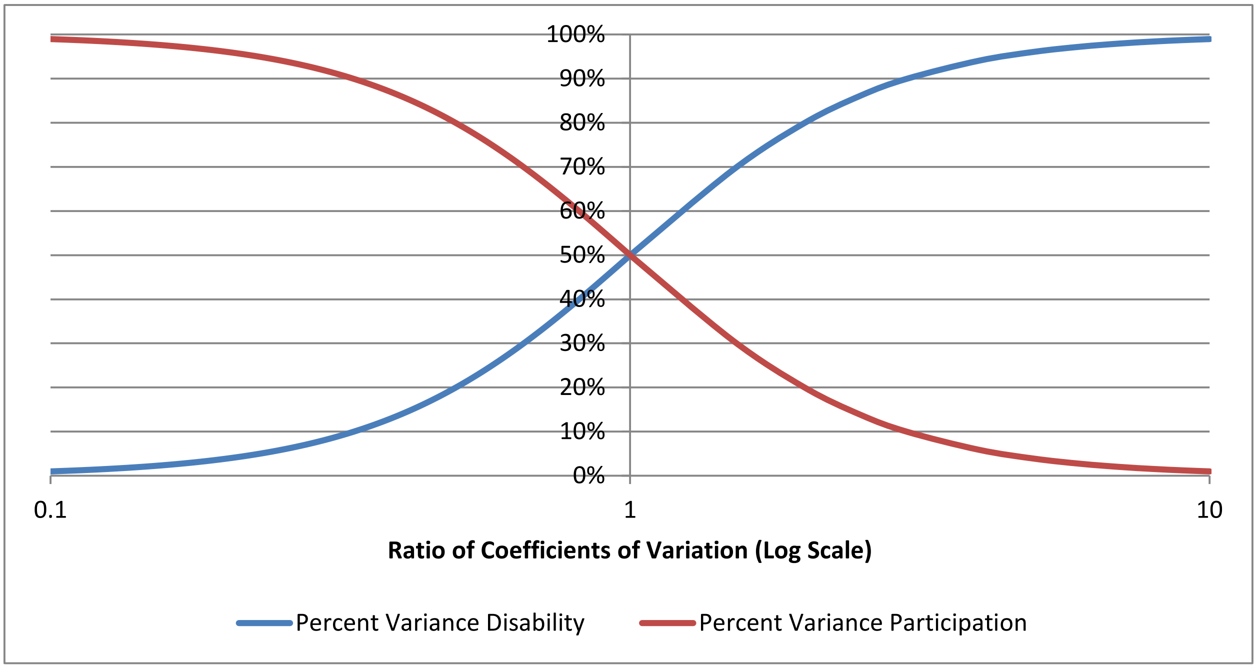
As is evident from equations 11 and 12, the magnitudes of $PV_{disability}$ and PV_{ssad} are dependent on the ratio of $CV_{pdisability}$ and CV_{pssad} (RCV) as shown in Figure 1.[3] $PV_{pdisability}$ increases with increasing RCV and there is a complementary relationship between $PV_{disability}$ and PV_{ssad}. Increases in $PV_{disability}$ correspond to decreases in PV_{ssad} and vice versa. For this case

[2] Because of the approximation that Var[ln(pdisability)] equals $CV^2_{pdisability}$. The relationships of equation 11 and 12 approximate the decomposition relationships based on equation 4. The approximation is derived from the first-order Taylor expansion for the variance of non-linear function.

[3] For Figure 1, $CV_{pdisability}$ is assumed constant, 0.3 and CV_{pssad} varies from 3.0 to .03.

where *pdisability* and *pssad* are uncorrelated, the sum of $PV_{disability}$ and PV_{ssad} is approximately 100%. For the circumstance where the standardized dispersions are equal ($CV_{pdisability} = CV_{pssad}$), $PV_{disability}$ and PV_{ssad} are equal with values of approximately 50%.

Figure 1. Relationship between percent variance disability, percent variance participation and RCV; zero correlation case



For the case where *pdisability* and *pssad* are correlated, the sum of the variance decomposition measures is not equal to 100%. For the case of positive correlation, the sum $PV_{disability}$ and PV_{ssad} will be less than 100% and the sum decreases with increasing positive correlation. The sum will be greater than 100% for the case of negative correlation and the sum increases with increasing negative correlation. In essence, the vertical scale of Figure 1 contracts for positive correlation and expands for negative correlation without corresponding contraction or expansion of the line graphs. The variance decomposition characteristics for various correlation and RCV conditions are summarized in Table 1.

Table 1. Variance decomposition characteristics, correlation and RCV conditions

Correlation between *pdisability* and *pssad*	RCV	Comparison of $PV_{pdisability}$ and PV_{pssad}	Sum of $PV_{pdisability}$ and PV_{pssad}
None	RCV > 1	$PV_{pdisability} > PV_{pssad}$	≈100%
None	RCV < 1	$PV_{pdisability} < PV_{pssad}$	≈100%
Positive	RCV > 1	$PV_{pdisability} > PV_{pssad}$	< 100%
Positive	RCV < 1	$PV_{pdisability} < PV_{pssad}$	< 100%
Negative	RCV > 1	$PV_{pdisability} > PV_{pssad}$	>100%
Negative	RCV < 1	$PV_{pdisability} < PV_{pssad}$	>100%

C. Effects of Measurement Error

In this section we describe how measurement error may bias the variance decomposition estimates. This information supports our assessment of possible bias in the variance decomposition estimates (see *Results Section*). We use SSA administrative data and Census Bureau data to determine $pssa_g$ (see equation 1) and we assume that the measurement error $pssa_g$ is small. Survey data was used to estimate the number of person with disabilities ($ndisability_g$) and we assume estimates for $pssad_g$ and $pdisability_g$ are measured with error (see equations 2 and 3). For example, the measurement error in *pdisability* is described by equation 13.

$$pdisability_g = \pi pdisability_g^* + \nu_g \qquad (13)$$

In this equation, $pdisability_g*$ is the true disability prevalence without error, π is error that is consistent across geographic areas, and ν_g is error that varies across geographic areas. The measurement error in *pdisability* may bias some or all of the following estimates: $PV_{disability}$, PV_{ssad}, the sum of the $PV_{disability}$ and PV_{ssad}, and the correlation between *pdisability* and *pssad*.

We examine the possible bias effects of three types of *pdisability* error: (a) error that biases the mean of *pdisability* consistently across geographic areas (π), (b) random error, and (c) non-random error.[4] Random error is error in *pdisability* that is not correlated with *pssa* (v_g *not correlated with* $pdisability_g$). Non-random error is error in *pdisability* that is correlated with *pssa* (v_g *correlated with* $pdisability_g$). The effects of these types of error on the variance decomposition estimates are summarized in Table 2.

Error that is consistent across geographic areas ($\pi \neq 1$) biases the mean and variances of *pdisability* and *pssad.* However, this error does not bias the CVs and thus, the variance decomposition estimates are not biased (see Figure 1).

Random error increases the variance of *pdisability* and *pssad* and this results in overestimates of $PV_{disability}$ and PV_{ssad}. We show empirically in the Results Section that because the biases in $PV_{disability}$ and PV_{ssad} are both in the same direction, random error does not substantially bias comparisons between $PV_{disability}$ and PV_{ssad}. Random error also biases estimates of the correlation between *pdisability* and *pssad* toward negative correlation (see equations 2 and 3).

The bias associated with non-random error is more complex and is dependent on whether the error is positively or negatively correlated with *pssa* and whether RCV is greater than or less than 1. A positive correlation between the error and *pssa* results in overestimates of $PV_{disability}$ and underestimates of PV_{pssad}. Because these biases are in opposite directions, this will also bias a comparison between $PV_{disability}$ and PV_{ssad}. An error that is positively correlated with *pssa* may either positively or negatively bias the correlation between *pdisability* and *pssad*. If the RCV is greater than 1, the correlation is biased in the positive direction; if the RCV is less than 1, the

[4] To simplify the description, we only describe errors in *pdisability*. Because the errors in both *pdisability* and *pssad* are the result of the error in *pdisability*, the error types and the descriptions also apply to errors in *pssad*.

bias is in the negative direction. The bias effects for a negative correlation between the error and *pssa* are described in Table 2.

Table 2. Effects of bias in **pdisability** on variance decomposition estimates

pdisability Bias Type			Bias in Variance Decomposition Estimates			
Mean, RCV not biased	Random Error	Non-Random Error	$PV_{disability}$	PV_{pssa}	Sum of $PV_{disability}$ and PV_{pssa}	Correlation *pdisability, pssad*
Overestimate	None	None	None	None	None	None
Underestimate	None	None	None	None	None	None
None	Yes	None	Overestimate	Overestimate	Overestimate	Negative
None	None	Positively Correlated with *pssa*, RCV >1	Overestimate	Underestimate	Overestimate	Negative
None	None	Positively Correlated with *pssa*, RCV <1	Overestimate	Underestimate	Underestimate	Positive
None	None	Negatively Correlated with *pssa*, RCV >1	Underestimate	Overestimate	Underestimate	Positive
None	None	Negatively Correlated with *pssa*, RCV <1	Underestimate	Overestimate	Overestimate	Negative

Positive indicates a bias toward positive correlation. Negative indicates a bias toward negative correlation

D. Explaining the Variation in Disability Prevalence and DI/SSI Participation

The functional form of equation 4 (above) has implications for the methods used in research to explain the variation in *pssa*. In this section we describe why separate analyses of the variation of *pdisability* and *pssad* are preferable to the direct analysis of variation in *pssa*. Equation 4 implies there is a process generating variation in *pdisability* and also a process generating variation in *pssad.* For illustrative purposes, we assume the data generating processes are of the forms represented by equations 14 and 15.

$$pdisability_g = \beta X_g \qquad (14)$$

$$pssad_g = \alpha Z_g \qquad (15)$$

In equation 14, X_g is a vector of variables affecting *pdisability* and β is a vector of their effects. In equation 15, Z_g is a vector of variables affecting *pssad* and α is a vector of their effects. Some variables may affect both *pdisability* and *pssad* and some variables may separately affect only *pdisability* or *pssad*. For a variable that affects both, the magnitude and sign of the effect on *pdisability* may be different from the magnitude and sign of the effect on *pssad*. For variables that separately only affect *pdisability* or *pssad*, they may differ in their degree of correlation to other variables.

Using a research method that examines the variation in *pssa* directly, for example by using regression methods to estimate the parameters of equation 16, will reveal associations; however, the associations will not reveal the effects of the separate data generating processes (see equations 14 and 15). It will be impossible to know whether the associations occur because of the effects on *pdisability* or the effects on *pssa* or some combination of both. In addition, it is possible that some factors may have large effects on both *pdisability* and *pssad* and minimal effects on *pssa* because of negative correlation. These effects on *pdisability* and *pssad* would not be evident from estimates of equation 16. The shortcomings of this method would limit researchers' ability to explain the variation in DI/SSI participation, to interpret the results and to assess policy implications.

Separate analysis of *pdisability* and *pssad*, for example by using regression methods to separately estimate equations 17 and 18, would be more informative. With this approach, we expect that the explanations of the variation in DI/SSI participation, the interpretation of results and assessment of policy implications will be more straightforward. Estimates of equation 18 also have an additional advantage compared to estimates of equation 16. Area characteristics (Z_g) can readily be identified that are exogenous. For example, an estimate of the association between $pssad_g$ and poverty could be made using an estimate of the poverty level of persons without disabilities as an exogenous proxy of area-level poverty. Estimates of equation 16 are generally made with measures of area characteristics that are endogenous, for example the overall poverty rate of persons with and without disabilities. Because this measure of poverty is affected by DI/SSI participation, the estimates of the association between $pssad_g$ and poverty would be biased.

$$pssa_g = \gamma Y_g + \varepsilon_g \qquad (16)$$

$$pdisability_g = \beta X_g + \eta_g \qquad (17)$$

$$pssad_g = \alpha Z_g + \zeta_g \qquad (18)$$

In this study, we estimate equation 18 at the local-area level to include sub-state variation and to provide a sufficient number of observations to support a range of explanatory factors. This study does not include estimates of equation 17 which we have deferred to future research.

There are numerous area-level variables (vector Z_g in equation 18) that are expected to be associated with DI/SSI participation (Coe et al., 2011). We use ACS data to estimate the area-level variables and we group these variables into the following categories: (a) demographics, (b) disability, (c) income and poverty, (d) labor market, and (f) public assistance and health insurance (see Table 3). Age and income are measured as means and the remaining variables as proportions. The variables are based on persons with disabilities when the characteristic is exogenous (e.g. gender). The variables are based on the population without disabilities when we expect that the characteristic based on the full population of persons with and without disabilities is endogenous (e.g. poverty, labor force participation rate). We expect that the area-level conditions reflected in characteristics based on persons without disabilities are also experienced by persons with disabilities. For example a high labor force participation rate among persons without disabilities is likely an indication of the area labor market conditions experience by both persons with and without disabilities.

Table 3. Local-area characteristics and population basis

Category	Characteristic	Population Basis
Demographics		
	Average	With disabilities
	Female	With disabilities
	Never married	With disabilities
	High school or less	With disabilities
	Hispanic	With disabilities
	Black	With disabilities
	Non-English at home	With disabilities
	Native born	With disabilities
	U.S. citizen	With disabilities
Disability		
	Self-care difficulty	With disabilities
	Hearing difficulty	With disabilities
	Vision difficulty	With disabilities
	Independent living difficulty	With disabilities
	Ambulatory difficulty	With disabilities
	Cognitive difficulty	With disabilities
Income and Poverty		
	Below 100% federal poverty	Without disabilities
	Annual income	Without disabilities
	Annual household income	Without disabilities
	Annual earned income	Without disabilities
Labor Market		
	Male labor force participation	Without disabilities
	Female labor force participation	Without disabilities
	Self-employment	Without disabilities
	Usual hours worked per week	Without disabilities
	Worked 26 weeks or less among workers	Without disabilities
	Service occupations	Without disabilities
	Production occupations	Without disabilities
	Sales occupations	Without disabilities
	Construction and maintenance occupations	Without disabilities
	Management occupations	Without disabilities
	Manufacturing industry	Without disabilities
	Education and health services industry	Without disabilities
	Prof. and business services industry	Without disabilities
	Other industries	Without disabilities
	Wholesale and retail trade industry	Without disabilities
	Leisure and hospitality	Without disabilities
Public Assistance and Health Insurance		
	Public Assistance	Without disabilities
	Health insurance	Without disabilities

Note. Other industries includes agricultural and related industries; mining, quarrying, and oil and gas extraction; construction; transportation and utilities; information; financial activities; other services and government workers.

We estimate two versions of equation 18. The first is estimated with the constraint that the intercept is constant across states. The second does not assume a constant intercept across states and is estimated with state-specific intercepts (fixed effects model). The fixed effects model accounts for between-state variation that is not accounted for by the variation in local area characteristics (Table 3). The between-state variation may exist for two reasons. The first is unobserved factors that vary between states because they are determined by state policy, for example, insurance regulation or DI/SSI disability determination services. The second is unobserved factors that vary between states but are not determined by state policy, possible examples include stigma, discrimination, and attitudes about employment.

The estimates of equation 18 (vector α) provide associations; however, because the regression equation does not include all factors associated with SSI participation, these estimates cannot be interpreted as causal effect estimates. It is possible that the association is biased by correlations of the predictors with unobserved factors that independently affect DI/SSI participation.

E. Principal Components Analysis

The regression analysis (equation 18) will provide estimates of the percentage of variation in DI/SSI participation among persons with disability that is accounted for, in total, by the characteristics listed in Table 3. However, because many characteristics are correlated, regression does not provide an estimate of the variance attributed to individual characteristics. To decompose the variance into mutually-exclusive components, we use principal components analysis.

Principal components analysis is a procedure that transforms the variables into a set of components that are uncorrelated. The number of components is generally considerably smaller than the original number of variables without substantially reducing the total variation of the original variables. Each component is a linear combination of the original variables. The linear combinations are determined to maximize the variance contribution among ordered components. For example, the variance contribution of the first component is maximized based on the total variance and the variance contribution of the second component is maximized based on remaining variance. This allows for the selection of principal components that capture most of the variance while reducing the number of analysis variables (components). Because the components are uncorrelated, regression may be used to determine the variance contribution of each principal component to the variance of the outcome variable (e.g. SSI participation among persons with disabilities). The interpretation of each component is determined by assessing the correlation between the components and the original variables.

We determine the principal components of the variables listed in Table 3. Principal components analysis methods are not appropriate for models including categorical variables and we do not include the state fixed effects in the analysis. We use regression to estimate the association between the DI/SSI participation and the principal components (see Equation 19).

$$pssad_g = \beta C_g + \varepsilon_g \qquad (19)$$

In equation 19, C_g is a vector of the area-level principal components and β is a vector of the associations between $pssad_g$ and the principal components. Because the components are uncorrelated, the variance contribution of a component to the variance of $pssad_g$ is determined by the square of the correlation between $pssad_g$ and the component.

F. Potential Bias in Regression and Principal Components Analysis

The regression estimates of equation 18 and the variance decomposition estimates may be vulnerable to bias. DI/SSI participation among persons with disabilities is estimated with error because it is based on ACS-based disability prevalence estimates. Similarly, the area-level characteristics are also ACS-based and measured with error. Thus, the measurement error exists in both the dependent variable and independent variables of equation 18. The bias associated with the measurement error will depend on whether the dependent-variable measurement error is correlated with the measurement errors of the independent variables; in our case, whether the error in DI/SSI participation among persons with disabilities is correlated with the error in area characteristics. If the errors are uncorrelated, the estimates of the regression coefficients will be biased toward zero (Hyslop and Imbens, 2001) and the variance accounted for by the independent variables will also be biased toward zero.

The bias for the case where errors are correlated is dependent on whether the correlation is positive or negative.[5] Depending on the correlation, it is possible for either over- or underestimation of the regression coefficients and the variance contribution estimates. We are unable to determine if the errors are correlated; however, correlation is a possibility. For example, an under-reporting of disability prevalence among Hispanics would result in error in the estimated DI/SSI participation among persons with disabilities that is correlated with the error in the estimated proportion of Hispanics among persons with disabilities.

In summary, we expect there to be measurement error in both the dependent and independent variables. We do not have data to determine whether the errors are correlated and we are not able to determine the direction of the bias. To assess whether the findings of the regression and principal components analysis may be an artifact of bias, we assess the potential error in disability prevalence (See Results Section).

G. Data

We used data from the 2009-2011 American Community Survey (ACS) Public Use Microdata Sample to estimate the number of persons with disabilities, working-age (18-64) population counts across states and Public Use Microdata Areas (PUMAs), and area socioeconomic characteristics. The 2009-2011 time period was chosen because the sub-state geographic boundaries and the disability questions were consistent during the period. We do not include individuals living in institutional group quarters because

[5] See Hyslop and Imbens, 2001 for a description of the biases in the two variable case.

DI/SSI participation is precluded for a large majority of the group quarters population, those who are incarcerated, and the data do not allow us to differentiate that population for those living in other institutional group quarters, such as nursing homes. .

We used the 2009-2011 ACS summary table, S1810, “Disability Characteristics” estimates for county disability prevalence. These estimates are based on the civilian noninstitutionalized population. Disability prevalence is not publicly available for counties with less than 10,000 persons. The 2009-2011 S1810 table includes disability prevalence estimates for 1844 counties.

To determine the number of persons with disabilities, an ACS respondent is considered disabled if she answered ‘yes’ to any of the following questions:

- Is this person deaf or does he/she have serious difficulty hearing?
- Is this person blind or does he/she have serious difficulty seeing even when wearing glasses?
- Because of a physical, mental, or emotional condition, does this person have serious difficulty concentrating, remembering or making decisions?
- Does this person have serious difficulty walking or climbing stairs?
- Does this person have difficulty dressing or bathing?
- Because of a physical, mental, or emotional condition, does this person have difficulty doing errands alone such as visiting doctor’s office or shopping?

We used Social Security Administration administrative data to determine the number of DI, SSI, and combined participants in states and counties. These data sources are described in Appendix Table A1. DI participants include disabled workers but do not include disabled widows or disabled adult children because disabled widows and disabled adult children data was not available across geographic areas. In 2011, there were approximate 8.5 million disabled workers and one million disabled widows and disabled adult children (SSA, 2015). SSI recipients include both federal SSI and federally administered state supplementation. Data was not available to separately identify federal SSI recipients and federally administered state supplementation only recipients across geographic areas. In December 2010, there were approximately 6.5 million federal SSI recipients and 167 thousand state supplementation-only recipients (SSA, 2012). DI/SSI participation (DI, SSI, or combined) was defined as the ratio of the number of participants and the population count.

We use public health data to assess possible bias in the variance decomposition estimates. The Centers for Disease Control and Prevention WONDER online database was used to determine mortality rates.[6] The 2011 Behavioral Risk Factor Surveillance System (BRFSS) is the data source for diabetes prevalence, proportion with poor or fair health, and smoking rates.[7]

H. Limitations of ACS Disability Estimates

Our analysis assumes that DI and SSI participants exist within the subpopulation of persons with disabilities (see equation 4). This is not necessarily the case when using ACS survey data to identify persons with disabilities. Using Current Population Survey (CPS) data matched with SSA administrative data, Burkhauser et al. (2012) found that approximately 66 percent of DI and SSI recipients are captured by the ACS questions as administered within the CPS. This suggests that ACS-based disability prevalence estimates and the estimates for DI/SSI participation among persons with disabilities may be biased.[8] We assess this in the *Results Section* below.

I. Geographic Regions

We conduct our analysis on four geographic levels: states, counties, Public Use Microdata Areas (PUMAs) and county-aligned-PUMAs (CAPUMAs) (see Table 4). PUMAs are within-state regions of approximately 100,000 to 200,000 people. We define CAPUMAs as the smallest

[6] http://wonder.cdc.gov/

[7] http://www.cdc.gov/brfss/annual_data/annual_2011.htm

[8] See Ben-Shalom & Stapleton (2014) for additional information on the bias in estimates of SSI and DI participation among persons with disabilities.

areas that contain both complete PUMAs and complete counties. The relationships between PUMAs, counties and CAPUMAs are listed in Table 5.

Disability prevalence data is available on state, CAPUMA and PUMA levels and also for a subset of counties. DI/SSI participation data granularity is available at the state and county levels. [9] We determine DI/SSI participation at the CAPUMA level by combining PUMA and county-level data. We conduct the decomposition analyses at the state and CAPUMA levels and for the subset of counties where both disability prevalence and DI/SSI participation data are available. We conduct the analysis of the association between SSA participation and area-level characteristics at the CAPUMA level. In addition, we estimate disability prevalence statistics at the PUMA level.

Table 4. Geographic areas of analysis

Geography	Description	18 to 64 Population Range	Number	SSA Program Participation Data	Disability Prevalence Data
States and DC		357,431 to 23,767,298	51	Yes	Yes
Counties	Administrative division within-states	52 to 6,372,275	3142	Yes	Subset (1844 counties)
Public Use Microdata Areas	Within-state areas of approximately equal population size	24,729 to 205,129	2069	No	Yes
County Aligned Public Use Microdata Areas	Smallest areas that contain both complete PUMAs and Counties	50,200 to 6,329,545	937	Derived from county data	Yes

Table 5. Relationships between PUMAs, Counties and CAPUMAs

Relationship between PUMA and County	CAPUMA relationship to PUMA	CAPUMA relationship to County
One PUMA equals one county	CAPUMA equals PUMA	CAPUMA equals County
One PUMA equals many counties	CAPUMA equal PUMA	CAPUMA equals Counties
Many PUMAs equals one county	CAPUMA equals PUMAs	CAPUMA equals County
Many PUMAs equals many counties	CAPUMA equals PUMAs	CAPUMA equals Counties

[9] SSI participation rates are not available for 288 counties. To calculate CAPUMA statistics, we imputed SSI participation rates as the state means for these counties.

A. Geographic Variation in DI/SSI Participation

Table 6 shows descriptive statistics for DI/SSI participation among working-age individuals (18-64) at the state, CAPUMA and county levels. The average state combined participation rate is 6.5%. The average state DI participation rate is higher than the SSI participation rate, 4.6% vs. 2.5%. The standard deviation, coefficient of variation (ratio of standard deviation and mean) and percentile statistics all indicate wide variation in participation rates across states, CAPUMAs, and counties. There is considerable within-state variation in DI/SSI participation. Approximately 2/3 of variance is within-state (county or CAPUMA) and 1/3 is between-states.[10]

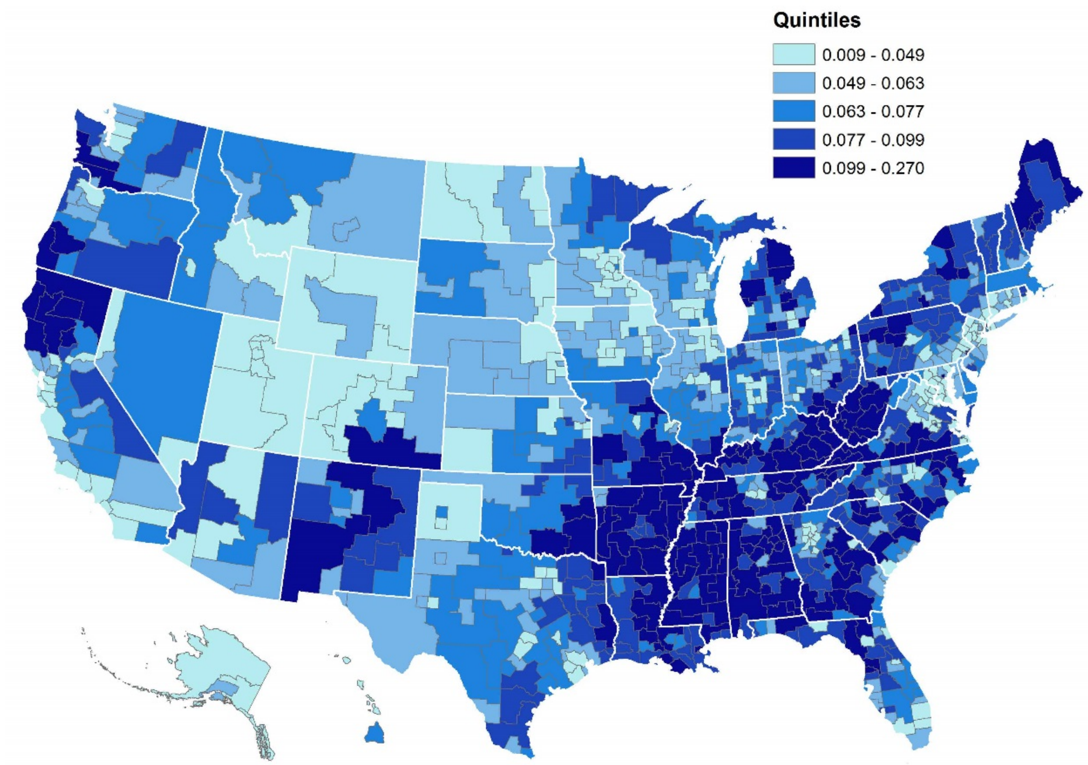
Table 6. Proportion of working-age individuals participating in DI/SSI: state, CAPUMA and county levels

Program	Statistic	States and DC (n=51)	CAPUMA (n=937)	County
SSI				(n=3038)
	Mean	.025	.028	.030
	Standard Deviation	.009	.016	.020
	Coefficient of Variation	.36	.57	.67
	75th Percentile	.029	.035	.039
	Median	.021	.025	.025
	25th Percentile	.018	.017	.017
	Proportion Within-State Variance	N/A	.66	.62
DI				(n=3142)
	Mean	.046	.054	.058
	Standard Deviation	.014	.021	.025
	Coefficient of Variation	.304	.389	.431
	75th Percentile	.054	.065	.071
	Median	.043	.051	.054
	25th Percentile	.036	.039	.040
	Proportion Within-State Variance	N/A	.61	.56
Combined				(n=2854)
	Mean	.065	.076	.085
	Standard Deviation	.020	.033	.040
	Coefficient of Variation	.308	.434	.470
	75th Percentile	.076	.093	.105
	Median	.062	.071	.077
	25th Percentile	.049	.053	.057
	Proportion Within-State Variance	N/A	.61	.59

Note: There is missing data for some small counties. Missing county participation rates are imputed as the state mean participation when determining CAPUMA participation rates.

[10] The proportion of within-state variation was approximated by Analysis of Variance as the ratio of the within-state sum of squares to the sum of the within-state sum of squares and the between-state sum of squares.

Figure 2. Proportion of working-age persons participating in SSI and/or DI, CAPUMA level



The variation in the combined participation rate across CAPUMAs is shown in the thematic map of Figure 2. The categories in Figure 2 are defined by quintiles. The within-state and between-state variation are apparent. Many states contain CAPUMAs ranging from low to high participation illustrating within-state variation. For example in Georgia, the rates are high in southern areas and low in the vicinity of Atlanta. Similarly in Minnesota, the rates are higher in the northern areas and lower in the vicinity of Minneapolis. It is also apparent from the map that the states vary in average program participation indicating between-state variation. For example, Arkansas has high participation with nearly all areas having participation in the 4th and 5th quintiles. In contrast, Utah has low participation with all areas having participation in the 1st quintile. The variations in DI and SSI participation rates across CAPUMAs are shown in the Appendix, Figures A1 and A2.

B. Geographic Variation in Disability Prevalence

Table 7 shows descriptive statistics for disability prevalence among working-age individuals (18-64) at the state, CAPUMA, PUMA, and county levels. The mean state disability prevalence is 10.6%. The mean disability prevalence rates for the other geographic levels are comparable. The statistics indicate wide variation in participation rates across states, CAPUMAs, PUMAs and counties. Approximately 2/3 of the variance is within-state (CAPUMA, PUMA, and county) and 1/3 is between-states.

Table 7. Proportion of Working-Age Individuals with Disabilities: State, CAPUMA and PUMA and County Levels.

Statistic	States and DC (n=51)	CAPUMA (n=937)	PUMA (n=2069)	County (n=1844)
Mean	.106	.120	.103	.127
Standard Deviation	.024	.039	.040	.045
Coefficient of Variation	.226	.325	.388	.354
75^{th} Percentile	.115	.142	.127	.153
Median	.103	.114	.097	.121
25^{th} Percentile	.088	.091	.072	.095
Proportion Within-State Variance	N/A	.64	.70	.67

Note: County-level data does not include counties with fewer than 10,000 persons.

The variation in disability prevalence across PUMAs is shown in the thematic map of Figure 3.[11] The land area of a PUMA is a rough estimate of population density because the range of PUMA populations is relatively narrow (See Table 4) Similar to Figure 2, the existence of wide within-state and between-state variation is apparent. The maps suggest that there are regions of high prevalence that transcend state borders. The region defined by the states of Louisiana, Mississippi, Alabama, Arkansas, Missouri (southern), Tennessee, and West Virginia has high

[11] The CAPUMA-level thematic map of disability prevalence is Figure A3.

disability prevalence. There also appear to be patterns associated with major cities. There are regions of low disability areas near most major cities. For example, areas in the vicinity Atlanta, Boston, Chicago, Dallas, Los Angeles, Minneapolis, and San Francisco and the I-95 corridor from Washington DC to Connecticut (See Figure 4). The major U.S. cities are shown on Appendix Figure A6. Most major cities also have areas of high disability in their vicinity but in some cases, the number of high disability prevalence areas is small compared to the number of low disability areas, for example Washington, DC and New York (See Figure 4). In many urban regions, areas of high and low disability prevalence are in close proximity, for example, the areas including and surrounding Washington, Philadelphia, and New York (see Figure 4).

Figure 3. Proportion of working-age persons with disabilities, PUMA level

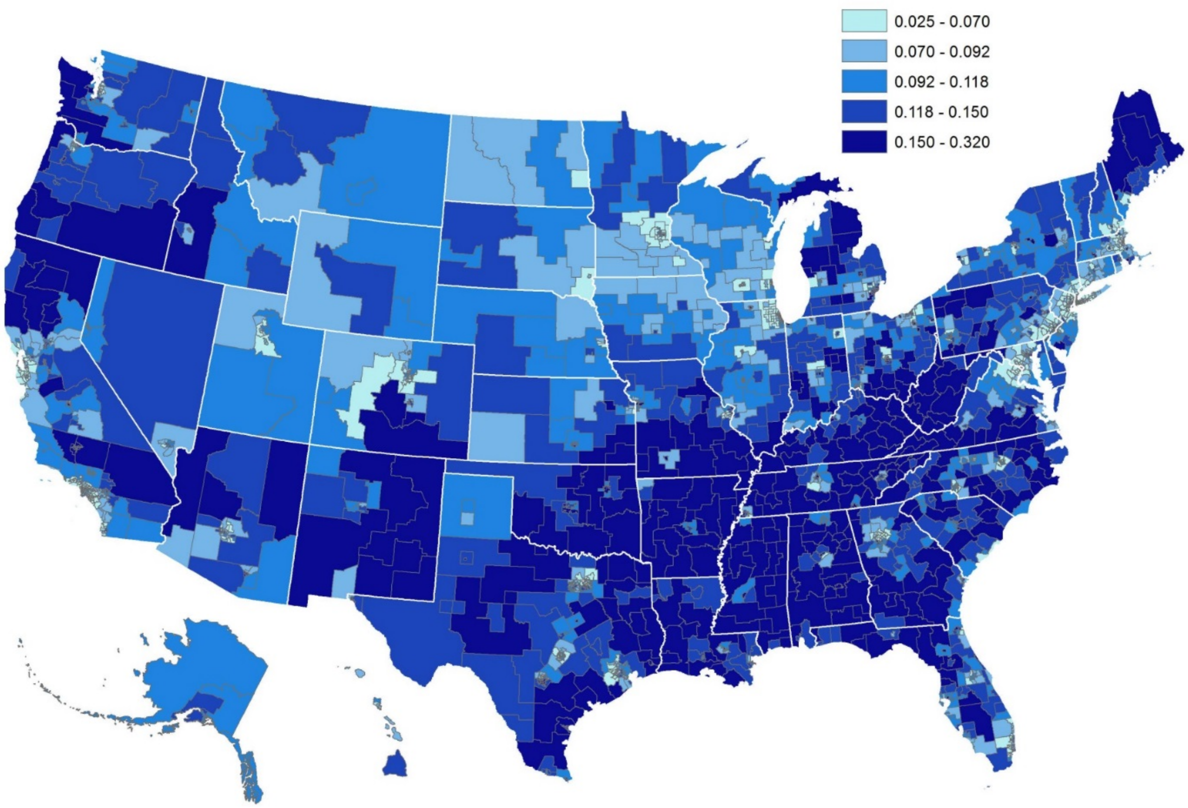
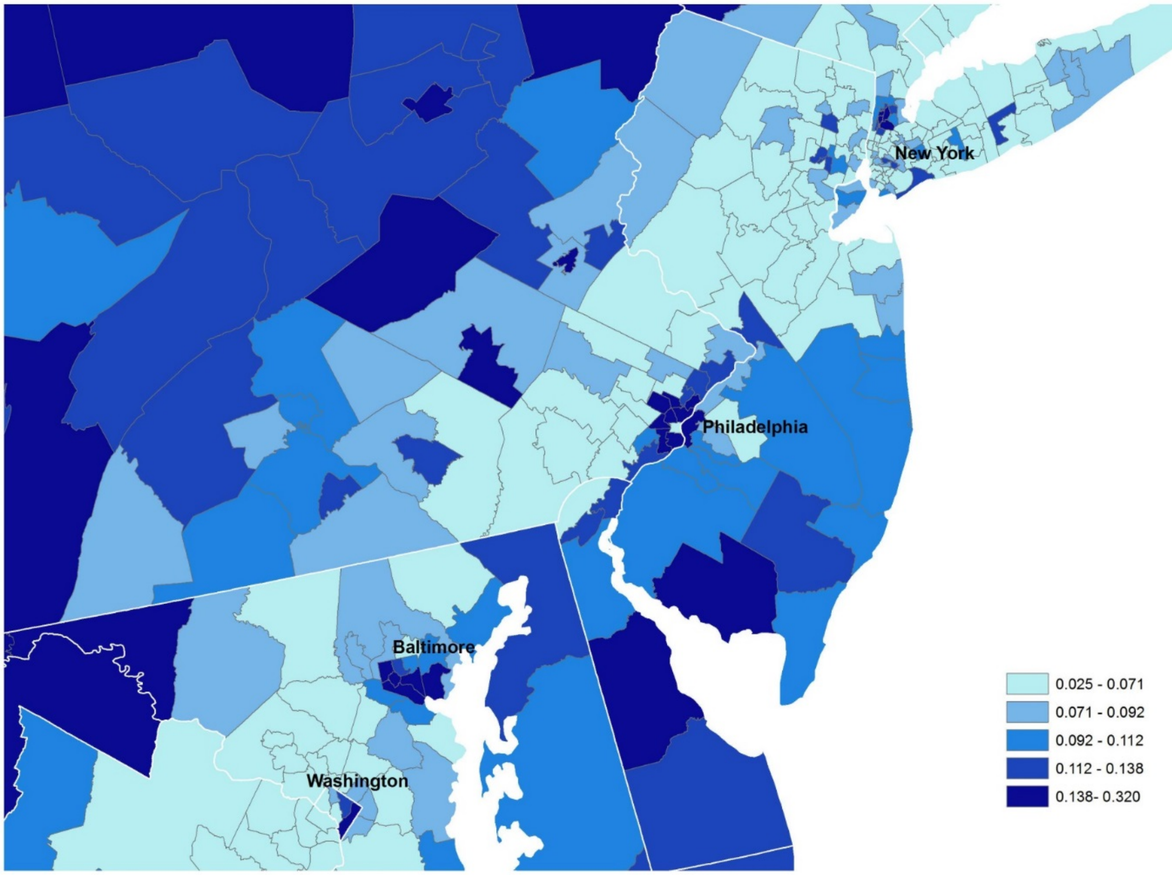


Figure 4. Proportion of working-age persons with disabilities Washington to New York, PUMA level



C. Geographic Variation in DI/SSI Participation among Persons with Disabilities

Table 8 shows descriptive statistics for DI/SSI participation among working-age persons with disabilities (18-64) at the state, CAPUMA and county levels. The average state combined participation rate is 61.5%. The average state DI participation rate is higher than the SSI participation rate, 43.6% vs. 23.1%. The participation rates are comparable at the CAPUMA and county levels. The statistics all indicate wide variation in participation rates across states, CAPUMAs, and counties. The variation is greater across sub-state regions compared to states. Approximately 2/3 of the variance is within-state (CAPUMA, PUMA, and county) and 1/3 is between-states.

Table 8. Proportion of working-age individuals with disabilities participating in DI/SSIs: state, CAPUMA and county levels

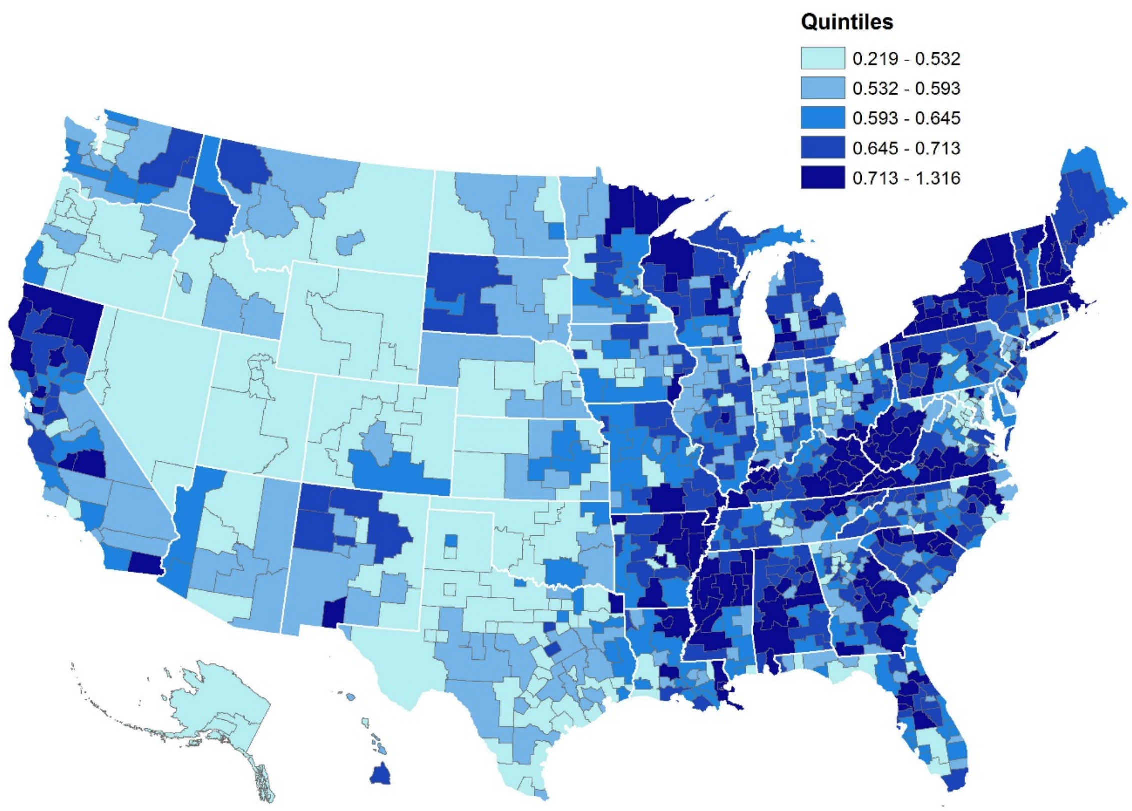
Statistic		States and DC (n=51)	CAPUMA (n=937)	County (n=1844)
SSI				
	Mean	.231	.227	.22
	Standard Deviation	.06	.079	.082
	Coefficient of Variation	.26	.348	.373
	75th Percentile	.261	.263	.263
	Median	.215	.212	.207
	25th Percentile	.189	.174	.164
	Proportion Within-State Variance	N/A	.67	.72
DI				
	Mean	.436	.451	.452
	Standard Deviation	.061	.079	.092
	Coefficient of Variation	.14	.175	.204
	75th Percentile	.481	.502	.514
	Median	.437	.451	.448
	25th Percentile	.403	.398	.389
	Proportion Within-State Variance	N/A	.61	.67
Combined				
	Mean	0.615	0.626	0.622
	Standard Deviation	0.088	0.118	0.134
	Coefficient of Variation	0.143	0.188	0.215
	75th Percentile	0.681	0.693	0.704
	Median	0.614	0.618	0.61
	25th Percentile	0.549	0.553	0.538
	Proportion Within-State Variance	N/A	.64	.71

Note. Participation data is not estimated for counties with fewer than 10,000 persons because disability prevalence is not available for these counties.

The variation in combined participation across CAPUMAs is shown in the thematic map of Figure 5. Similar to Figures 2 and 3, the existence of wide within-state and between-state variation is apparent. With the exceptions of some CAPUMAs within California, New Mexico, Washington, Idaho, Montana, South Dakota, the participation rates are much lower in the west than the east. In this case, the west is the region defined by states to the west of and including the following states: Texas, Oklahoma, Kansas, Nebraska, South Dakota and Northern Dakota.

There is no readily apparent regional pattern of participation in the east. The pattern of geographic variation for DI participation is similar (see Appendix Figure A4). The geographic variation in SSI participation is shown in Appendix Figure A5. There is high participation in portions of Appalachia including: Eastern Kentucky, West Virginia, Southeastern Ohio, Western Pennsylvania and Southwest New York. There are also non-Appalachia states with regions of predominantly high participation, including: New York, California, and Massachusetts. States with regions that are predominantly low in SSI participation include: Nevada, Utah, Colorado, Kansas, Nebraska, and Wyoming.

Figure 5. Proportion of working-age persons with disabilities participating in SSI and/or DI, CAPUMA level



D. Variance Decomposition Estimates

To determine the contributions of the variance of the disability component and the variance of the participation component among persons with disabilities to the total variance of DI/SSI participation, we estimated the variance decomposition measures, $PV_{disability}$ and PV_{ssad}, as described in the *Methods Section*. Table 9 shows the estimates at the state, CAPUMA and county levels. Table 9 also includes estimates of the correlation between disability prevalence (*pdisability*) and DI/SSI participation among persons with disabilities (*pssad*) and estimates of the ratio of the coefficients of variation (RCV).

For SSI participation at the CAPUMA and county levels, the variance contributions of the disability component and the participation component are approximately equal. At the state level, the participation component is greater than the disability component. There is a weak positive correlation between disability prevalence and SSI participation among persons with disabilities.

For DI and combined participation, the variance contribution of the disability component is greater than the contribution of the participation component. $PV_{disability}$ is more than twice the value of PV_{ssad} at all geographic levels. There is a weak positive correlation between disability prevalence and DI (and combined) participation.

The variance decomposition results are consistent across state, CAPUMA and county levels.

These results provide an indication of the variation in DI/SSI participation that would exist if there was only variation in either the disability component or the participation component. For example, if disability prevalence were constant across states, the variation in DI participation across states would be approximately 21% of the actual variation. Correspondingly, if DI participation among persons with disabilities was constant across states, the variation in DI participation across states would be approximately 54% of the actual variation.

Table 9. Variance decomposition estimates

Geographic Level	Participation Rate	RCV	$PV_{disability}$	PV_{ssad}	Correlation *pdisability, pssad*
State	SSI	0.86	36.8%	48.0%	0.16
State	DI	1.60	55.1%	25.8%	0.27
State	Combined	1.57	54.9%	20.9%	0.28
CAPUMA	SSI	0.95	37.5%	36.7%	0.31
CAPUMA	DI	1.86	67.9%	21.2%	0.13
CAPUMA	Combined	1.74	60.5%	20.4%	0.26
County	SSI	0.94	37.0%	37.9%	0.29
County	DI	1.73	72.7%	25.6%	0.00
County	Combined	1.64	65.3%	24.0%	0.15

Note. The county estimates are based on counties with at least 10,000 persons (1844 counties).

E. Bias Analysis

In this section we assess the potential bias in variance decomposition estimates using empirical data and the information described in *Effects of Measurement Error Section* above (see Table 2). Because we used ACS survey data to estimate the number of persons with disabilities, the disability prevalence estimates are vulnerable to bias in the mean (under- or over-reporting), random error, and non-random error.

Bias in the Mean. As we describe in the *Effects of Measurement Error Section* above, bias in the mean of disability prevalence will not bias the variance decomposition estimates provided the standardize dispersion (coefficient of variation) is not biased. For example, if disability prevalence is consistently under-reported across geographic areas, the variance decomposition estimates would not be biased. Prior research suggests that disability prevalence is under-reported (Burkhauser et al., 2012); however, this research does not assess whether the under-reporting is consistent across geographic areas. The *Non-Random Error Section* below assesses the case where the bias in disability is not consistent across geographic areas.

Random Error. The main vulnerability to random error is sampling error. We estimate the magnitude of the sampling error and assess how the sampling error may bias the variance decomposition estimates. As we discussed in the *Effects of Measurement Error Section* above, random error will bias the variance decomposition estimates; however, we do not expect that it will substantially bias the relative comparison between $PV_{disability}$ and PV_{ssad}.

To assess possible bias in decomposition estimates because of sampling error, we estimate the variance of disability prevalence ($pdisability_g$) and the variance of the ACS sampling error and compare the magnitudes. We used the Census Bureau method for approximating the standard error of ACS disability prevalence estimates (Census Bureau, 2012) and we use the standard error to estimate the sampling variance for each estimate. Because the sampling variance may be different for each disability prevalence estimate, we used the mean sampling variance as an estimate of the sampling variance. Table 10 shows the estimates at the state, CAPUMA and county levels. Comparing the magnitude of the sampling variance and the magnitude of the disability prevalence variance, the sampling variance comprises less than 1% of the disability prevalence variance at the state level, approximately 4% at the CAPUMA level, and approximately 12% at the county level.

Table 10. State, state age group, and CAPUMA sampling and disability prevalence variances

	State	CAPUMA	County
Number	51	937	1844
Disability Prevalence Range	.074, .174	.032, .302	.027, .354
Population Range	353,814, 23,480,775	50,200, 6,329,545	6,983, 6,318,129
Sampling Variance Mean	.0000036	.000057	.00014
Disability Prevalence Variance	.00056	.0015	.0012
Ratio Sampling Variance Mean and Disability Prevalence Variance	.006	.038	.117

How will this level of sampling error affect the variance decomposition measures? The sampling error will result in overestimates for both $PV_{disability}$ and PV_{ssad} as described in the

Effects of Measurement Error section above. To assess the possible extent of the overestimation, we simulated the circumstances matching the county combined participation with an assumed sampling error that increased the variance of disability prevalence by 10%. This value is consistent with the sampling error at the county-level. This sampling error increased the estimate of $PV_{disability}$ by 8 percentage points and the estimate of PV_{ssad} by 11 percentage points. However, these same-direction biases do not substantially bias the relative comparison between $PV_{disability}$ and PV_{ssad}. The results of this simulation suggest that sampling error does not substantially bias the relative comparison of the variance decomposition measures. The biases at the state and CAPUMA levels would be smaller because the sampling error at these geographic levels is smaller compared to the county-level.

Sampling error will bias the correlation between *pdisability* and *pssa* toward a negative correlation (see Table 2). Thus, the positive correlation we observe (see Table 9) is not an artifact of sampling error.

Non-Random Measurement Error. The main vulnerability to non-random error is justification bias. Justification bias will occur if DI/SSI participants are more likely to report their disability compared to similar persons with disabilities that are not SI/SSI participants. We indirectly assess justification bias by examining the correlation between disability prevalence and population characteristics (e.g. mortality) that we expect are both correlated with disability prevalence and less vulnerable to DI/SSI justification bias. We use the correlation results to assess whether justification bias is a plausible explanation for the variance decomposition results. To further examine justification bias, we also examine disability prevalence and DI/SSI participation across age groups.

The variables that we expect are both correlated with disability prevalence and less vulnerable to justification bias are the following: mortality per 100 persons, proportion of persons with diabetes, the proportion of persons with fair or poor health, and the proportion of smokers.

The descriptive statistics for these variables are shown in Table 11. With the exception of mortality, the magnitude of the disability prevalence mean is comparable with the magnitudes of the other variable means. The standardized dispersion (coefficient of variation) of disability prevalence is comparable to the standardized dispersion of the other variables.

We used ordinary least squares regression to estimate the association between the ACS estimate of disability prevalence [$(pdisability^* + v)_g$] and the respective variables (equation 20).

$$(pdisability^* + v)_g = \beta prev_g + \varepsilon_g \quad (20)$$

In equation 20, *pdisability** is true disability, ν is justification bias, and *prev* is the prevalence (or rate) for the respective variable. We expect that *pdisability** and *prev* are correlated and that ν and *prev* are not correlated. Thus, the amount of variation that is explained by *prev* will depend on the relative variances of *pdisabliity** and ν. If the variance of *pdisability** is large relative to the variance of ν, we would expect a high proportion of the variation in *pdisability* would be explained by variation in *prev*. In other words, we would expect a high value for R-Squared for the regression estimate of equation 20.

The regression estimates (equation 20) are shown in Table 12. The values of R-Squared are high and this suggests that the variance of true disability prevalence (*pdisability**) is large relative to the unexplained variance. The unexplained variance may include justification bias and other factors.

Table 11. Descriptive statistics: disability, mortality, diabetes, poor or fair health, and smoking

	Statistic	States and DC (n=51)	County (n=1844)
Disability Prevalence			
	Mean	.106	.127
	Standard Deviation	.024	.045
	Coefficient of Variation	.226	.354
	75th Percentile	.115	.153
	Median	.103	.121
	25th Percentile	.088	.095
Mortality per 100 persons			
	Mean	.353	.403
	Standard Deviation	.069	.122
	Coefficient of Variation	.195	.303
	75th Percentile	.391	.480
	Median	.340	.390
	25th Percentile	.294	.315
Proportion with diabetes			
	Mean	.070	NA
	Standard Deviation	.012	NA
	Coefficient of Variation	.171	NA
	75th Percentile	.078	NA
	Median	.070	NA
	25th Percentile	.061	NA
Proportion with fair or poor health			
	Mean	.151	NA
	Standard Deviation	.031	NA
	Coefficient of Variation	.205	NA
	75th Percentile	.170	NA
	Median	.147	NA
	25th Percentile	.124	NA
Proportion Smokers			
	Mean	.239	NA
	Standard Deviation	.041	NA
	Coefficient of Variation	.171	NA
	75th Percentile	.263	NA
	Median	.236	NA
	25th Percentile	.212	NA

Table 12. Association between disability prevalence and mortality, diabetes, poor or fair health and smoking

Level	n	Slope (β)	R-Squared
Mortality per 100 persons (state level)	51	.31	.80
Mortality per 100 persons (county level)	1844	.31	.68
Proportion with diabetes (state level)	51	1.48	.58
Proportion with fair or poor health (state level)	51	.60	.61
Proportion smokers (state level)	51	.47	.65

To assess the effects of justification bias on the variance decomposition estimates, we simulate a justification bias that is perfectly correlated with *pssa* and increases the variance of *pdisability* by 10% .[12] We chose the value of 10% to represent some but not all of the unexplained variation in the results of Table 11. We do not have data to estimate the variance associated with justification bias; however, we would not expect the variance to be substantially higher given the high values of R-Squared (Table 12). The simulation indicates that this level of justification bias would increase the estimate for $PV_{disability}$ by 6 percentage points and decrease the estimate for PV_{ssa} by 3 percentage points. This biases the comparison between $PV_{disability}$ and PV_{ssad}; however, the bias is relatively small. Thus, the simulation suggests that the effects of justification bias, if existent, are relatively small and it is unlikely that the findings that $PV_{disability}$ is greater than PV_{ssad} for DI and combined participation are an artifact of justification bias.

Justification Bias and Age: The relationships between age and mortality, age and diabetes prevalence, and age and poor or fair health provide information about justification bias. If justification bias were large, we would expect the age profiles for mortality, diabetes prevalence and fair or poor health to be different from the age profile for disability prevalence. Specifically, we would expect the increase in disability prevalence with age to exceed the increase in mortality, diabetes prevalence or poor health with age because the increase in disability prevalence would reflect the justification bias associated with increased DI participation.

Figure 6 shows disability prevalence and mortality by age groups. Figure 7 shows disability prevalence, poor or fair health, and diabetes prevalence by group. Mortality, poor or fair health and diabetes all increase with age and the age profiles are comparable with the disability

[12] The mean and variance of *pdisability* and the mean and variance of *pssad* match those of the state-level estimates for DI participation, see Tables 6 and 7.

prevalence age profile. This suggests that 'true' disability prevalence also increase with age and that the increase in disability prevalence with age is not simply an artifact of justification bias. These age profiles also suggest that justification bias, if existent, is not a large component of disability prevalence estimates.

Figure 6. Disability prevalence and mortality by age group

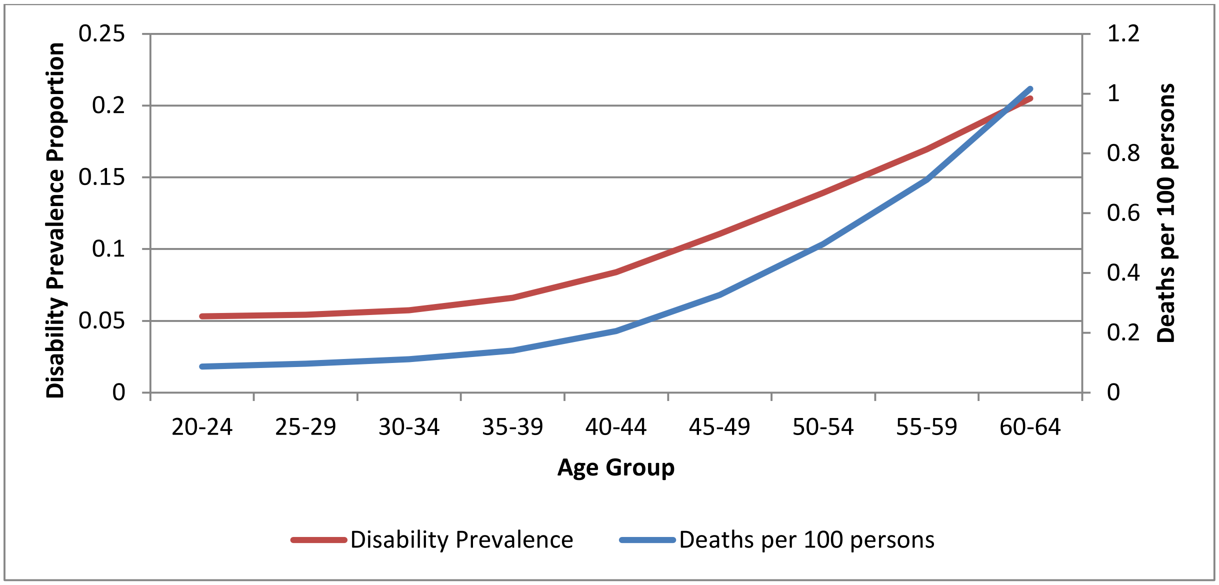
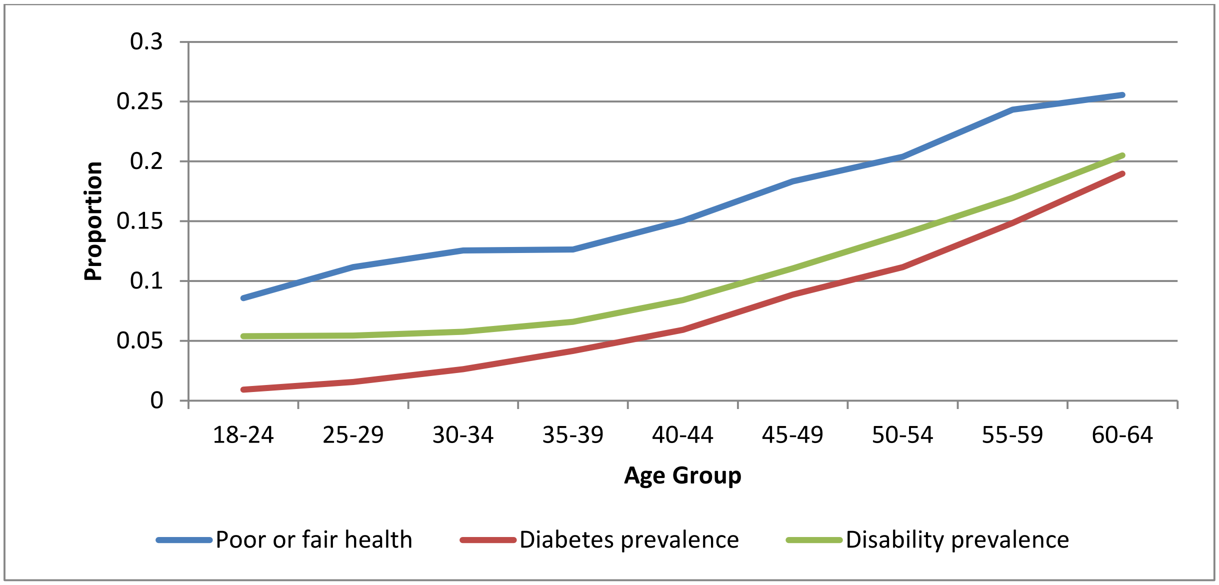


Figure 7. Disability prevalence, fair or poor health, and diabetes prevalence by age group



F. Associations, DI/SSI Participation among Persons with Disabilities and Area Characteristics

In this section we describe the results of the regression examining the associations between DI/SSI participation and CAPUMA-level characteristics (see Methods Section, Equation 18 and Table 3). Table 13 shows descriptive statistics for the characteristics. There is wide variation in the characteristics across areas. For example, the proportion Hispanic ranges from 0 to 0.97; the proportion with cognitive difficulty ranges from 0.19 to 0.57; the proportion living in poverty ranges from 0.03 to 0.40; the proportion working in manufacturing ranges from 0.02 to 0.36; and the proportion with health insurance ranges from 0.44 to 0.94.

Table 13. Descriptive statistics of CAPUMA-level characteristics

Characteristic	Mean	Std. Dev	Coefficient of Variation	Min	Max
Demographics[a]					
Age	47.05	1.37	0.03	38.45	50.87
Female	0.49	0.04	0.07	0.32	0.59
Never married	0.29	0.07	0.24	0.13	0.55
High school or less	0.59	0.09	0.15	0.31	0.83
Hispanic	0.08	0.12	1.57	0.00	0.97
Black	0.14	0.16	1.14	0.00	0.80
Non-English at home	0.10	0.12	1.24	0.00	0.94
Native born	0.95	0.07	0.07	0.50	1.00
U.S. citizen	0.97	0.04	0.04	0.69	1.00
Disability[a]					
Self-care difficulty	0.18	0.03	0.19	0.08	0.31
Hearing difficulty	0.22	0.05	0.21	0.11	0.50
Vision difficulty	0.17	0.05	0.27	0.06	0.60
Independent living difficulty	0.34	0.05	0.13	0.12	0.47
Ambulatory difficulty	0.52	0.06	0.12	0.33	0.76
Cognitive difficulty	0.41	0.05	0.11	0.19	0.57
Income and Poverty[b]					
Below 100% federal poverty	0.15	0.05	0.37	0.03	0.40
Annual income (thousands)	34.84	7.80	0.22	18.95	75.47
Annual household income (thousands)	71.14	16.33	0.23	39.92	145.05
Annual earned income (thousands)	38.31	8.31	0.22	25.97	87.35
Labor Market[b]					
Male labor force participation	0.86	0.04	0.05	0.68	0.94
Female labor force participation	0.75	0.05	0.06	0.55	0.86
Self-employment	0.07	0.02	0.30	0.03	0.18
Usual hours worked per week	38.95	1.32	0.03	33.68	45.40
Worked 26 weeks or less among workers	0.12	0.03	0.22	0.06	0.30
Service occupations	0.16	0.03	0.17	0.09	0.30
Production occupations	0.14	0.05	0.35	0.03	0.31
Sales occupations	0.24	0.02	0.10	0.15	0.33
Construction and maintenance occupations	0.11	0.03	0.29	0.02	0.25
Management occupations (omitted category)	0.34	0.06	0.18	0.21	0.68
Manufacturing industry	0.12	0.06	0.48	0.02	0.36

TABLE 13 (*continued*)

Characteristic	Mean	Std. Dev	Coefficient of Variation	Min	Max
Education/health services industry	0.21	0.03	0.16	0.12	0.41
Prof and business services industry	0.16	0.05	0.33	0.06	0.42
Other industries	0.20	0.05	0.23	0.11	0.39
Wholesale and retail trade industry	0.15	0.02	0.12	0.06	0.28
Leisure and hospitality industry (omitted category)	0.09	0.03	0.28	0.05	0.28
Public Assistance and Health Insurance[b]					
Public Assistance	0.01	0.01	0.51	0.00	0.04
Health insurance	0.79	0.07	0.09	0.44	0.94

Note. Other industries includes agricultural and related industries; mining, quarrying, and oil and gas extraction; construction; transportation and utilities; information; financial activities; other services and government workers.

[a] Categories of variables in which the population of persons with disabilities is the denominator.

[b] Categories of variables in which the population of persons without disabilities is the denominator.

The estimates of the ordinary least squares regression of DI/SSI participation versus the characteristics are shown in Table 14. The results suggest that variation in area-level characteristics account for much of the variation in DI/SSI participation among persons with disabilities; approximately 63% of the variation in SSI participation and approximately 50% of the variation in DI.

SSI participation is associated with a variety of area-level demographic characteristics of persons with disabilities. Higher SSI participation is associated with areas of higher average ages, lower education levels, higher proportions of never married persons, higher proportions of Blacks, higher levels of U.S. citizens, and lower levels of native born persons. The seemingly contradictory associations between SSI participation and citizenship and between SSI participation and nativity may exist because of the properties of regression; the association between SSI participation and nativity is determined holding all other variables constant including citizenship. Area-level citizenship and nativity are highly correlated (correlation coefficient, 0.95).

Table 14. Regression estimates, association of DI/SSI participation and area characteristics

Area Characteristic	SSI Estimate	SSI P Value	DI Estimate	DI P Value
Intercept	-0.604	0.0056	-0.163	0.5246
Demographics[a]				
Average age	0.007	0.0001	0.018	<.0001
Female	-0.002	0.9624	-0.024	0.6951
Never married	0.173	0.0001	0.032	0.5523
High school or less	0.175	<.0001	0.100	0.0052
Hispanic	0.046	0.287	0.036	0.4858
Black	0.080	<.0001	0.018	0.4277
Non-English at home	-0.041	0.4484	-0.096	0.1305
Native born	-0.432	<.0001	-0.230	0.0453
U.S. citizen	0.387	0.0273	0.310	0.1328
Disability[a]				
Self-care difficulty	-0.122	0.0586	-0.011	0.8851
Hearing difficulty	-0.064	0.2294	-0.192	0.0022
Vision difficulty	-0.087	0.0527	-0.269	<.0001
Independent living difficulty	0.108	0.0377	0.161	0.0084
Ambulatory difficulty	0.060	0.1991	-0.032	0.564
Cognitive difficulty	0.114	0.0106	-0.011	0.831
Income and Poverty[b]				
Below 100% federal poverty	0.371	<.0001	-0.134	0.1288
Average annual income	0.002	0.5378	0.007	0.0146
Average annual household income	-0.004	<.0001	-0.002	0.0025
Average annual earned income	0.008	<.0001	-0.003	0.2619
Labor Market[b]				
Male labor force participation	-0.264	0.0034	-0.588	<.0001
Female labor force participation	-0.049	0.4974	0.000	0.9963
Self-employment	0.176	0.0986	-0.211	0.0928
Usual hours worked per week	-0.001	0.5447	-0.005	0.0648
26 or fewer weeks worked	-0.383	0.0002	-0.554	<.0001
Service occupations	0.341	0.005	0.303	0.0343
Production occupations	0.193	0.0794	-0.054	0.6749
Sales occupations	0.150	0.2087	0.076	0.5891
Construction and maintenance occupations	0.280	0.017	0.484	0.0005
Management occupations (omitted category)	N/A	N/A	N/A	N/A
Manufacturing industry	0.142	0.1803	0.437	0.0005
Education/health services industry	0.556	<.0001	0.476	0.0001
Prof and business services	0.152	0.2178	0.201	0.165
Other industries	0.381	<.0001	-0.114	0.2515
Wholesale and retail trade industry	0.387	0.0077	0.628	0.0002
Leisure and hospitality (omitted category)	N/A	N/A	N/A	N/A
Public Assistance and Health Insurance[b]				
Public Assistance	2.780	<.0001	-0.436	0.2301
Health insurance	0.122	0.0076	0.172	0.0014

[a] Categories of variables in which the population of persons with disabilities is the denominator.

[b] Categories of variables in which the population of persons without disabilities is the denominator.

DI participation is also associated with a variety of demographic characteristics. Similar to SSI, higher DI participation is associated with areas of higher average ages, lower education levels, higher levels of U.S. citizens (not statistically significant) and lower levels of native born persons. In contrast to the SSI findings, we were unable to detect an association between DI participation and the proportion Black or the proportion never married.

SSI and DI participation are associated with the variation in disability types across CAPUMAs. Higher SSI and DI participation is associated with areas of higher prevalence of independent living difficulties and cognitive difficulties. Higher DI participation is also associated with areas of lower prevalence of vision difficulties and hearing difficulties.

Among persons with disabilities, SSI is more strongly associated with income and poverty characteristics compared to DI. A 10 percentage point increase in the poverty rate among persons without disabilities is associated with a 2.7 percentage point increase in SSI participation. We are unable to detect an association between DI participation and poverty. Higher levels of SSI and DI participation are associated with lower levels of average area income among persons without disabilities. A decrease in average household income, among households without an adult with a disability, of $10,000 is associated with increase of 4.1 percentage points in SSI participation and 2.3 percentage points in DI participation.

Among persons with disabilities, SSI and DI participation are associated with a variety of area labor market characteristics. Higher levels of both SSI and DI participation are associated with lower levels of male labor force participation; however, the association is stronger for DI. A 10 percentage point decrease in male labor force participation is associated with a 5.8 percentage point increase in DI participation and 2.6 percentage point increase in SSI participation. Relative to management occupations, higher proportions of workers without disabilities working service in construction and maintenance occupations are associated with higher levels of SSI and DI participation. Relative to the proportion of persons employed in the leisure and hospitality industry, employment in the education and health care services industry and other industries is associated with higher levels of SSI participation. The industry associations are different for DI. Relative to the proportion of persons employed in the leisure and hospitality industry, employment in manufacturing, education and healthcare services, and wholesale and retail trade are associated with higher levels of DI participation.

There is a strong association between area SSI participation and public assistance participation among persons without disabilities. A 1 percentage point increase in area public assistance participation is associated with a 2.8 percentage point increase in SSI participation among persons with disabilities. We are unable to detect an association between DI participation and public assistance participation. Areas of higher health insurance rates among persons without disabilities are associated with higher levels of participation in both SSI and DI.

Predicted Versus Actual, DI/SSI Participation among Persons with Disabilities: The variation in area characteristics accounts for much but not all of the variation in DI/SSI participation among persons with disabilities across areas. To assess whether there are regional patterns to the unexplained variation, we created thematic maps of CAPUMA differences between the predicted program participation and the actual program participation (See Figures 8 and 9). We refer to a positive difference between the actual and the predicted as an under-prediction.

Figure 8. Difference between actual and predicted: proportion of working-age persons with disabilities participating in SSI, CAPUMA level

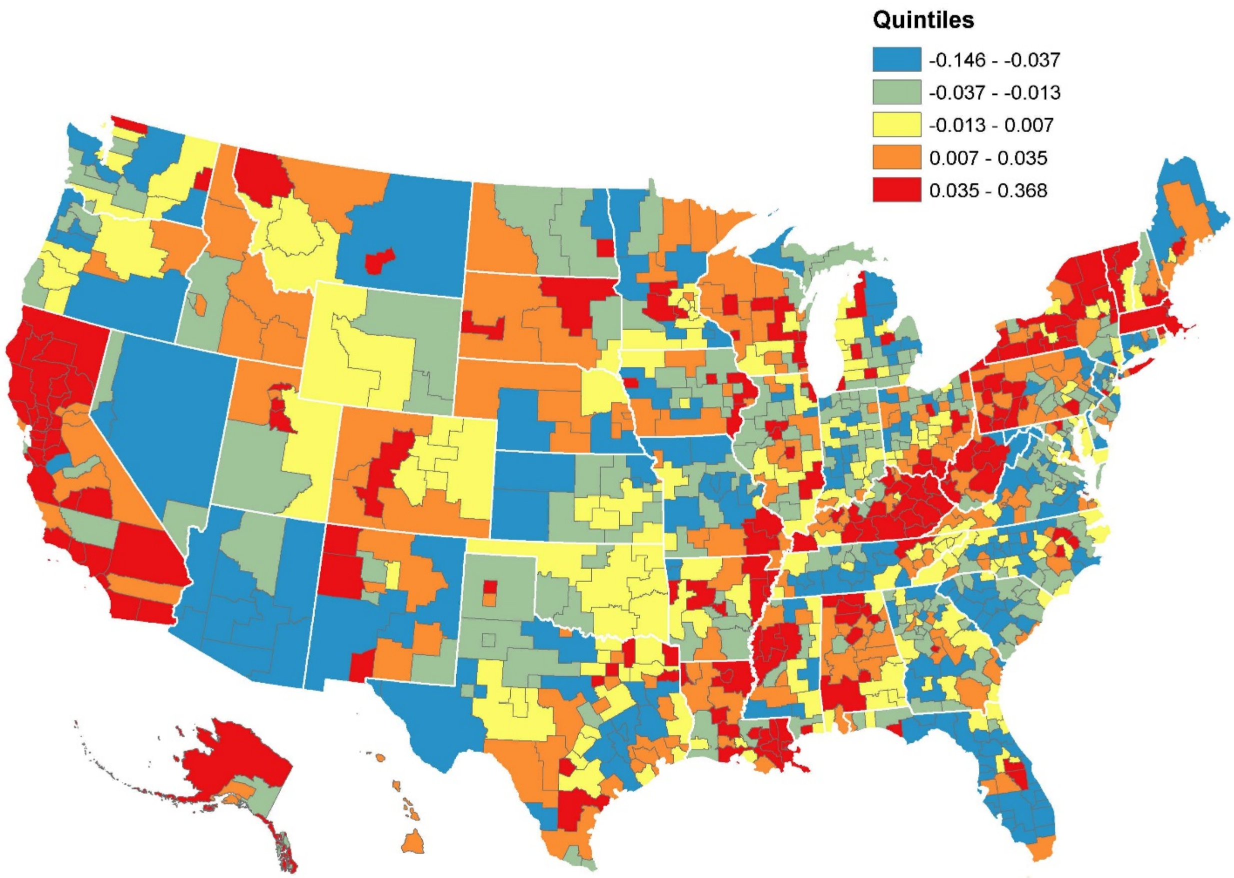
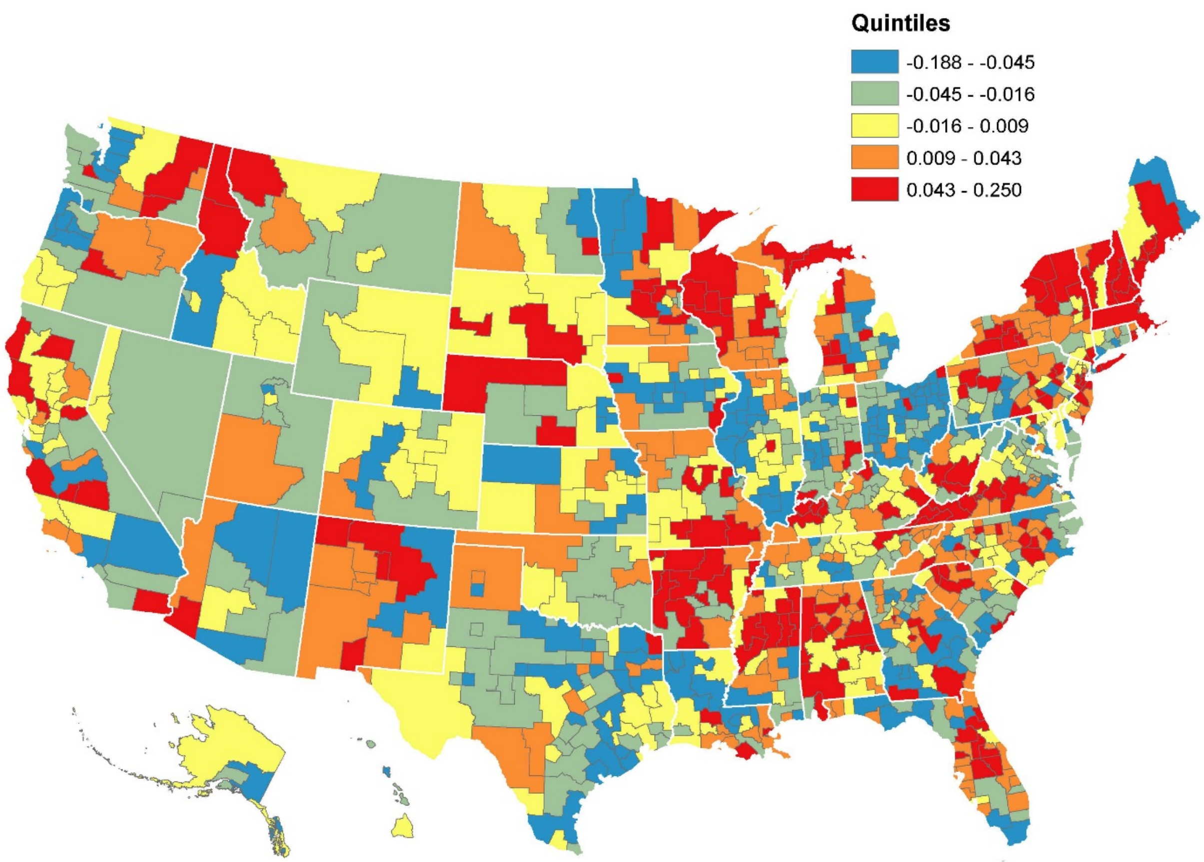
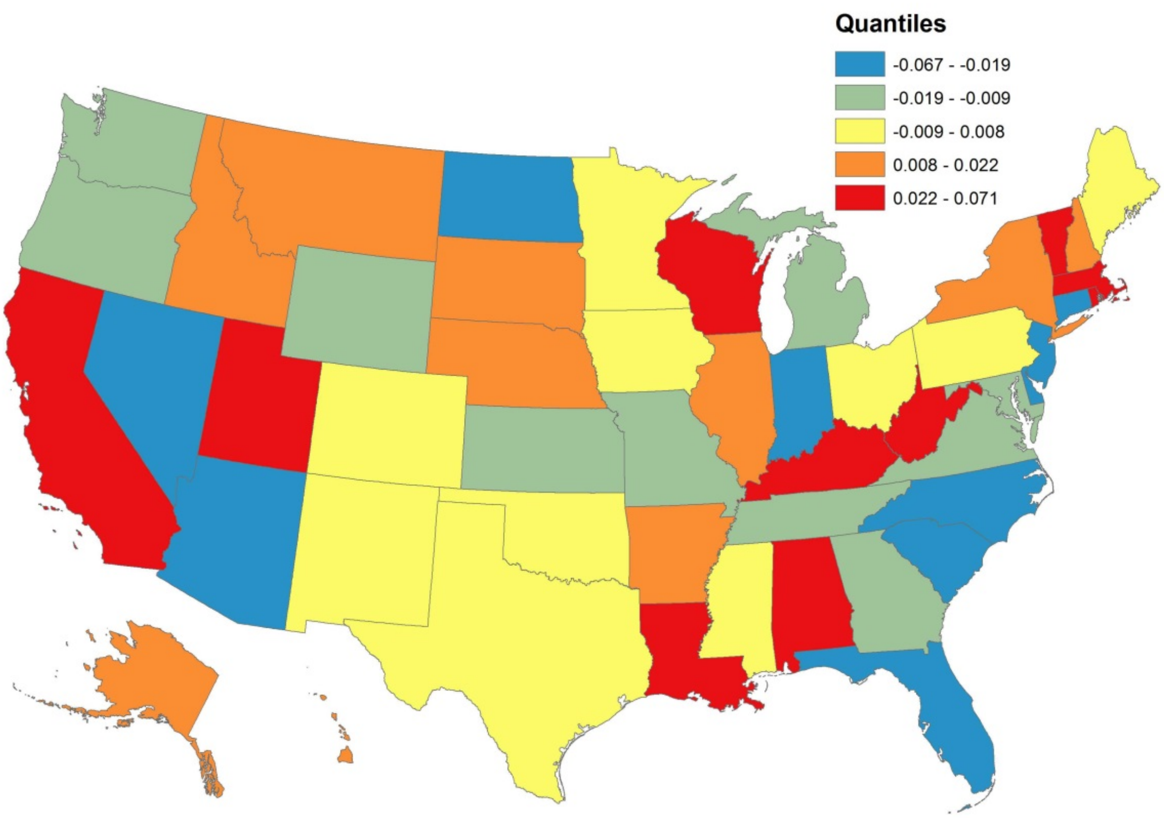


Figure 9. Difference between actual and predicted: proportion of working-age persons with disabilities participating in DI, CAPUMA level



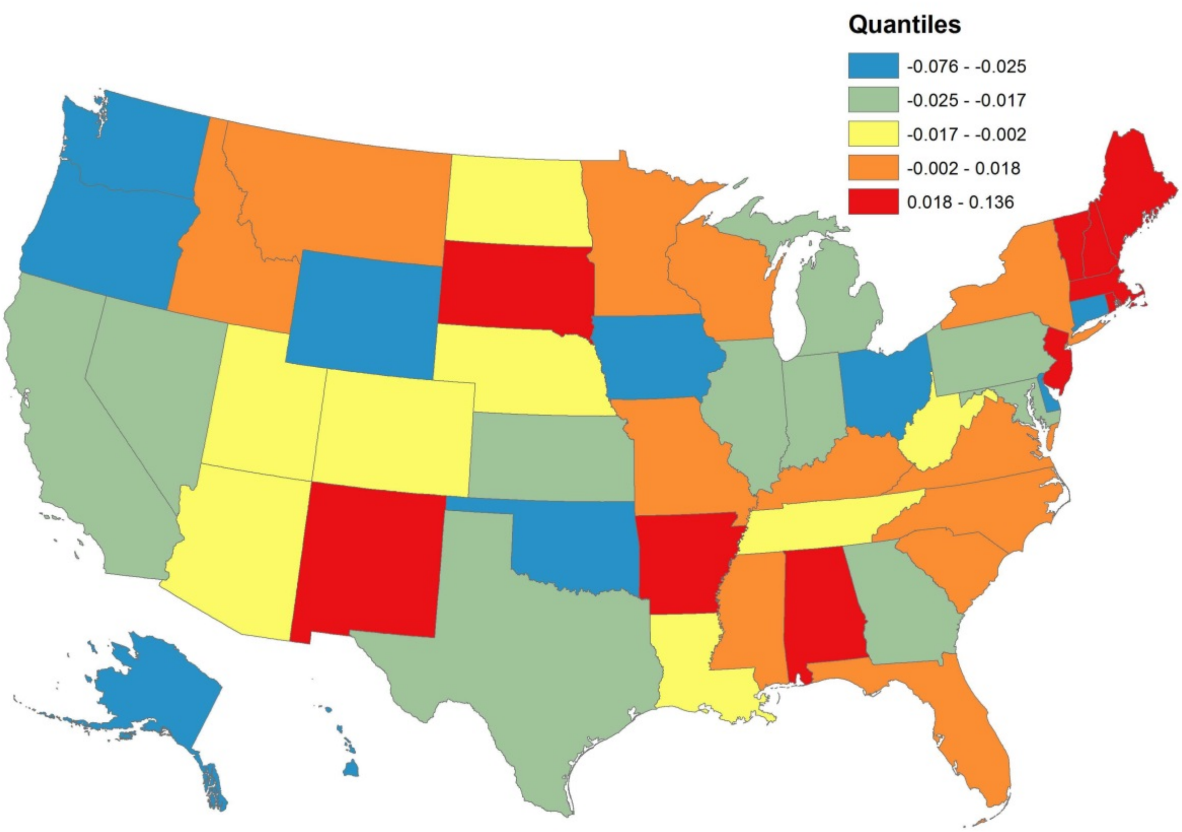
For SSI, there are both under-predicted and over-predicted CAPUMAs throughout the U.S. without strong regional patterns with the exception of a region of under-predicted SSI participation in the area containing Kentucky, West Virginia, southeastern Ohio, western Pennsylvania, western New York and western and northern Vermont (See Figure 8). A portion of this region (eastern Kentucky, West Virginia, southeastern Ohio and southwestern Pennsylvania) has a history of coal industry volatility and this history may account for at least part of regional difference between actual and predicted (Black, Kermit, and Sanders, 2002). Many states contain both under-and over-predicted CAPUMAs, for example, New Mexico, Texas, Arkansas, Missouri, Georgia, North Carolina, Virginia, Pennsylvania, Maine, Ohio, Illinois, Michigan, , Iowa, Utah, Oregon and Washington. There are some states with predominantly under-predicted CAPUMAs or predominantly over-predicted CAPUMAs. Nevada, Arizona, Kansas, and South Carolina predominantly have areas of lower than predicted SSI participation. Vermont, Kentucky, and California predominantly have areas of higher than predicted SSI participation. The CAPUMA actual and predicted participation rates were aggregated on a CAPUMA population basis to determine the state-level differences between actual and predicted SSI participation (See Figure 10).

Figure10. Difference between actual and predicted: proportion of working-age persons with disabilities participating in SSI, state level



For DI, no regional patterns are obvious. Under- and over-predicted CAPUMAs exist in all regions. Many states contain both under- and over-predicted CAPUMAs, for example Florida, Georgia, South Carolina, North Carolina, Tennessee, Virginia, Pennsylvania, Michigan, Minnesota, Iowa, Missouri, Kansas, Nebraska, Idaho, Montana, Washington, Oregon, and California. There are some states with predominantly under- or over-predicted CAPUMAs. New Hampshire, Wisconsin, Arkansas, and Alabama, predominantly have CAPUMAs with higher than predicted DI participation. Ohio's CAPUMAs are predominantly lower than predicted. The state-level differences between actual and predicted DI participation are shown in Figure 11.

Figure 11. Difference between actual and predicted: proportion of working-age persons with disabilities participating in DI, state level



State Fixed Effects Model: The estimates of the previous section assume there is a single intercept in the model of the association between DI/SSI participation and CAPUMA-level characteristics (see Methods Section, Equation 18). Under this assumption, the area-level factors accounted for 50% of the variation in DI participation among persons with disabilities and 63% of the variation in SSI. The fixed effects model relaxes the single-intercept constraint and includes state-specific intercepts. The state-specific intercepts account for between-state variation that is not accounted for by the variation in local area characteristics. Estimates of the fixed effects model indicate that together, variation in area-level factors and variation between-states account for 66% (increase of 16 percentage points) of variation in DI participation and 80% (increase of 17 percentage points) of the variation in SSI participation. As described in the Methods Section, the increase may be attributed to unobserved between-state variation in state-determined policy (e.g. health insurance regulation or DI/SSI disability determination services) or to between-state variation in unobserved factors that are not determined by state policy (e.g. stigma, discrimination, attitudes about employment). We are not able to determine how much of the increase is attributed to state policy and how much is attributed to other factors. As an approximation, the increase (approximately 17 percentage points) represents the upper-limit of the variance contribution of unobserved state policy.[13]

G. Principal Component Analysis, DI/SSI Participation among Persons with Disabilities

The regression analysis provides information on the association between DI/SSI participation and area characteristics; however, it does not provide an indication of the variance contributions of separate area characteristic. To do this, we use principal components analysis

[13] The upper limit applies to unobserved state policy. Unobserved state policies are those policies that are not reflected in the variation of local area characteristics. For example variation poverty, public assistance participation and health insurance rates may in part be due to variation in state policy.

(See *Methods Section*). Principal component analysis resulted in a reduction from 35 independent variables to 12 components (see Table 14). The 12 components account for 84.1% of the total variation in the original variables. Variables that have strong or moderate correlations (0.5 or greater) with the respective components are shown in Table 15. The component names were chosen to summarize the general meaning of the component based on the variables correlated with the component. For consistency we group the components into the same categories we used for the full-variable regression analysis. This grouping is approximate because some components have variables in more than one category.

Table 15. Principal components

Component Name	Variables with Positive Correlation (correlation coefficient)	Variables with Negative Correlation
Demographics		
Hispanic/Non-English	Hispanic (.93), Non-English at home (.95)	Native born (-.90), U.S. citizen (-.93)
Black/Low Self Employment	Black (.89), Never married (.53)	Self-employment (-.60), Hearing difficulty (-.56)
Age/Fewer Cognitive Disabilities	Average age (.77), Ambulatory difficulty (.54)	Cognitive difficulty (-.70)
Female	Female (.93)	
Disability		
Personal Assistance Needs	Self-care difficulty (.85), Independent living difficulty (.83)	
Income		
Income	Average annual income (.93), Average annual household income (.93), Average annual earned income (.95), Professional and business service (.77), Health insurance (.50)	Poverty Level (-.65)
Labor Market		
Manufacturing	Production occupations (.77), manufacturing industry (.83), High School or less (.53)	Service occupations (-.60)
Labor force participation	Male labor force participation (.68), Female labor force participation (.83)	26 or fewer weeks worked (-.62)
Construction/Work Hours	Hours worked per week (.74), construction and maintenance occupations (.67), other industries (.80)	
Sales	Sales occupations (.84), Wholesale and retail trade industries (.86)	
Education/Health Services	Education and health care services industry (.77)	
Public Assistance		
Public Assistance	Public Assistance (.82)	

The results of an OLS regression of DI/SSI participation versus the principal components are shown in Table 16 (See equation 19). The components account for approximately 53% of area variation in SSI participation among persons with disabilities and approximately 43% of variation in DI participation. This is less than the variance accounted for by the regression based on the original variables because the components account for most (84%) but not all of the total variance of the variables. Because the components are uncorrelated, we are able to calculate the percent of the total variance in DI/SSI participation that is associated with each component.

For SSI, the largest contributions to the area variance are the Black/low self-employment component (18.3%), the public assistance component (10.3%), and the income component (8.1%). The Black/low self-employment component and the public assistance component are associated with higher levels of SSI participation and the income component is associated with lower levels of SSI participation. Taken together, these components are likely indicative of the economic conditions in an area. Area with high levels of the Black/low self-employment and public assistance components and low levels of the income component are likely economically disadvantaged. Smaller but significant contributions to the area variation in SSI participation are associated with the following components: personal assistance needs (5.7%), education/health services (4.2%), labor force participation (3.7%), Hispanic/non-English (1.6%) and the construction/work-hours component (0.9%).

For DI, the largest contributions to the variance are the Hispanic/non-English component (10.4%), personal assistance needs component (7.4%), the education/health services component (7.0%), and the age/fewer cognitive difficulties component (6.0%). The Hispanic/non-English component is associated with lower levels of DI participation and the personal assistance needs, education/health services, and age/fewer cognitive difficulties components are associates with higher levels of DI participation. Smaller but significant contributions to variation in DI participation are associated with the following components: manufacturing (3.6%), construction/work-hours (3.1%), public assistance (2.4%), income (1.0%), labor force participation (0.6%), Black/low self-employment (0.5%), and sales (0.5%).

Table 16. Principal components regression results

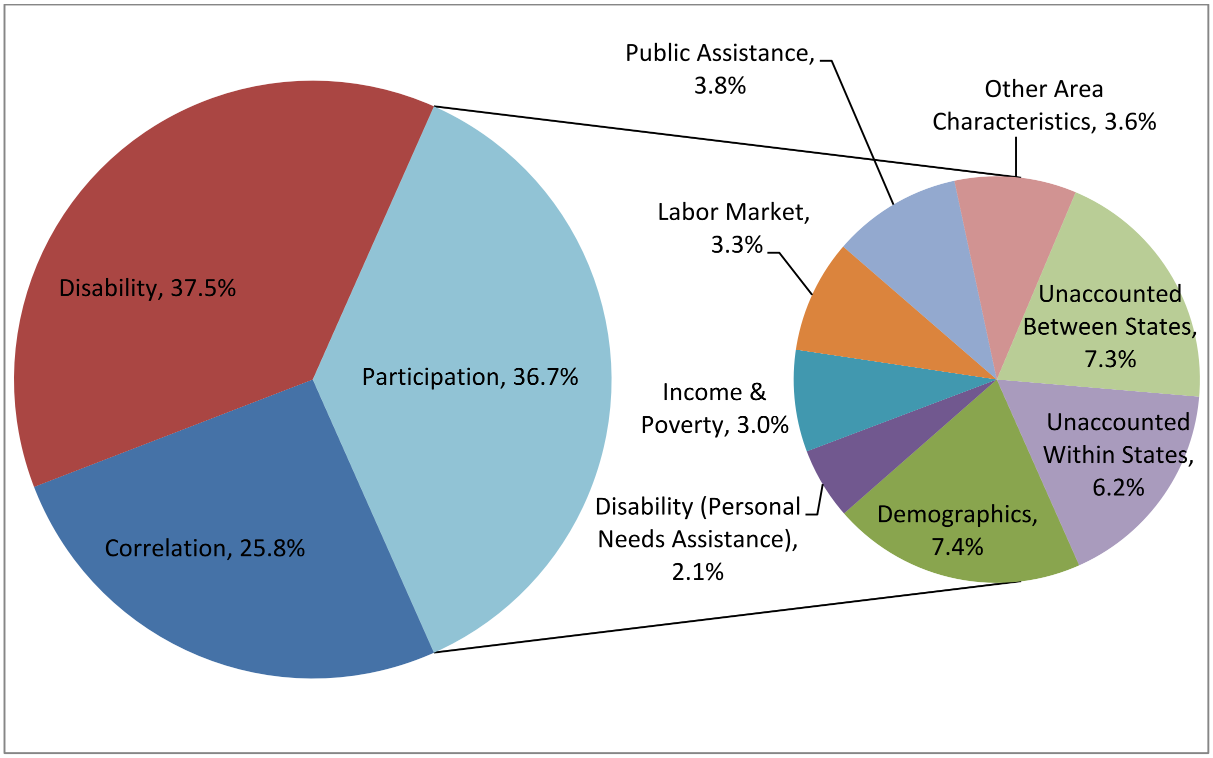
	SSI			DI		
Component	Estimate	P Value	% Variance	Estimate	P Value	% Variance
Intercept	0.227	<.0001		0.451	<.0001	
Demographics						
Hispanic	0.010	<.0001	1.60%	-0.025	<.0001	10.40%
Black	0.033	<.0001	18.30%	0.006	0.0031	0.50%
Age	-0.002	0.2734	0.10%	0.02	<.0001	6.10%
Female	0.004	0.0328	0.20%	0.002	0.3484	0.10%
Disability						
Personal assistance needs	0.019	<.0001	5.70%	0.022	<.0001	7.40%
Income						
Income	-0.022	<.0001	8.10%	-0.008	<.0001	1.00%
Labor Market						
Manufacturing	-0.003	0.1428	0.10%	0.015	<.0001	3.60%
Labor force participation	-0.015	<.0001	3.70%	0.006	0.0015	0.60%
Construction	0.007	<.0001	0.90%	-0.014	<.0001	3.10%
Sales occupations	-0.003	0.1228	0.10%	0.006	0.0044	0.50%
Education/Health Services	0.016	<.0001	4.20%	0.021	<.0001	7.00%
Public Assistance						
Public Assistance	0.025	<.0001	10.30%	0.012	<.0001	2.40%

H. Decomposition Summary, CAPUMA Level

In this section, we combine the CAPUMA-level disability/participation variance decomposition and the principal components variance decomposition to obtain an overall decomposition in chart form (see Figures 12 and 13). In order to summarize the principal components variance decomposition, we sum the component variances into categories.

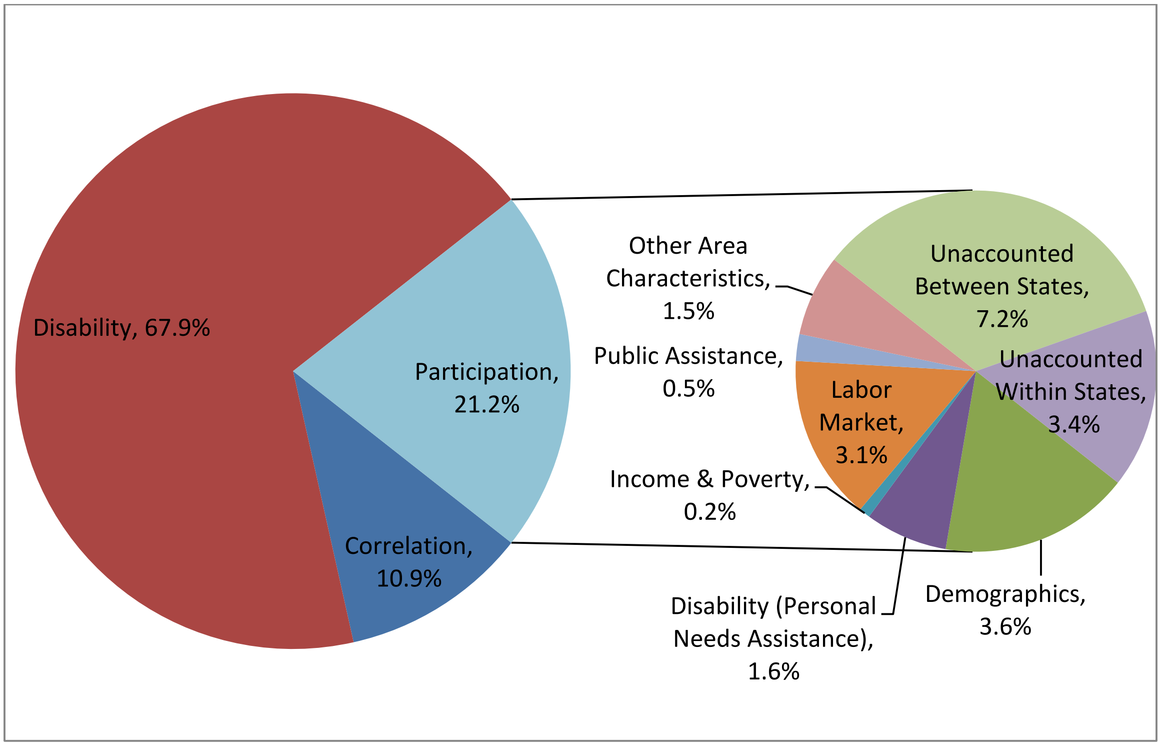
For SSI, the variation in disability prevalence and SSI participation among persons with disabilities contribute approximately equally to the total variance, 37.5% and 36.7%. Because disability prevalence and SSI participation among persons with disabilities are weakly correlated, the correlation contributes approximately 25.8% to the total variance. The 36.7 percentage point participation contribution is further decomposed; approximately 23.2 percentage points are attributed to variation in observed area characteristics and 13.5 percentage points are unaccounted.

Figure 12. Variance decomposition summary, SSI participation among working-age persons



For DI, the variation in disability prevalence accounts for more of the variation (approximately 67.9%) in total participation among working-age persons compared to variation in DI participation among persons with disabilities (approximately 21.2%). Correlation accounts for approximately 21.2% of the total variance. The 21.2 percentage point participation contribution is further decomposed; approximately 10.6 percentage points are attributed to variation in observed area characteristics and 10.6 percentage points are unaccounted.

Figure 13. Variance decomposition summary, DI participation among working-age persons



IV. DISCUSSION

There is wide geographic variation, at the state level and the sub-state level, in DI/SSI participation among working-age persons. Mathematically the variation is the product of two components, the geographic variation in disability prevalence and the geographic variation in DI/SSI participation among persons with disabilities. Both components contribute substantially to the variation in DI/SSI participation among working-age persons. We further decomposed variation in DI/SSI participation among persons with disabilities into the variation in area socioeconomic characteristics. Area characteristics vary widely and area characteristics account for much of the variation. In total, nearly all (86.4% for SSI, 89.4% for DI) of the variation in DI/SSI participation can be accounted for by the following: (a) variation in disability prevalence, (b) correlation between disability prevalence and DI/SSI participation among persons with disabilities, and (c) variation in area socioeconomic characteristics.

Variation in disability prevalence accounts for a large portion of the variation in DI/SSI participation, particularly for DI and combined participation. For example, if there was no variation in disability prevalence across states, the variation in DI participation would be reduced by approximately 80%. The reasons for the wide geographic variation in disability prevalence across states are not known. Future research is needed. The correlations between disability prevalence and DI/SSI participation among persons with disabilities are weak suggesting that the factors associated with the variation in disability prevalence will be different than the factors associated with DI/SSI participation. The variation in disability prevalence may exist because of variation in the incidence of disability and/or variation in net-migration of persons with disability relative to persons without disabilities across geographic areas. The incidence of disability could vary across areas because of geographic variation in demographics, health care access or quality, health or disability risk behaviors, risk of injury, or disease prevalence. Net-migration could vary because of labor market, cost-of-living or other area differences. For instance, if those without disabilities are more likely than those without disabilities to migrate from economic weak areas to areas that are strong, that would increase prevalence in economically weak areas relative to prevalence in strong areas. There are strong regional patterns in disability prevalence. Analysis of these patterns may help to identify the reasons for the variation.

There are differences in the decomposition of SSI participation among persons with disabilities compared to DI. SSI is a means-tested program and for the most part, only individuals living in very low income households are eligible. Thus, we expect that area variation in SSI participation would be associated with variation in area economic conditions with higher participation in economically disadvantaged areas. Our findings are consistent with this. We find the largest contributions to the variance in SSI participation among persons with disabilities are from the following components: Blacks/low self-employment component, the income component, and the public assistance component. Area labor market and disability characteristics also account for variation in SSI participation among persons with disabilities; however, to a much smaller extent.

In contrast with SSI, little of the variance in DI participation is associated with characteristics indicative with economic disadvantage. The largest DI variance contribution is attributed to the Hispanic/non-English component. Areas with higher proportions of Hispanics, people who speak something other than English in the home, non-native born persons, or non U.S. citizens have lower DI participation. This could be because of access limitations, for example, an absence of qualifying work quarters, language barriers or discrimination. It is also possible that the demand for DI varies by ethnicity or country of origin. Further research is needed to determine the explanations. The DI variance is also attributed to components related to area industries and occupations as indicated by the following components: education/health care services component, construction/work-hours component and manufacturing component. Areas with higher proportions of workers in education and health care services, production and manufacturing, and construction have higher levels of DI participation. Further research is needed to explain this.

There is a strong regional pattern in DI/SSI participation among working-age persons with higher program participation in southeastern states. This is due to the regional pattern of disability prevalence and, to a lesser extent, the regional patterns of DI/SSI participation among persons with disabilities. Controlling for observed area socioeconomic characteristics, the regional patterns of DI/SSI participation among persons with disabilities are either eliminated as is the case for DI (See Figure 9) or greatly reduced as is the case for SSI (See Figure 8). Thus, nearly all of the regional patterns in DI/SSI participation appear to be the result of regional patterns in disability prevalence and regional patterns in observed area characteristics. This suggests that, if existent, the contributions of other unobserved regional factors are relatively small compared to the overall regional variation in DI/SSI participation. For example, this suggests that variance contribution of regional variation in DI/SSI administration (e.g. disability determination services) is small compared to the total variation.

The decomposition also provides insight into how much unobserved state policy may contribute to geographic variation in DI/SSI participation among persons with disabilities. Some state policies may be reflected in the variance decomposition estimates, for example welfare policy may be reflected in the variance contribution of the public assistance component. We expect that other state policies are not captured by the factors included in the decomposition, for example insurance regulation and disability determination services. We are not able to estimate the variance contribution of these unobserved factors. Our findings suggest these unobserved state policies may contribute; however, the contributions appear to be relatively small. We base this on the fixed effects estimates and the within-state area differences between predicted and actual DI/SSI participation among persons with disabilities (See Figures 8 and 9). The fixed effects results suggest that the upper-limit of the contribution of unobserved state policies is approximately 17%. Also, most states have mix of within-state areas that have both higher and lower participation than predicted. If unobserved state policies substantially contributed to geographic variation, we would expect the within-state predictions to be more consistent, either consistently under-predicted or consistently over-predicted. There are states that match this pattern; however, most states include a mix of under- and over-predicted areas.

The decomposition accounts for approximately 90% of the variation in DI/SSI participation (See Figures 12 and 13). Why is there unexplained variation? There are a number of possible explanations. Our methods rely on cross-sectional data that provide a current snapshot of area characteristics. However, DI/SSI participation depends on both current and past characteristics, for example long-term labor market characteristics. We are unable to account for characteristics in prior time periods that are uncorrelated with current characteristics. We are also not able to account for migration. People’s DI/SSI participation may be affected by characteristics of their prior residence area and we are not able to account for this. Also, there may be area variation in characteristics that affect DI/SSI participation that we were unable to observe (e.g. employment discrimination, population density). Lastly, part of the unexplained variation is likely due to measurement error. Disability prevalence and area characteristics are estimated with survey data and these are estimated with error. This error may contribute to the unexplained variation.

Publicly available ACS and SSA data representing sub-state areas made this analysis possible. Additional access to sub-state level data would improve this analysis and facilitate future disability research. In this analysis, we merged ACS PUMA-level statistics with county-level SSA administrative data to generate CAPUMA-level data. A shortcoming of this approach is that some of the CAPUMAs represent large populations, generally because some counties have large populations. This merging of PUMAs and counties obscures some of the local-area variation, particularly in urban areas with dense populations. The analysis would be improved if it were conducted at the PUMA-level. This would more than double the number of observations, reveal urban-area variations, and provide consistent population sizes across areas. Currently, the PUMA-level approach is not possible because DI/SSI participant counts are not available at the PUMA-level. In addition, this analysis would be improved if it were conducted on subgroups that vary in DI/SSI participation rates, for example subgroups based on age and sex. This subgroup analysis was not possible because substate DI/SSI participation counts by age and sex are not publicly available. Publicly available, PUMA-level DI/SSI participant counts, by age and sex, would further this research and facilitate additional sub-state disability research.

This study is a decomposition of geographic variation in DI/SSI participation; however, the findings have implications beyond accounting for area variation. We discuss three. First, the decomposition suggests that changes in disability prevalence, if they occur over time, will be reflected in DI participation changes. Disability prevalence is the predominant source of the variation in DI participation across areas even though there is also wide variation in labor markets and economic conditions. It is likely that similar association would exist between changes in disability prevalence over time and changes in DI participation. This suggests that future changes in disability prevalence will proportionally change DI participation. Little is known about the long-term trends in disability prevalence and further research is needed. One potential research approach is the analysis of area variation in disability prevalence. There is wide variation in disability prevalence across areas and explanations for this variation may provide insight into the long-term trends in disability.

Second, the decomposition suggests that demographics and labor market characteristics affect DI participation. Prior research has shown the importance of the time-based changes in the age/sex composition of the labor force as explanations for changes in DI participation (Daly, 2013; Liebman, 2015). In addition to these characteristics, the decomposition suggests that time-based changes in the proportions of Hispanics, non-citizens, persons born outside of the U.S, and persons speaking a language other than English at home could also affect DI participation. Further research is needed to evaluate the time trends of these characteristics. Prior research indicates that changes in industrial composition of the labor market affect DI participation (Autor and Duggan, 2003). The decomposition results also suggest this.

Lastly, the geographic analysis illustrates the wide between-state and within-state variation in socioeconomic conditions experienced by persons with disabilities. In some urban areas there is substantial disparity in conditions between geographic areas that are in very close proximity. Because of this heterogeneity, the effects of DI/SSI reforms will likely vary across locations. The design of DI/SSI reforms, pilot programs, and evaluations will be strengthened by taking this heterogeneity into account.

V. LIMITATIONS

There are four limitations to this analysis. The first is possible bias in the variance decomposition measures because of measurement error. Persons with disabilities may be more likely to report their disability when participating in DI or SSI and this could bias the estimates. Our analysis suggests that justification bias, if existent, is not of a magnitude that would change our overall findings.

The variance decomposition of DI/SSI participation into the disability component and the participation component is based on a mathematical relationship (DI/SSI participation = disability prevalence x DI/SSI participation among persons with disabilities). Thus, no causal inference is necessary. However, this is not the case for the regression-based decomposition of DI/SSI participation among persons with disabilities. In this case, the decomposition is a descriptive association rather than a causal association. For example, the variation in SSI participation is associated with the variation in the prevalence of independent living difficulties; however, it is possible that the association is caused by something that is correlated with the variation in independent living difficulties (e.g. access to personal care services) rather than the direct variation. Thus, the second limitation is the descriptive, rather than causal, associations for the decomposition of DI/SSI participation among persons with disabilities.

Third, the DI participation rates used in this study include disabled workers but do not include disabled widows or disabled adult children because the data was not available. In 2011, there were approximate 8.5 million disabled workers and one million disabled widows and disabled adult children (SSA, 2015).It is possible that the variance decomposition would change with the inclusion of disabled spouses and disabled adult children. Thus, the findings of this study only apply to DI disabled workers and not to disabled spouses or disabled adult children.

This limitation would be alleviated if PUMA-level disabled widow and disabled adult children counts were publicly available.

Finally, the SSI participation rates used in this study include federal SSI and federally administered state supplementation participants. The inclusion of federally administered state supplementation participations will cause variation SSI participation across states because federal state supplementation does not exist in all states and there is variation in eligibility criteria across states. The variance decomposition did not account for this variation. Because participation in SSI is much higher relative to participation in state supplementation-only, 6.5 million vs 167 thousand (SSA, 2012), we do not expect this to substantially affect our findings.

VI. CONCLUSIONS

There is wide geographic variation in DI/SSI participation. Approximately 90% of the geographic variation can be accounted for by geographic variation in disability prevalence, area socioeconomic characteristics, and correlation between disability prevalence and DI/SSI participation among persons with disabilities.

There are differences in the accounting for DI and SSI. More of the variation in DI participation is accounted for by variation in disability prevalence and less by socioeconomic characteristics compared to SSI. Compared to DI, more variation in SSI participation is accounted for by characteristics associated with economically disadvantaged areas. The explanations for the geographic variation in disability prevalence are not known and further research is needed. The correlations between disability prevalence and DI/SSI participation among persons with disabilities are weak suggesting that the factors associated with the variation in disability prevalence will be different than the factors associated with DI/SSI participation described in this study.

There are strong regional patterns in the geographic variation in DI/SSI participation with higher levels of participation in southeastern regions. Our findings suggest that these regional patterns occur because of regional variation in disability prevalence and socioeconomic characteristics. This suggests that, if existent, the contributions of other regional factors that were not included in our analysis are relatively small compared to the overall regional variation in DI/SSI participation; for example, the contribution of regional variation in DI/SSI administration.

REFERENCES

Autor, D. H. & Duggan, M. G. (2003). The Rise in Disability Rolls and the Decline in Unemployment. Quarterly Journal of Economics. Quarterly Journal of Economics, 118(1), 157-206.

Ben-Shalom, Y. & Stapleton, D. (2014). *National and State Program Participation Ratios for Working-Age People with Disabilities.* Washington, DC: Mathematica Policy Research. http://www.disabilitypolicyresearch.org/~/media/publications/pdfs/disability/program_participation_ratios.pdf

Black,D., Daniel, K. & Sanders, S. (2002). The Impact of Economic Conditions on Participation in Disability Programs: Evidence from the Coal Boom and Bust. American Economic Review, 92(1), 27-50.

Burkhauser, R. V., Fisher, T. L. Houtenville, A. J. & Tennant, J. R. (2012). *Using the 2009 CPS-ASEC-SSA Matched Dataset to Show Who is and is Not Captured in the Official Six-Question Sequence on Disability*. Paper presented at the 14th Annual Joint Conference of the Retirement Research Consortium, Washington, DC, August 2012. http://www.nber.org/programs/ag/rrc/rrc2012/papers/4.3%20Burkhauser%20Fisher%20Houtenville.pdf

Coe, N. B., Haverstick, K., Munnell, A. H. & Webb, A. (2011). *What Explains State Variation in SSDI Application Rates?* Chestnut Hill, MA: Center for Retirement Research at Boston College. http://crr.bc.edu/wp-content/uploads/2011/11/wp_2011-23-508.pdf

Census Bureau. (2012) 2009-2011 *PUMS Accuracy of the Data*. http://www2.census.gov/programs-surveys/acs/tech_docs/pums/accuracy/2009_2011AccuracyPUMS.pdf

Daly, M. C, Lucking, B., & Schwabish, J. A. (2013). *FRBSF Economic Letter: The Future of Social Security Disability Insurance.* San Francisco: Federal Reserve Bank of San Francisco. http://www.frbsf.org/economic-research/files/el2013-17.pdf].K

Hyslop, R. & Imbens, G. W. (2001). Bias in Classical and Other Forms of Measurement Error. Journal of Business and Economic Statistics, 19(4), 475-481

Liebman, J. B. (2015). Understanding the Increase in Disability Insurance Benefit Receipt in the United States. Journal of Economic Perpectives, 29(2), 123-150

Social Security Administration (2011*) Annual Statistical Supplement, 2011*. https://www.ssa.gov/policy/docs/statcomps/supplement/2011/index.html

Social Security Administration (2015*) Annual Statistical Report on the Social Security Disability Insurance Program*, *2014*. https://www.ssa.gov/policy/docs/statcomps/di_asr/

APPENDIX

Table A.1. SSA Administrative Data Sources

Variable	Geographic Region	Source	Table	Column
SSI participation	State	SSI Annual Statistical Report, 2011	Table 10. Recipients, by state or other area, eligibility category, and age, December 2011	Age 18-64
SSI participation by age group	State	Supplemental Security Income (SSI) recipient by geographic area, sex, age, eligibility and diagnostic group, 2010 data http://www.ssa.gov/policy/docs/data/ssi-2010/	SSI-State-Age-2010.csv	Recipients
SSI participation	County	SSI Recipients by State and County, 2011	Table 3. Number of recipients in state (by eligibility category, age, and receipt of OASDI benefits) and amount of payments by county, December 2011	Age 18-64
DI participation	State	OASDI Beneficiaries by State and County, 2011	Table 2. Number of beneficiaries in current-payment status, by state or other area, type of benefit, and sex of beneficiaries aged 65 or older, December 2011	Disabled Workers
DI participation by age group	State	Annual Statistical Report on the Social Security Disability Insurance Program, 2011	Table 27. Number, by sex, state or other area, and age, December 2011.	
DI participation	County	OASDI Beneficiaries by State and County, 2011	Table 4. Number of beneficiaries in current payment status, by county, type of benefit, and sex of beneficiaries aged 65 or older	Disabled workers (no option for total)
Combined participation	State	SSI Annual Statistical Report, 2011	Table 10. Recipients, by state or other area, eligibility category, and age, December 2011	Age 18-64
	State	SSI Annual Statistical Report, 2011	Table 16. Persons aged 18-64 receiving both Social Security and SSI on the basis of disability and their average monthly Social Security benefit and SSI payment, by state or other area and type of beneficiary, December 2011	Number of SSI recipients with Social Security disability Workers
	State	OASDI Beneficiaries by State and County, 2011	Table 2. Number of beneficiaries in current-payment status, by state or other area, type of benefit, and sex of beneficiaries aged 65 or older, December 2011	Disabled Workers
Combined participation	County	SSI Recipients by State and County, 2011	Table 3. Number of recipients in state (by eligibility category, age, and receipt of OASDI benefits) and amount of payments by county, December 2011	Age 18-64

Variable	Geographic Region	Source	Table	Column
	County	Custom SSA Report, ssistco3supp.xlsx; December 2014	Comparable to Table 16, SSI Annual Statistical Report, 2014	Total
	County	OASDI Beneficiaries by State and County, 2011	Table 4. Number of beneficiaries in current payment status, by county, type of benefit, and sex of beneficiaries aged 65 or older	Disabled workers (no option for total)

Figure A.1. Proportion of working-age persons participating in DI, CAPUMA level

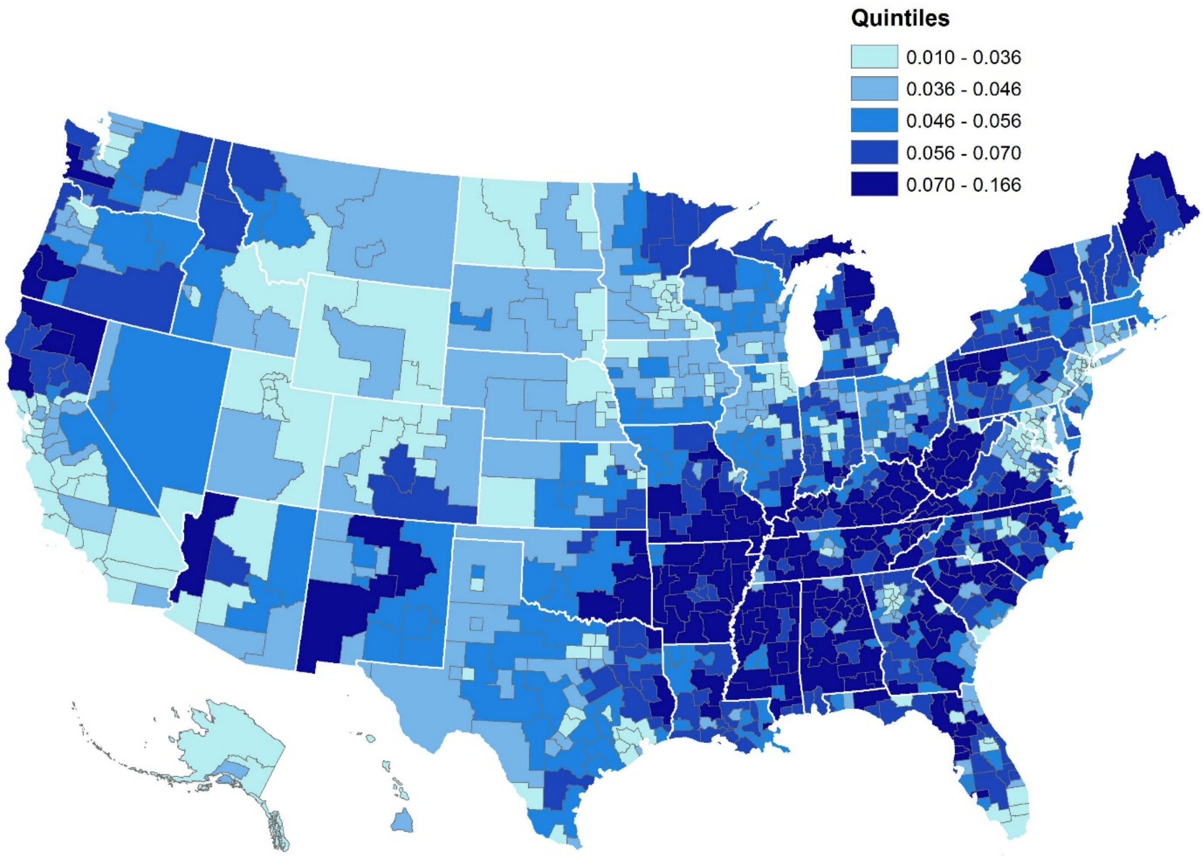


Figure A.2. Proportion of working-age persons participating in SSI, CAPUMA level

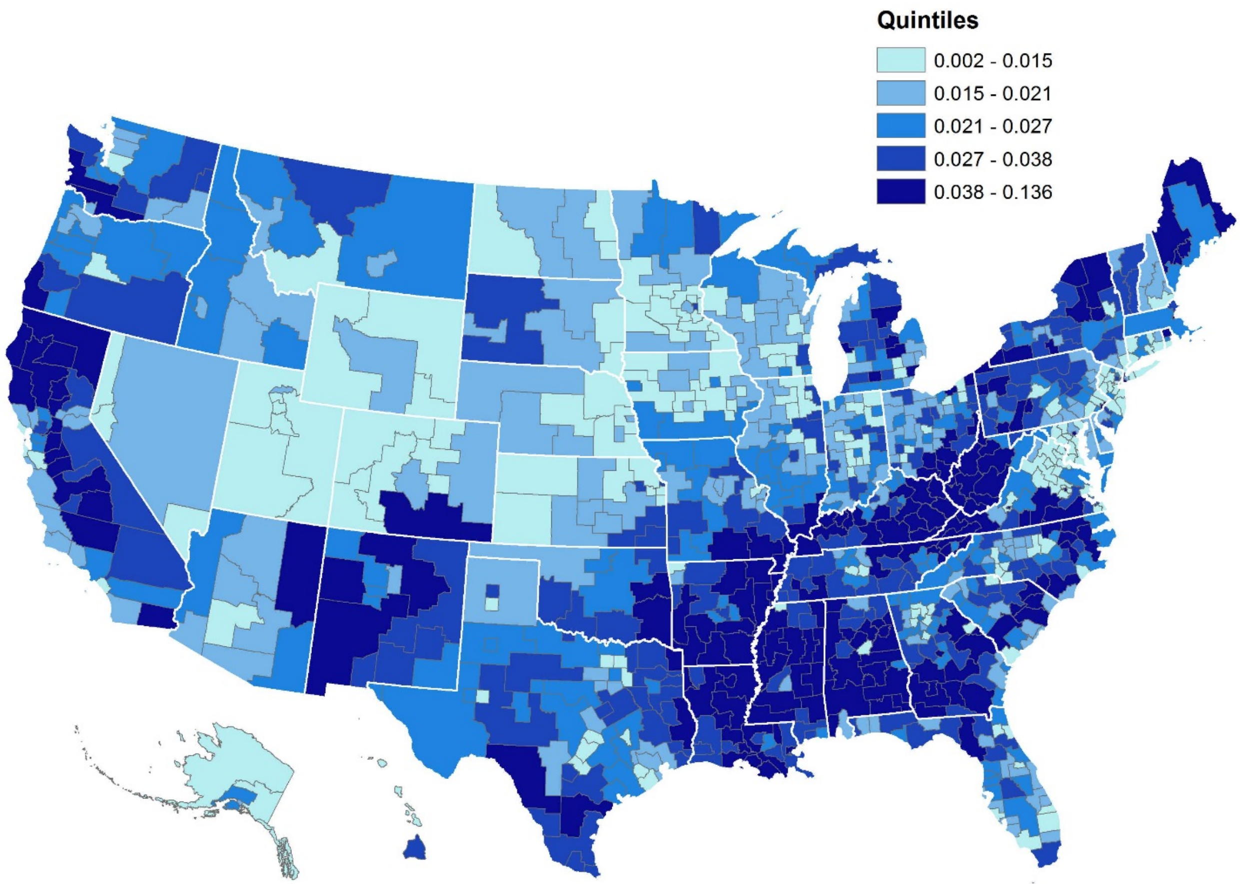


Figure A.3. Proportion of working-age persons with disabilities, CAPUMA level

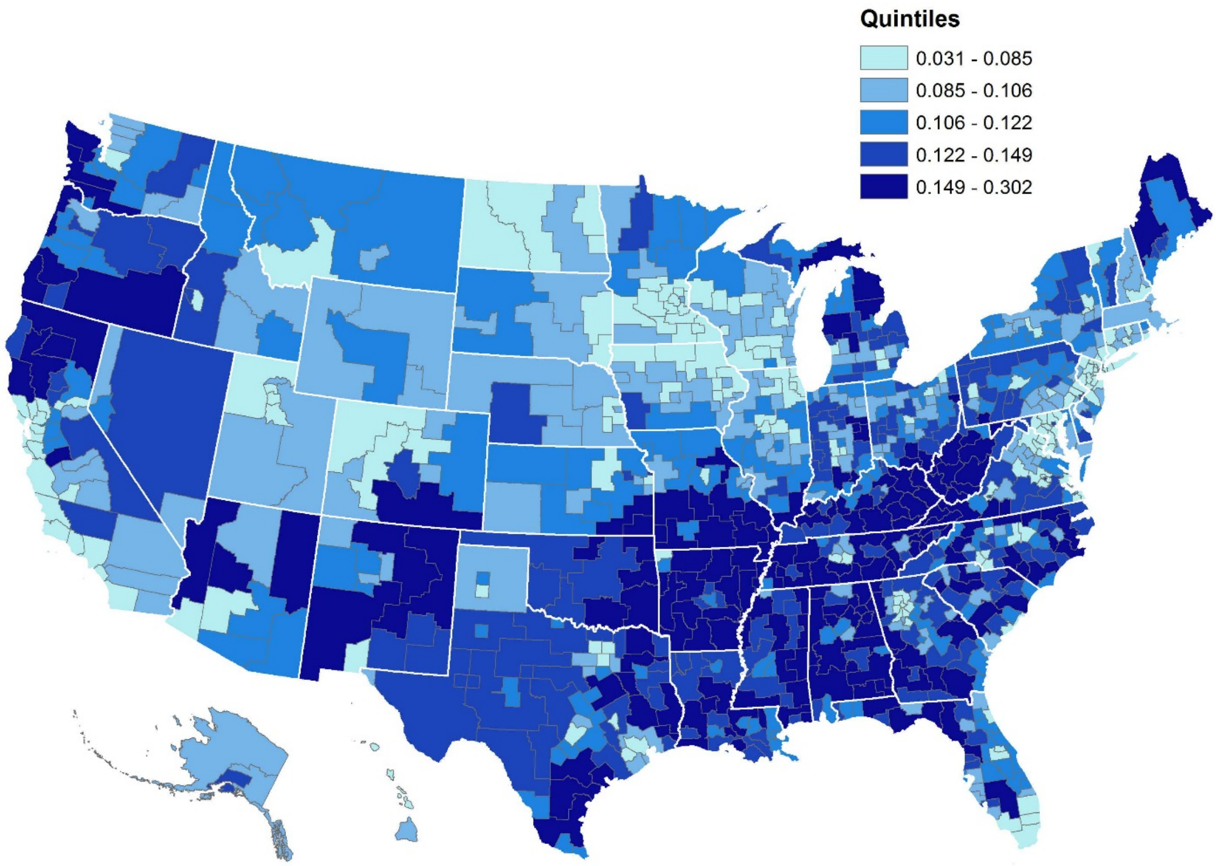


Figure A.4. Proportion of working-age persons with disabilities participating in DI, CAPUMA level

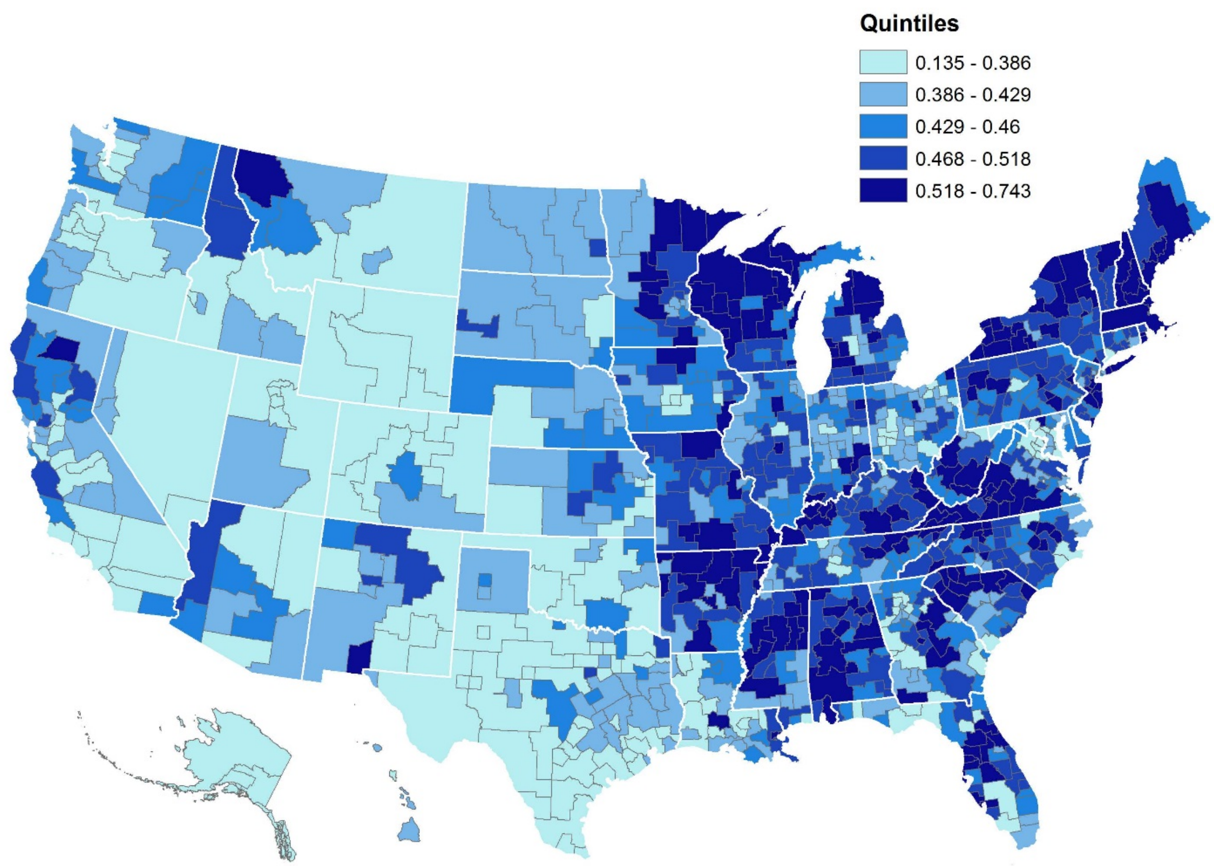


Figure A.5. Proportion of working-age persons with disabilities participating in SSI, CAPUMA level

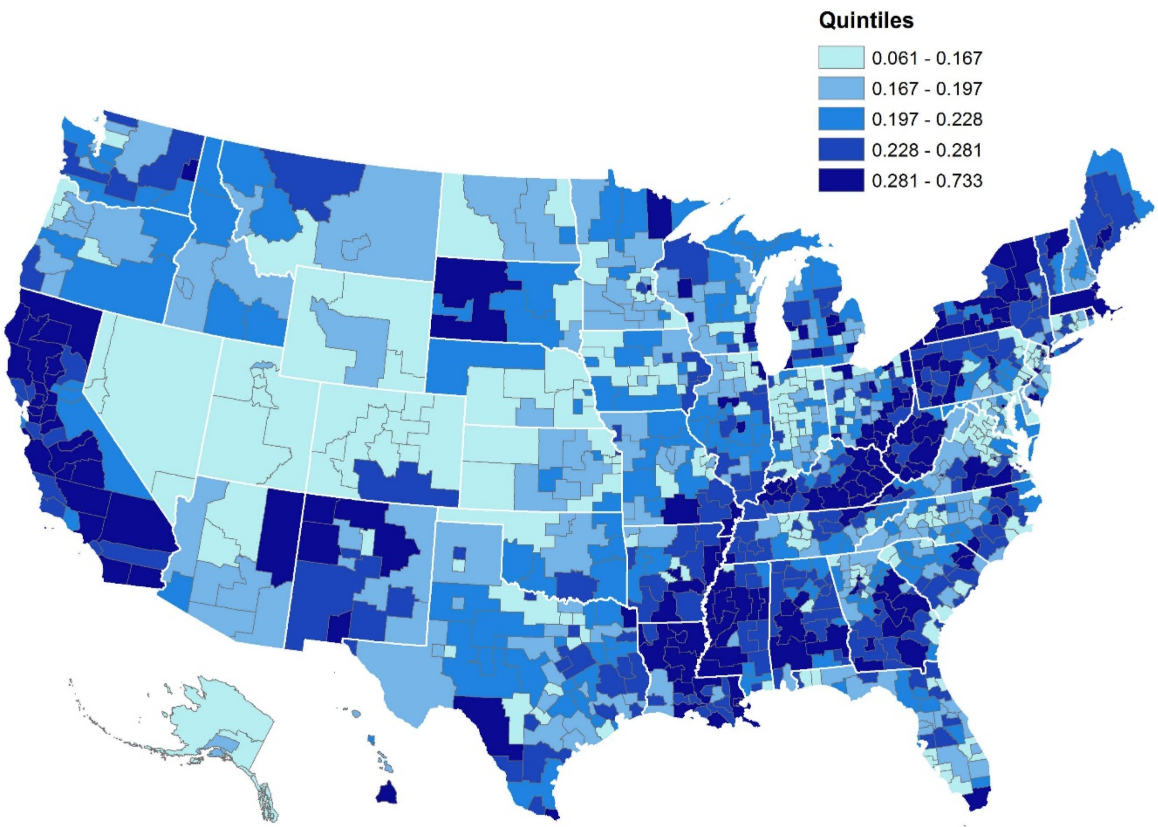


Figure A.6. PUMA boundaries and major cities

www.mathematica-mpr.com

Improving public well-being by conducting high quality, objective research and surveys

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ WASHINGTON, DC

MATHEMATICA CENTER FOR STUDYING DISABILITY POLICY