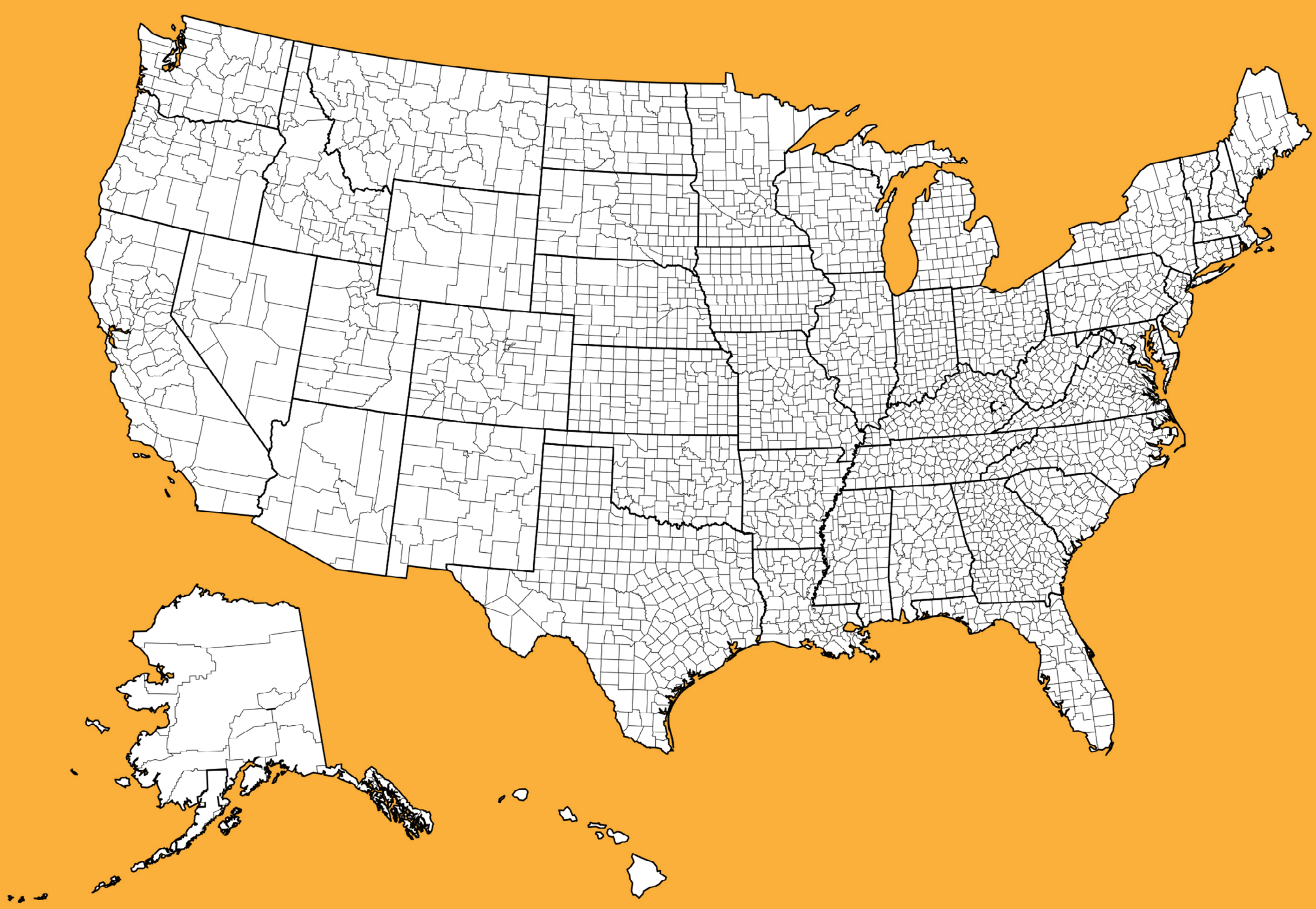
IES Institute of Education Sciences

Program for the International Assessment of Adult Competencies (PIAAC)

State and County Indirect Estimation Methodology Report

NCES 2020-225
U.S. DEPARTMENT OF EDUCATION

A Publication of the National Center for Education Statistics at IES



Program for the International Assessment of Adult Competencies (PIAAC)

State and County Indirect Estimation Methodology Report

APRIL 2020

Tom Krenzke
Leyla Mohadjer
Jianzhu Li
Andreea Erciulescu
Robert Fay
Weijia Ren
Wendy Van de Kerckhove
Lin Li
Westat

J.N.K. Rao
Carleton University, Ottawa, Canada

Holly Xie
Stephen Provasnik
Project Officers
National Center for Education Statistics

NCES 2020-225
U.S. DEPARTMENT OF EDUCATION

IES Institute of Education Sciences

U.S. Department of Education
Betsy DeVos
Secretary

Institute of Education Sciences
Mark Schneider
Director

National Center for Education Statistics
James L. Woodworth
Commissioner

The National Center for Education Statistics (NCES) is the primary federal entity for collecting, analyzing, and reporting data related to education in the United States and other nations. It fulfills a congressional mandate to collect, collate, analyze, and report full and complete statistics on the condition of education in the United States; conduct and publish reports and specialized analyses of the meaning and significance of such statistics; assist state and local education agencies in improving their statistical systems; and review and report on education activities in foreign countries.

NCES activities are designed to address high-priority education data needs; provide consistent, reliable, complete, and accurate indicators of education status and trends; and report timely, useful, and high-quality data to the U.S. Department of Education, the Congress, the states, other education policymakers, practitioners, data users, and the general public. Unless specifically noted, all information contained herein is in the public domain.

We strive to make our products available in a variety of formats and in language that is appropriate to a variety of audiences. You, as our customer, are the best judge of our success in communicating information effectively. If you have any comments or suggestions about this or any other NCES product or report, we would like to hear from you. Please direct your comments to

NCES, IES, U.S. Department of Education
Potomac Center Plaza (PCP)
550 12th Street SW
Washington, DC 20202

April 2020

The NCES Home Page address is https://nces.ed.gov.
The NCES Publications and Products address is https://nces.ed.gov/pubsearch.

This report was prepared for the National Center for Education Statistics under Contract No. ED-IES-12-C-0072 with Westat. Mention of trade names, commercial products, or organizations does not imply endorsement by the U.S. Government.

Suggested Citation
Krenzke, T., Mohadjer, L., Li, J., Erciulescu, A., Fay, R., Ren, W., Van de Kerckhove, W., Li, L., and Rao, J.N.K. (2020). *Program for the International Assessment of Adult Competencies (PIAAC): State and County Indirect Estimation Methodology Report* (NCES 2020-225). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office. Retrieved [date] from https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2020225.

Content Contact
Holly Xie
(202) 245-8481
Holly.Xie@ed.gov

ACKNOWLEDGMENTS

The authors are grateful for the opportunity and guidance provided by the National Center for Education Statistics, in particular Holly Xie and Stephen Provasnik. An expert panel was convened in the early stages that provided some important feedback. Panel members included William Bell, Andreea Erciulescu (prior to employment at Westat), Partha Lahiri, Danny Pfeffermann, Dan Sherman, and Avi Singh. At Westat, the efficient programming skills of Ying Chen, who led the statistical programming necessary for the production runs, was invaluable.

CONTENTS

CONTENTS—Continued

CONTENTS—Continued

LIST OF TABLES

CONTENTS—Continued

List of Tables—Continued

List of Tables—Continued

CONTENTS—Continued

List of Tables—Continued

LIST OF FIGURES

CONTENTS—Continued

List of Figures—Continued

CONTENTS—Continued

List of Figures—Continued

CONTENTS—Continued

List of Figures—Continued

1. INTRODUCTION

The Program for the International Assessment of Adult Competencies (PIAAC) is a multicycle survey of adult skills and competencies sponsored by the Organization for Economic Cooperation and Development (OECD). The survey examines a range of basic skills in the information age and assesses these adult skills consistently across participating countries. The first cycle of PIAAC included three rounds: 24 countries participated in 2011–12 (round 1); 9 additional countries participated in 2014–15 (round 2); and 5 additional countries participated in 2017–18 (round 3).

The United States has participated in all three rounds of the first cycle of PIAAC. The round 1 (PIAAC 2012) survey design was consistent with the international requirements (OECD 2016). In round 2 (PIAAC 2014), a supplemental sample was drawn to enhance the round 1 sample (Hogan et al. 2016). The combined PIAAC 2012/2014 sample is nationally representative of the U.S. adult population 16–74 years old. The round 3 (PIAAC 2017) data collection had two core objectives. First, it was designed to produce a nationally representative sample of the U.S. adult population 16–74 years old. Second, the sample was designed to arrive at a large enough sample size that, when combined with the 2012 and 2014 samples, can produce small area estimates for the U.S. counties.

PIAAC is the sixth of a series of adult skills surveys, sponsored by the National Center for Education Statistics (NCES), that have been implemented in the United States. It is preceded by the Young Adult Literacy Survey (YALS) in 1985, the National Adult Literacy Survey (NALS) in 1992, the International Adult Literacy Survey (IALS) in 1994, the National Assessment of Adult Literacy (NAAL) in 2003, and the International Survey of Adult Literacy and Life Skills (ALL) in 2003.[1] In 2009, NCES published model-dependent estimates for states and counties using the NALS and NAAL data (available at https://nces.ed.gov/naal/estimates/Index.aspx). These estimates were called "indirect" estimates to distinguish them from standard or "direct" estimates that do not depend on the validity of a statistical model. The 2009 state and county estimates were produced using small area estimation (SAE) techniques that rely on survey data as well as data from other sources such as decennial censuses (1990 and 2000) for each of the two survey years.

The demand for reliable small area estimates has greatly increased in the past decades (see for example, Czajka, Sukasih, and Maccarone 2014). Many federal agencies have been producing estimates for states and counties so that policymakers can plan and allocate resources and target interventions as necessary. Some examples include: Small Area Income and Poverty Estimates (SAIPE); Small Area Health

[1] A history of the NCES-sponsored adult literacy surveys is available at https://nces.ed.gov/surveys/piaac/history.asp.

Insurance Estimates (SAHIE); state and local area employment and unemployment statistics; county estimates of diabetes prevalence, incidence, and risk factors; state and county estimates of cancer risk factors and screening; and state estimates from the National Survey of Drug Use and Health. Similarly, access to information about proficiency levels of adults' literacy and numeracy skills at the state and county levels has been essential for policymakers, educators, and researchers to evaluate the distribution of proficiency levels across various areas, understand variables impacting literacy, and develop programs aimed at improving the proficiency levels of adults in the United States. See Czajka, Sukasih, and Maccarone (2014) for general descriptions of these SAE programs as well as additional examples of other programs across different agencies, and refer to sections 2.1 and 2.2 for further descriptions of a selected number of SAE programs across federal agencies. As the demand for reliable small area estimates has increased in the past decades, the SAE literature and research findings also grew rapidly, with significant enhancements made in the methodology and approaches (see Rao and Molina 2015) since the last round of SAE was produced using data from the 1992 NALS and 2003 NAAL surveys. This report describes the advanced statistical methodology used to produce state and county indirect estimates of various proficiency levels of adults for individual states and counties using data from the first cycle of PIAAC, using the SAE approach. As mentioned above, SAE is a model-dependent approach that produces indirect estimates for areas where survey data are insufficient for direct estimation. SAE models "borrow strength" across related small areas through auxiliary information to produce reliable "indirect" estimates for small areas. The SAE models rely on covariates available at the small area level and PIAAC survey data. In addition, small area models make specific allowance for between-area random effects that account for between-area variances beyond what is explained by covariates (e.g., percentage with less than a high school diploma) included in the model. Based on the results of these models, NCES derived small area estimates for all states and counties in the United States and produced a tool called the "U.S. PIAAC Skills Map: State and County Indicators of Adult Literacy and Numeracy" to view heat maps (i.e., data values represented by shades of colors) and to compare proficiency estimates across states or counties. The precision measures associated with each indirect estimate are based on the sophisticated statistical methodology that attempts to account for all sources of error.

In the absence of any other literacy assessment data available for individual states and counties, the indirect estimates provide a general picture of literacy for all states and counties. Lacking these model-dependent estimates, covariates highly related to literacy and numeracy, such as educational attainment and poverty, have generally been used as proxy indicators of state and county proficiency levels. The estimates presented in the Skills Map website were developed using data from the actual assessments administered in the PIAAC survey and covariate data from the American Community Survey, which is administered by the Census Bureau. The estimates are thus predictions of how the adults in a state or county would have performed had they been administered the PIAAC assessment.

The remainder of this chapter contains a brief review of the PIAAC sample design, a description of proficiency measures in PIAAC, and the type of estimates and local areas for which small area estimates have been produced. A brief description of the website that hosts the small area estimates is also included in this chapter.

1.1 PIAAC Sample Design

Over the course of the years 2012 to 2017, PIAAC surveyed individuals 16–74 years old[2] in the United States as part of an international study involving over 35 countries. Data collected and combined over these years were, through the use of sophisticated statistical methods, used to produce indirect (also referred to as small area) estimates of literacy and numeracy proficiency. The combined household sample was created from three data collection efforts that took place in 2012, 2014, and 2017. In each year, a four-stage stratified area probability sample was selected. In the first stage, primary sampling units (PSUs) were selected, consisting of individual counties or groups of contiguous counties. The 2014 sample was designed as a supplement to 2012 and used the same sampled PSUs as 2012. The PSUs for 2017 were selected in such a way as to reduce overlap with the PSUs in 2012/2014. In the second stage, secondary sampling units (SSUs) were selected, consisting of Decennial Census blocks or block groups. In the third stage, dwelling units (DUs) were selected. A Screener interview was administered to each DU, and used to identify the eligible persons within selected DUs. At the fourth stage, a sampling algorithm was implemented within the computer-assisted personal interviewing (CAPI) system to select one or more sample persons among those identified to be eligible. Once selected, the Background Questionnaire (BQ) interview was completed. Upon completion of the BQ, the respondents were provided either the paper-and-pencil or computer-based assessment, based on whether they reported having any previous computer experience during the BQ interview or whether they refused the computer-based assessment as well as their performance on the Information and Communication Technology (ICT) core instrument. Together, the combined sample has 12,330 respondents from 185 counties. A set of weights for the combined PIAAC 2012/2014/2017 sample was created for the purpose of creating small area estimates. The weights for the PIAAC 2012/2014/2017 sample were created by combining the final PIAAC 2012/2014 and PIAAC 2017 weights and calibrating to population totals. The PIAAC Technical Report provides details of the sample designs for each data collection year in chapter 3, and discusses the weighting processes in chapter 8.

[2] The PIAAC 2012 sample was limited to 16- to 65-year-olds.

Table 1-1 provides a summary of the sample sizes in the 2012, 2014, and 2017 surveys, which results in 12,330 completed cases in the 2012/2014/2017 combined sample.

Table 1-1. Number of completed cases for PIAAC samples: 2012/2014/2017

Age group	All years	2012	2014	2017
Total	**12,330**	**5,010**	**3,660**	**3,660**
16–65	11,217	5,010	2,911	3,296
66–74	1,113	†	749	364

† Not applicable
SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

As shown in table 1-2, of the 3,142 U.S. counties, 100 were at least a part of a sampled PSU in 2012/2014 and 103 in 2017. Table 1-2 also provides the number of counties in the 2012/2014 and 2017 samples with one or more completed cases. There are 99 counties in PIAAC rounds 1 and 2 with one or more completed cases, 99 counties with one or more completed cases in round 3, and a total of 185 unique counties with one or more completed cases. Table 1-3 provides a breakdown of the county sample sizes for the combined 2012/2014/2017 sample.

Table 1-2. Number of counties with at least one completed case: 2012/2014/2017

Sample year	Number of counties as part of a sampled PSU	Number of counties with completed cases
2012/2014	100	99
2017	103	99
Combined 2012/2014/2017	190	185

SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

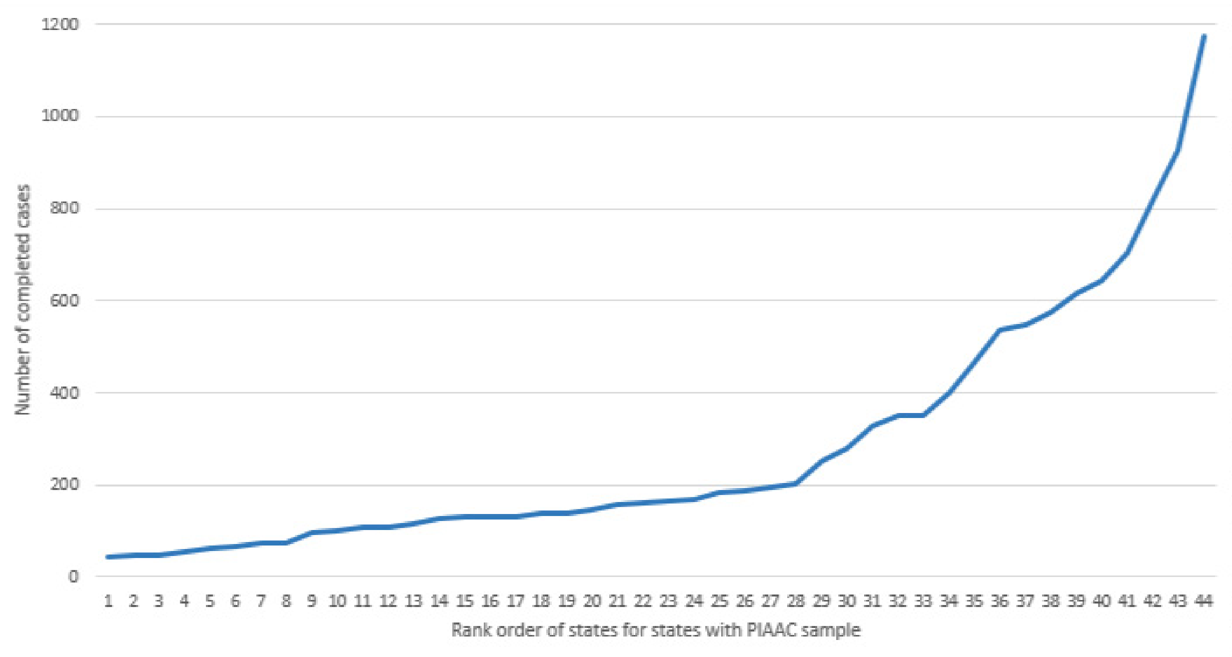
Table 1-3. Number of completed cases per county: 2012/2014/2017

Number of completed cases	Number of counties
Total	**185**
Less than 5	4
5 to 10	14
11 to 20	10
21 to 50	58
51 to 100	56
101 or more	43

SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

Of the 50 states, plus the District of Columbia, 44 have completed cases in the combined 2012/2014/2017 sample. The number of completed cases by state ranges from just over 40 to almost 1,200. This is illustrated in figure 1-1.

Figure 1-1. Number of completed cases by states with at least one completed case, sorted by number of completed cases: 2012/2014/2017



SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

1.2 Proficiency Measures in PIAAC

The first cycle of PIAAC assessed three domains of cognitive skill:

- literacy (including reading component);
- numeracy; and
- problem solving in technology-rich environment.

The test design for PIAAC is based on an approach most common to the major large-scale surveys. This design, called matrix sampling, entails administering a subset of items from a larger item pool, with different groups of respondents answering different sets of items. This design allows a reduction in the response burden for an individual, and at the same time makes it inappropriate to use any statistic based on the number of correct responses in reporting the survey results. This limitation is overcome through item response theory (IRT) scaling used to derive scores for each domain. To provide a measure of the uncertainty of the cognitive measurement, PIAAC uses 10 plausible values (multiple imputations) drawn from a posterior distribution based on the IRT scaling of the cognitive items with a latent regression model using information from the BQ in a population model. For more details about the IRT models and the model equations see the PIAAC Main Study technical report (Yamamoto, Khorramdel, and von Davier 2013).

1.3 Types of Estimates, Areas of Interest, and Results Website

County- and state-level estimates of adult literacy and numeracy proficiency are produced for the proportion at or below Level 1, the proportion above Level 1 and below Level 3 (referred to as "proportion at Level 2"), proportion at Level 3 and above, and the average. Discussion and illustrations in the remaining chapters mainly focus on one estimate (percentage at or below Level 1 of literacy). Results pertaining to other small area estimates are given in appendix C.

The county and state estimates are published in the Skills Map website at http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2020047. The website uses a visualization-based system that allows users access to the small area estimates through heat maps and summary card displays. This user-friendly website provides precision estimates and facilitates statistical comparisons among counties and states. The summary cards provide descriptions of the small areas in terms of all eight outcomes, and can show two selected states and/or two selected counties in a side-by-side comparison. There is also ability to see a comparison of distributions (from the American Community Survey) for select demographics to help in explaining the SAE model-based results. Summary text discusses results from statistical comparisons between areas on eight proficiency outcomes. That is, for both Literacy and Numeracy, comparisons will be conducted on the proportion at or below Level 1, proportion at Level 2, proportion at Level 3 and above, and the average score. The areas involved in the comparisons cover the following:

- state-to-nation;
- state-to-state;
- county-to-state;
- county-to-county within state; and
- county-to-county across states.

Reports are available on the website that visually display the estimates for all counties within a state, and all states within the United States.

The remainder of this state and county indirect estimates methodology report is divided into seven chapters and three appendixes.

- Chapter 2 provides a review of the methodology (as relevant to PIAAC) with focus on new developments in the SAE area since the NAAL model development in 2003, and summarizes the results of our review of such recent literature, including both unit-level and area-level modeling. In addition, this section includes a literature review of the methodologies used in recently published SAE estimates produced for some major federal statistics in the United States, and discusses the relationship and relevance to PIAAC SAE. An important component of SAE modeling is the direct survey estimates.

- Chapter 3 provides a summary of the approaches used in the computation of direct estimates, their variances, and the steps involved in preparing the estimates for the SAE modelling step including the application of survey regression estimation (SRE) approach for calibrating the weights to county level totals, pooling plausible values into one point estimate and one variance estimate to reflect the plausible value variations in the estimation, and smoothing the variance to improve the stability of the variance estimates.

- Another key component of SAE modeling is access to predictor variables that are measured consistently across all counties and states, and that are effective predictors of the estimated proportion of adults at different levels or average of literacy and numeracy. Because of the small number of data points (counties with PIAAC sample) for the model development, a small number of covariates needed to be selected from the large pool of covariates. Chapter 4 contains a listing of the sources of covariates for model development, along with a description of the approach used for arriving at a final set of variables for the model development stage.

- SAE model estimation, and prediction and aggregation are described in chapter 5. The model estimation is conducted on the small number of counties with PIAAC sample. The resulting model is used to make predictions for all 3,142 counties. Weighted aggregations of the county predictions lead to the state estimates. The model estimation and prediction process produce precision estimates that account for various sources of error, including sampling error, imputation error and modeling error. A brief description of a simulation study designed to evaluate the relative performance of estimates under a selected subset of various models and software is also given in this chapter. Often, users may be interested in conducting multiple comparisons to make simultaneous inferences. Therefore, the PIAAC state and county indirect estimates website provides comparisons between areas on eight outcomes. That is, for each of literacy and numeracy, comparisons are conducted on the proportion at or below Level 1, proportion at Level 2, proportion at Level 3 and above, and the average. Finally, section 5.6 contains a description of the methodology used for making the simultaneous inferences included in the PIAAC website.

- Large-scale SAE programs generally employ an extensive model evaluation process since models are never a perfect fit to the data, and systematic errors can manifest themselves. It is especially important for PIAAC to conduct a thorough evaluation of the model development since over 90 percent of the county estimates rely solely on the model predictions. Chapter 6 contains the highlights of the full range of model diagnostics, sensitivity analysis, and evaluation results. In the chapter are some important evaluation results, such as in figure 6-16, which shows that the model predictions are generally close to the survey regression estimates. Another example is figure 6-17, which shows the reduction in mean square error from the model when compared to the survey regression estimates.

The three appendixes include the list of potential covariates (appendix A), the simulation study results (appendix B), and the preliminary study results (appendix C). A list of references follows chapter 7.

2. BACKGROUND

The term small area estimation (SAE) refers to a variety of methods or statistical techniques to estimate information or, more precisely, parameters for subpopulations or smaller areas of interest. SAE uses survey data in combination with auxiliary data at the small area level from other sources to model the estimates of interest. A wide variety of models have been developed for generating small area estimates; the two major types of models are known as *area level* and *unit level*.[3] The area-level approach models the small area estimate of interest in terms of auxiliary data at the area level, whereas the unit-level approach models the underlying variable of interest in terms of unit-level auxiliary data, and then aggregating the individual predictions for each small area. Two major schools of thought underlie the proliferation of techniques and models: frequentist and Bayesian. Among the frequentist is the Empirical Best Linear Unbiased Predictor (EBLUP), which can be used to estimate random effects, such as through SAS Proc Mixed. The Linear Mixed Model (LMM) may be used when the dependent variable follows a normal distribution. The Generalized Linear Mixed Model (GLMM) is considered when the dependent variable has a nonnormal error distribution. Empirical Bayes is considered a frequentist approach and is conducted when the prior distribution is based on the data itself. Among the Bayesian approaches, Hierarchical Bayes (HB) is applied when there is a fixed prior distribution. The reader can find many details on the various types of SAE models in Rao and Molina (2015).

A thorough literature review was conducted prior to developing the Program for the International Assessment of Adult Competencies (PIAAC) SAE methodology, in addition to organizing a U.S. PIAAC Summit of International SAE Experts (including William Bell, Partha Lahiri, Danny Pfefferman, Jon Rao) and those on the PIAAC team (e.g., Robert Fay) to discuss modeling issues relevant to PIAAC SAE (more discussion is in section 2.4). A summary of the literature review, including reviews of the federal SAE programs is given in Krenzke et al. (2018). The remainder of this chapter contains the highlights of the literature review as relevant to this report. First, a brief summary of model-based approaches applied to literacy data is given in section 2.1. Next, a short summary of our review of recent SAE developments and methodologies used to produce SAE official statistics published by some major federal statistical agencies in the United States is given in section 2.2. Section 2.3 provides a brief summary of the considerable amount of research and development in SAE methods that has taken place in the past decade.

[3] "Unit level" typically refers to an individual sample unit level (e.g., person or household), or it could mean a geographic area that is constitutive of the target "small area."

A systematic decisionmaking process was critical for the research and development of the PIAAC SAE models. It began with the literature review of recent advances in SAE models, followed by the U.S. PIAAC Summit of International SAE Experts, which provided various recommendations (e.g., determine level-of-effort for modeling different levels of proficiency, incorporate cross-validation into the simulation, investigate different models, look into variable selection through various approaches). Further investigation occurred, including a simulation study (described in appendix B) that provided key findings, and follow-ups led to decisions related to the recommendations made by the International SAE Experts. Section 2.4 provides more information about the topics that were addressed during the research and development that led to the final PIAAC SAE models.

2.1 Approaches Applied to Literacy Data

The National Center for Education Statistics (NCES) produced state- and county-level estimates of the proportions of adults lacking basic prose literacy skills for the 2003 National Assessment of Adult Literacy (NAAL) and 1992 National Adult Literacy Survey (NALS) (Mohadjer et al. 2009) using SAE methodology. The model was developed using a HB unmatched area-level model based on the 2003 NAAL and auxiliary data from the 2000 census (Mohadjer et al. 2011).[4] The aim of the NAAL SAE model was to estimate the true proportion of adults who are lacking basic prose literacy skills (as evaluated by the NAAL instrument) for all counties and states in the United States. The NAAL and NALS county and state estimates of the proportions of adults lacking basic prose literacy skills are available at https://nces.ed.gov/naal/estimates/Index.aspx. The website allows users to compare the proportions of adults lacking basic prose literacy skills across any two states or counties that are specified in advance, and across years for a single state or county. It should be pointed out that the proficiency assessment instruments and scales used in NAAL and NALS are different from those used in PIAAC, and thus the small area estimates for counties and states from NAAL and NALS SAE are not comparable with the corresponding estimates from PIAAC. The only two years available in the data tool are the 2003 NAAL and 1992 NALS.

Other applications of model-based methods applied to literacy data include Gibson and Hewson (2012), where the authors used a unit-level, nonlinear (logistic) model, for literacy and numeracy binomial outcomes for the 2011 Skills for Life Survey that was conducted in the United Kingdom. Yamamoto (2014) presents a model-dependent approach to produce estimates of skill distribution for provinces based on population parameters derived from the Canadian PIAAC data and auxiliary information such as

[4] The process was repeated for the 1992 National Adult Literacy Survey (NALS).

census. In addition, Bijlsma et al. (2017) use the Netherlands' PIAAC data and focus on obtaining the literacy estimates at the municipality level in the Netherlands using model-based SAE techniques in an HB framework. They used a basic unit-level model originally proposed by Battese, Harter, and Fuller (1988) to model the average literacy score since literacy scores are continuous per individual and area and assumed to have a linear relation with individual-level covariates. An area-level model originally proposed by Fay and Herriot (1979) was used to model the proportion of low literates.

2.2 Review of Major Federal SAE Programs

As noted in Czajka, Sukasih, and Maccarone (2014), a number of federal statistical agencies have developed SAE methods and have published SAE estimates. This section contains brief reviews of a select number of these programs that are thought to have some relevance to the SAE efforts for the PIAAC survey, and thus could be helpful in guiding the decisions about final plans for PIAAC.

The Census Bureau's Small Area Income and Poverty Estimates (SAIPE) program produces a number of income and poverty-related estimates for states, counties, and school districts using area-level models based on the American Community Survey (ACS) and small area auxiliary data from the Internal Revenue Service and other sources (https://www.census.gov/programs-surveys/saipe.html). Examples of the model-based small area estimates produced by the SAIPE program include median household income and the proportions and numbers of total persons, children, and school-age children below the poverty level. These annual estimates are used to allocate funding for many federal grant programs to state and local jurisdictions, including nutrition assistance, medical assistance, jobs training, housing, and education (see, for example, Citro and Kalton 2000, chapter 2).

The Office of Applied Studies at the Substance Abuse and Mental Health Services Administration regularly produces substate-level estimates of substance abuse using unit-level models based on the National Survey of Drug Use and Health (NSDUH) and a wide variety of sources for the auxiliary data.[5] As in the NAAL, a multistage sample design was used for the NSDUH. The primary sampling units (PSUs) are collections of adjacent census block groups. The NSDUH small area model differs from the SAIPE and NAAL models in that it is an individual person-level (unit-level) model. Refer to Folsom, Shah, and Vaish (1999) for more detail about the NSDUH approach to SAE.

[5] Documentation of the methodology is available at: https://www.samhsa.gov/data/sites/default/files/substate2k12-Methodology/NSDUHsubstateMethodology2012.htm.

Raghunathan et al. (2007) provide an application that includes a large national survey (National Health Interview Survey [NHIS]) and a supplementary survey (Behavior Risk Factor Surveillance Survey [BRFSS]) where small area parameters are jointly estimated in a small area model. Other specialized approaches include the SAE work that has been conducted for the National Crime Victimization Survey to produce state-level time series estimates of victimization and for large metropolitan areas. The sources for the covariates include the ACS and the Federal Bureau of Investigation's *Uniform Crime Reports*. More information can be found in Fay and Diallo (2015), Fay, Planty, and Diallo (2013), Fay and Diallo (2012), and Fay and Li (2011).[6] Also, as mentioned in Bauder, Luery, and Szelepka (2016), the Small Area Health Insurance Estimates (SAHIE) program at the Census Bureau produces SAE estimates of numbers and proportions of those with and without health insurance coverage for demographic groups within states and counties. The demographic groups are defined by age, sex, and income, and in addition, for states by race and ethnicity. Income groups are defined in terms of income-to-poverty ratio (IPR), which is the family income divided by the appropriate federal poverty level.

2.3 Review of Recent Developments in SAE

A considerable amount of research and development in SAE methods has taken place since the development of SAE procedures for the 2003 NAAL. The text by Rao (2003) presents a comprehensive overview of the methods, history, and applications of SAE methods. The book has since been updated (Rao and Molina 2015). As the demand for reliable small area estimates has greatly increased in the past decades, the SAE literature and research findings also grew rapidly. Pfefferman (2013) reviews and discusses some of the important new developments in SAE method since the publication of Rao's *Small Area Estimation* in 2003. The review covers both design-based (frequentist) and model-dependent (frequentist and Bayesian) methods. Pfefferman (2013) also reviews new developments on SAE under informative sampling and nonresponse as well as model selection and checking. In the case of informative sampling or not-missing-at-random nonresponse, the model assumed for the population may not apply to the sample data. If not properly accounted for, the resulting predictions can be seriously biased.

The model developed by Fay and Herriot (1979) has been one of the most widely-used models in SAE. Benavent and Morales (2016) extended the Fay-Herriot model to multivariate models to estimate correlated descriptive measures. The new models have multivariate vectors of random effects with the same dimension as the target variables and allowing for different correlation structures. They found that the multivariate EBLUPs have lower mean square error (MSE) than the corresponding univariate model

[6] Published estimates can be found at: https://www.bjs.gov/index.cfm?ty=pbdetail&&iid=5499.

when the true generating model is multivariate. Univariate models may be good enough if a good set of auxiliary variables is available for modeling. Otherwise, multivariate models, taking into account additional data relationships, would work better to improve the precision of the estimates.

Pfeffermann, Terryn, and Moura (2008) considered the situation that the target response variable is a continuous variable with a large peak of zero values. This occurs in the assessment of literacy proficiency in developing countries where zero outcomes indicate illiteracy and positive scores measure the level of literacy. Their estimates of interest are average literacy scores and the proportion of positive scores in small areas. A two-part random effect model was developed and fitted to the data with mixed distribution. Part one assumes a linear mixed model for positive responses, and part two assumes a generalized linear mixed model for the probabilities of positive responses. Nonzero correlations are allowed between the random effects in the two parts. This paper concluded that fitting a linear mixed model without differentiating the large frequency of zero values from other positive values will result in highly biased predictors and wrong coverage rates of credibility intervals. Accounting for the correlations between the random effects of the two parts is the best choice, but it may only improve the predictions marginally. The magnitude of the correlations and the importance of accounting for them in the model mainly depend on the predictive power of the available covariates.

2.4 Key Features of PIAAC Indirect Estimation Methodology

The U.S. PIAAC International SAE Experts discussed the importance of striving toward a more full picture of the distribution for local areas. Therefore, estimating proportions in more than one level was conducted, specifically for the proportion at or below Level 1, proportion at Level 2, and the proportion at or above Level 3, in both literacy and numeracy, and in addition estimating proficiency averages, resulted in eight outcome measures for each state and county. An area-level bivariate HB linear three-fold model was developed for proportions (discussed in section 5.2.1), and for averages, an area-level univariate HB linear three-fold model was developed (discussed in section 5.2.2). There were many models considered to estimate the outcome measures, and the final models resulted from the following preliminary steps: (1) review applications of SAE to literacy data, review federal programs, and review recent developments in literature (as discussed above); (2) U.S. PIAAC Summit of International SAE Experts held in December 2017; (3) simulation work as documented in appendix B; and (4) working through the various steps of the process using prefinal data. Key features of the final models include the following:

- **Informative sampling and nonresponse were incorporated.** The sample of states and counties are not simple random samples; therefore, the sample design is informative. Also, weighting adjustments for nonresponse can reduce bias to the extent that the weighting variables are related to the proficiency scores. When nonresponse is not sufficiently explained by the weighting variables, informative nonresponse exists. The U.S. PIAAC International SAE Experts discussed being able to handle informative sampling and informative nonresponse because, otherwise, the process will lead toward biased estimates. The probability of selection of PSUs was included in the covariate selection process (described in chapter 4) but was not an important factor to include in the small area model. Relating to informative nonresponse, it was assumed that any literacy-related nonresponse was below Level 1, and for the estimation of averages, the first percentile of proficiency scores was imputed for literacy-related nonrespondents. This is a reasonable assumption because such nonrespondents could not complete the Background Questionnaire and assessment (conducted in English) due to a literacy-related reason (language barrier, reading/writing barrier, or mental disability). More discussion of literacy-related nonresponse is in section 3.

- **Area-level models were used.** In an area-level model, direct estimates produced at the local area-level are the prime elements in the modeling process. One part of an area-level model is a "sampling model," where survey-weighted estimates are produced for the small areas with sample-design based variance estimates. The regression model is developed using predictors at the small area level and could include variables at higher levels also. Unlike the area-level approach, the unit-level model is built at a much lower level such as individual persons or households. That is, a unit-level model uses covariates available at the person level to generate person-level values, which are aggregated to compute statistics at the area level. There is potential for smaller MSE and for producing estimates for a wide range of other subgroup of interest. The basic unit-level models ignore sample-design based variance estimates at this very low level. Extensions have included a random effect term as an attempt to capture the between area variation (see Rao and Molina 2015).

  In addition, the following discussion points occurred during the U.S. PIAAC Summit of International SAE Experts. If the model is linear, either the area-level or the unit-level approach could be used for a PIAAC small area program. The area-level approach is more design-based, since the basic building blocks are the sample (design-based) estimates at the targeted local level, as well as the sample-design based variance estimates at this level. Operationally, the area-level approach certainly works with a much simpler dataset, with one record for each local area rather than one record for each household or person, and in that sense is easier to work with in practice. While unit-level models were included in early research, for the various reasons stated above, a decision was made to move forward with the area-level models (details are provided in chapter 5).

- **HB was used.** Another discussion point during the U.S. PIAAC Summit of International SAE Experts was the use of HB models, over alternatives such as Empirical Bayes. The estimation of HB models generates Markov Chain Monte Carlo (MCMC) samples, which provide the ability to obtain good estimates of the variability of the indirect estimates by accounting for several sources of error. The model is written using a hierarchical form (a sampling level for the direct estimates of proportions and a linking level for the relationship between the target proportions and the covariates), prior distributions are adopted for the model parameters, and the Bayesian approach is used for inference, where credible intervals can accompany all point estimates. More discussion about the benefits of using HB models is provided in section 5.2.

- **Both univariate or multivariate models were used.** The area-level HB bivariate linear three-fold model estimated PIAAC proportions for at or below Level 1, and at or above Level 3. For each MCMC sample, the results were combined to estimate the proportion at Level 2, and provided credible intervals for all point estimates. The model takes advantage of the covariance between domains, which may result in reduced MSE. Due to the demands on the model fitting and small number of data points, it was decided to fit the bivariate model for proportions separately for literacy and numeracy. Relating to estimating averages, consideration was being given to estimate averages for literacy and numeracy simultaneously; however, it was decided to have separate univariate models with the same covariates, as supported by the covariate selection process.

- **Three levels of random effects were used.** Another feature of the model is the inclusion of three levels of random effects: county, state and census division.[7] The benefits of the three-fold model are (1) benchmarking[8] the estimates will not be necessary as estimates are controlled through the random effects (e.g., the aggregation of county indirect estimates within a state should align with the state indirect estimate), (2) estimates for states without sample will have some contribution from the PIAAC sample because all census divisions, have PIAAC sample, and (3) associations of counties within states, and states within divisions will have some impact while the same random effect is applied to those areas.

[7] Census divisions (New England, Middle Atlantic, East North Central, West North Central, South Atlantic, East South Central, West South Central, Mountain, and Pacific divisions) are groups of states defined by the Census Bureau.

[8] Benchmarking is sometimes conducted to align the indirect estimates to more stable national or regional estimates and to ensure consistency with known aggregate estimates.

3. DIRECT ESTIMATION

As a result of the research and development phase (summarized in section 2.4), the indirect estimation production process for the U.S. Program for the International Assessment of Adult Competencies (PIAAC) 2012/2014/2017 started with area-level modeling. The area-level modeling process began with computing direct estimates for the areas of interest. In section 3.1, we provide a summary of the methods used for the production of direct estimates and associated variances. The variances of the direct estimates and variance estimates can be large in counties with small sample sizes. As presented in section 3.2, we implemented model-assisted methods as tools for (a) improving stability of the direct estimates (through survey regression estimation [SRE] [Särndal and Hidiroglou 1989]) and (b) for smoothing the variances of the direct estimates (through generalized variance functions [Wolter 2007]). The direct estimates after the SRE and variance smoothing process served as inputs to the small area estimation (SAE) models.

Variance estimates are intended to account for the error associated with the item response theory (IRT) modeling in addition to the sampling error. As described in chapter 10 of Hogan et al. (2016), different PIAAC respondents took different sets of items that could be of various levels of difficulty, and it would be inappropriate to base the proficiency estimates simply on the number of correct answers obtained. Therefore, large-scale assessments using matrix sampling rely on IRT models. The PIAAC IRT modeling resulted in 10 plausible values (PVs) (Mislevy 1991) for each respondent, reflecting the uncertainty in the respondents' proficiency estimate. To capture this uncertainty, we implemented a multiple imputation approach for calculating direct estimates and the associated variance estimates, using the PVs. This is described further in section 3.1.

Another issue to account for in direct estimation is informative nonresponse; informative nonresponse is the condition when nonresponse is not sufficiently explained by the weighting variables. Being able to handle informative nonresponse needs to be considered and addressed because, otherwise, the process will lead toward biased estimates. For PIAAC, this means literacy-related nonresponse needs to be addressed, which is estimated to be about 5 percent of the population. Within a county, all cases receiving a final weight contribute to the direct estimate. This includes respondents to the background questionnaire (BQ) as well as sampled persons that did not respond to the BQ for a literacy-related reason (language barrier, reading/writing barrier, or mental disability). All respondents to the BQ have literacy scores, whereas the BQ literacy-related nonrespondents do not. For the estimation of proportions, we assumed any literacy-related nonresponse is below Level 1, and for the estimation of averages, we imputed a proficiency score for each PV using the first percentile of the respondents' scores for the corresponding

PV. The first percentile was chosen because it is similar to the average score for persons who completed the BQ but could not complete the assessment for a literacy-related reason.

3.1 Direct County Estimates

The first step in the area-level modeling process was to produce direct estimates for the 185 counties with sample. Under the multiple imputation approach, we first computed the survey estimate for the *m*-th PV for county *k* as:

$$\hat{y}_{km} = \sum_{l=1}^{n_k} w_{kl} y_{klm} \Big/ \sum_{l=1}^{n_k} w_{kl} \tag{3a}$$

where w_{kl}= the final PIAAC 2012/2014/2017 national weight for person l in county k, y_{klm} is the proficiency score (for average) or an indicator variable for the proficiency level (for proportions), and n_k = the number of cases in county *k*.

Then the county-level direct estimate ($\hat{y}_k$) was calculated as:

$$\hat{y}_k = \frac{1}{10} \sum_{m=1}^{10} \hat{y}_{km} \tag{3b}$$

The multiple imputation estimate of the variance (Rubin 1987) is:

$$\hat{\sigma}_k^2 = \hat{\sigma}_{Wk}^2 + (\tfrac{11}{10})\hat{\sigma}_{Bk}^2, \tag{3c}$$

where $\hat{\sigma}_{Wk}^2$ is the within-imputation variance and $\hat{\sigma}_{Bk}^2$ is the between-imputation variance. The within-imputation variance component was computed as the average of the sampling variance for each of the 10 PVs:

$$\hat{\sigma}_{Wk}^2 = \left(\textstyle\sum_{m=1}^{10} \hat{\sigma}_{km}^2\right)/10, \tag{3d}$$

where $\hat{\sigma}_{km}^2$ is the sampling variance of the estimated mean or proportion for PV m. The between-imputation component was calculated as:

$$\hat{\sigma}_{Bk}^2 = [\textstyle\sum_{m=1}^{10} (\hat{y}_{km} - \hat{y}_k)^2]/9. \tag{3e}$$

Sampling variances were calculated using the Taylor series method (Wolter 2007, p. 234), with primary sampling units (PSUs) as strata and secondary sampling units (SSUs) as variance units (clusters), where PSUs and SSUs are defined in section 1.1. Details about the computation of sampling variances using the Taylor series method for estimated mean and proportion can be found in SAS online documentation.[9]

Direct variance estimates could not be computed for 15 counties that only had one SSU with PIAAC data. The models described in section 3.3 were used to predict the variances for these counties. The remaining 170 counties had at least two SSUs and sample sizes of five or more.

Table 3-1 shows the distribution of the proportion of variance for direct estimates attributed to imputation error, as measured through multiple imputation (i.e., between-imputation variance) across the 170 counties with at least two SSUs. For literacy skills, multiple imputation contributes on average 11 percent of total variance for the average score, 22 percent of total variance for the proportion at or below Level 1, 35 percent of total variance for the proportion at Level 2, and 20 percent of total variance for the proportion at Level 3 and above. Across counties, the contribution to the total variance from multiple imputation ranges from nearly 0 percent to 90 percent. The distribution of the proportion for numeracy is similar to that for literacy. This illustrates that imputation variance can be a significant portion of the variance and cannot be ignored when producing variance estimates.

Table 3-1. Distribution of the proportion of variance associated with multiple imputation for direct estimates across counties: 2012/2014/2017

Proficiency domain	Proportion of variance due to multiple imputation for	N[1]	Mean	Minimum	Maximum	Standard deviation
Literacy	Average score	170	0.11	0.00	0.87	0.122
Literacy	Proportion at or below Level 1	170	0.22	0.02	0.90	0.154
Literacy	Proportion at Level 2	170	0.35	0.07	0.82	0.143
Literacy	Proportion at Level 3 and above	170	0.20	0.00	0.82	0.142
Numeracy	Average score	170	0.10	0.01	0.73	0.097
Numeracy	Proportion at or below Level 1	170	0.21	0.02	0.81	0.149
Numeracy	Proportion at Level 2	170	0.39	0.07	0.81	0.145
Numeracy	Proportion at Level 3 and above	169	0.21	0.03	0.70	0.139

[1]170 of the 185 counties in PIAAC sample have at least two clusters. The other 15 counties have only 1 SSU, thus their variance cannot be estimated. For Numeracy, one of the 170 counties has no respondents with scores at Level 3 and above, thus variance is 0 for that county and proportion of variance associated with multiple imputation is missing.
SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

[9] https://support.sas.com/documentation/cdl/en/statug/63962/HTML/default/viewer.htm#statug_surveymeans_a0000000224.htm
https://support.sas.com/documentation/cdl/en/statug/63962/HTML/default/viewer.htm#statug_surveyfreq_a0000000247.htm.

Finally, to estimate the covariance for proportions, let $\hat{y}_{1k}$, $\hat{y}_{2k}$, and $\hat{y}_{3k}$ be the estimated proportions at or below Level 1, at Level 2, and at Level 3 and above, respectively, and let $\hat{\sigma}_{1k}^2$, $\hat{\sigma}_{2k}^2$, and $\hat{\sigma}_{3k}^2$ be the corresponding variance estimates. Given that $\mathrm{Var}(y_{1k} + y_{2k}) = \mathrm{Var}(y_{3k})$, the covariance of the proportion at or below Level 1 and the proportion at Level 2 can then be computed using the following formula. In Section 3.3 we will smooth covariance using this formula based on smoothed variances.

$$\mathrm{Cov}(\hat{y}_{1k}, \hat{y}_{2k}) = (\hat{\sigma}_{3k}^2 - \hat{\sigma}_{1k}^2 - \hat{\sigma}_{2k}^2)/2. \tag{3f}$$

3.2 Survey Regression Estimation (SRE)

PIAAC was designed to be a nationally representative sample and does not produce efficient direct estimates at the county level; therefore, as is shown in table 3-2, the variances of the direct estimates can be large. This is particularly true for counties with small sample sizes. Therefore, we used SRE to reduce the variance associated with the survey estimates. Rao and Molina (2015, pp. 21–23) describe the use of these estimates in SAE, their derivation, and the usual Taylor series approach to estimating their variance. The SRE is a model-assisted approach that is used to bring survey estimated county population totals in line with county totals from a reliable external source and improve the stability of the survey estimates. The SRE process also helps to reduce variances that are used in the SAE modeling process.

In the modified form proposed by Särndal and Hidiroglou (1989), the survey regression estimate of the *m*-th PV for county *k* can be written as:

$$\hat{y}_{km}^{surv} = \mathbf{X}_{\boldsymbol{k}}^{\boldsymbol{T}}\widehat{\mathbf{B}}_{\boldsymbol{m}}/N_{\boldsymbol{k}} + \sum_{\boldsymbol{l}\in\boldsymbol{s_k}} w_{\boldsymbol{l}} e_{lm} \Big/ \sum_{l\in s_k} w_l,$$

where $\mathbf{X}_k$ is the vector of population totals in county *k* corresponding to the predictors in the unit-level regression model (defined below),

$$\widehat{\mathbf{B}}_m = \left(\sum_{l\in s} w_{\boldsymbol{l}}\mathbf{x}_l\mathbf{x}_l^T\right)^{-1} \sum_{l\in s} w_l \mathbf{x}_l y_{lm}$$

is the vector of survey-weighted regression coefficients from the unit-level regression based on the whole sample for PV *m*, N_k is the known size of the eligible population in the county, the $e_{lm} = y_{lm} - \mathbf{x}_l^T\widehat{\mathbf{B}}_m$ are the unit-level residuals from the regression fit in county *k* for PV *m*, the w_l are the corresponding survey weights, and s_k is the sample in county *k*. In later sections, the survey regression estimates for proportions are denoted by P_{ijk} and the survey regression estimates for averages are denoted by Y_{ijk}, for county k in state j of census division i.

The original survey regression estimator for a mean or proportion would have been instead:

$$\hat{y}_{km}^{surv*} = \mathbf{X}_{\boldsymbol{k}}^{\boldsymbol{T}}\widehat{\mathbf{B}}_{\boldsymbol{m}}/N_{\boldsymbol{k}} + \sum_{\boldsymbol{l}\in \boldsymbol{s_k}} w_{\boldsymbol{l}} e_{lm}/N_k,$$

which divides all terms in the original survey regression estimator of a total by N_k. In other words, the modification replaced N_k by a sample-based estimate $\sum_{l\in s_k} w_l$ in the correction term, giving the estimator generally better properties conditional on the estimated population size. In general, the sample estimate of the number of eligible respondents in a county would have been variable at the PSU level and particularly so at the county level in PSUs comprising more than one county.

The predictors for the unit-level regression model were chosen based on the availability of population totals that had the same definition and coverage as the corresponding PIAAC variables. Predictors were further limited to PIAAC variables that had a low level of item nonresponse (less than 5 percent), and imputation was used to fill in the missing values. The models for the eight literacy/numeracy estimates used the same set of predictors, and the final set consisted of 15 indicator variables for the following 15 categories:

- age groups: 18–19; 20–24; 25–34; 35–44; 45–54; 55–64; and 65–74;
- gender by age: Male of age 18–74;
- race/ethnicity by age: Black of age 18–74; and Hispanic of age 18–74;
- educational attainment by age: less than high school education of age 18–64; high school of age 18–64; some college of age 18–64; and bachelor's degree of age 18–64; and
- nativity by age: foreign-born of age 20–74.

Each indicator takes a value of 1 if the person falls into the category or 0 otherwise. For example, a 16-year-old would have a value of 0 for all 15 indicators. The corresponding county-level population totals were obtained from the American Community Survey (ACS) 2012–2016.[10] The age range for each indicator differs based on the availability of data in the ACS.

[10] As discussed in chapter 4, the covariates for the SAE model were based on ACS 2013-2017 data. However, these data were not available in time for the SRE process. Therefore, the SRE model relies on the data from the previous 5-year period of 2012–2016.

The variance of the SRE for each PV was estimated using the Taylor series method by applying the standard variance expression to the residuals (e_l), with PSUs as strata, and with SSUs as segments/clusters. Let $n_{seg,p}$ be the number of segments, c, sampled in PSU p, and let $n_{seg,pk}$ be the number of segments in county k within PSU p. If PSU p included more than one county, $n_{seg,pk}$ is generally a random variable. Because the PSU is treated as a stratum, each county falls in a single stratum, simplifying the notation for the variance estimator somewhat. After exclusion of counties with only one sample segment, the sampling variance for each PV was estimated with

$$var(\hat{y}_{km}^{surv}) = N_k^{-2}\frac{n_{seg,p}}{\left(n_{seg,p}-1\right)}\sum_{c\in p}\left(e_{pkcm}-e_{pkm}\right)^2$$

$$e_{pkcm} = \sum_{l\in c} a_{pkc}w_l e_{lm}$$

$$e_{pkm} = n_{seg,p}^{-1}\sum_{c\in p} e_{pkcm}$$

where a_{pkc} is an indicator with value 1 if segment c of PSU p is in county k and 0 otherwise. Rao and Molina (2015, p. 23) suggest this estimator. It is also similar and asymptotically equivalent to one of two versions offered by Särndal, Swensson, and Wretman (1992) (see formula [6.6.11] on p. 237 and $\hat{V}^*(\hat{t}_{dr})$ on p. 402); they also derive a more complex form dependent on computation of the g-weights ([6.6.10] on page 237 and [10.5.12] on page 401).

As a remark, the variance estimators do not incorporate an explicit account for variability in the estimation of $\widehat{\mathbf{B}}_m$. An explanation is that the individual segment-level residuals are $O_p(1)$, that is, of order 1 in probability, and $\mathbf{X}_k^T/N_k$ is $O(1)$, that is, of order 1, while the error in $\widehat{\mathbf{B}}_m$, that is, $\widehat{\mathbf{B}}_m - \mathbf{B}_m$, is $O_p(n^{-\frac{1}{2}})$, and consequently this term does not contribute to the first-order Taylor expansion.

In small counties, the variability in $\mathbf{X}_k^T$ is a possible consideration. Because the ACS samples approximately 1 in 8 households over the course of 5 years, roughly 250 households would be sampled in a county with a population of 5,000. If a county that size was a sampled PSU, then the ACS sample would only be about three times as large as the PIAAC sample. But counties this small may have been grouped with others in forming PSUs and would not receive the full PSU-level sample. In addition, such counties represent less than half of 1 percent of the U.S. population, and so few if any would be sampled as PSUs by themselves. Consequently, ignoring the contribution of ACS variance on the variance of the SRE has a negligible effect on the overall analysis.

The multiple imputation formulas in section 3.1 were then applied to account for imputation variance, using $\hat{y}_{km}^{surv}$ in place of $\hat{y}_{km}$ in equations (3b) and (3e) to obtain the survey regression estimates and variances for the sampled counties. As with the direct estimates, the variance could not be estimated for the 15 single-SSU counties. The result of this step is one point estimate and one variance estimate for each of the eight outcomes.

As shown in table 3-2, the SRE made a large impact on the variance estimates. That is, the median variance decreased substantially compared to the direct estimate for averages, the proportion at or below Level 1, and the proportion at Level 3 and above. For the proportion at Level 2, there was only a modest decline. The R^2 for the Level 2 models were of the order of 0.04 compared to 0.27 for the at or below Level 1 models, 0.25 for the Level 3 and above models, and 0.40 for averages.

3.3 Variance Smoothing

The Hierarchical Bayes models chosen for the PIAAC SAE process, discussed in chapter 5, assume that the variances of the SRE county estimates are known, whereas in practice they are unknown. Since the survey regression estimates of these variances are subject to substantial sampling error, the true variances have also been predicted using a modeling approach. An important feature of the development of the model for predicting the variances is that approximate values will suffice since the values of the variances affect the estimates of the PIAAC small area estimates in only a minor way. Their main impact is in stabilizing the widths of the credible intervals. Inspired by the generalized variance function (GVF) methods in chapter 7 of Wolter (2007), variance estimation smoothing models were developed separately for proportions (discussed first) and averages (discussed second).

Table 3-2. Summary of variance estimates prior to SRE, after SRE, and after smoothing: 2012/2014/2017

Proficiency domain	Sampling variance for	Stage	Number of counties	Minimum	Median	Mean	Maximum	Standard deviation
Literacy	Average score	Direct estimate	170	16.51	100.23	127.48	663.42	111.623
		SRE	170	8.73	47.46	61.93	476.57	58.305
		smoothed	170	15.86	45.55	56.20	275.15	40.635
	Proportion at or below Level 1	Direct estimate	170	0.0007	0.0058	0.0079	0.0566	0.00771
		SRE	170	0.0006	0.0036	0.0055	0.0594	0.00648
		smoothed	170	0.0009	0.0035	0.0054	0.0427	0.00595
	Proportion at Level 2	Direct estimate	170	0.0011	0.0069	0.0094	0.0600	0.00822
		SRE	170	0.0011	0.0067	0.0090	0.0602	0.00817
		smoothed	170	0.0011	0.0071	0.0090	0.0636	0.00803
	Proportion at Level 3 and above	Direct estimate	170	0.0040	0.0160	0.0246	0.2306	0.02586
		SRE	170	0.0010	0.0051	0.0065	0.0659	0.00640
		smoothed	170	0.0010	0.0054	0.0064	0.0309	0.00451
Numeracy	Average score	Direct estimate	170	18.00	127.99	164.41	771.03	138.954
		SRE	170	8.94	60.43	78.15	737.71	87.124
		smoothed	170	19.62	54.70	67.45	346.00	46.129
	Proportion at or below Level 1	Direct estimate	170	0.0016	0.0078	0.0101	0.0712	0.00916
		SRE	170	0.0008	0.0045	0.0066	0.0594	0.00715
		smoothed	170	0.0010	0.0045	0.0063	0.0454	0.00589
	Proportion at Level 2	Direct estimate	170	0.0010	0.0068	0.0097	0.1004	0.00983
		SRE	170	0.0008	0.0065	0.0095	0.0937	0.00949
		smoothed	170	0.0010	0.0072	0.0092	0.0749	0.00827
	Proportion at Level 3 and above	Direct estimate	170	0.0042	0.0201	0.0303	0.3254	0.03256
		SRE	170	0.0006	0.0047	0.0063	0.0637	0.00726
		smoothed	170	0.0010	0.0050	0.0058	0.0300	0.00407

SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

3.3.1 Variance Estimation Smoothing Model for Proportions

After evaluating a number of variance smoothing options (see appendix B), the variance of proportions was smoothed by fitting a weighted least square model for the effective sample size, which was computed as the sample size divided by the design effect. The number of clusters in a county and average cluster size in a county were used as the model predictors. And the weight was the number of clusters in a county minus 1. Note that counties with fewer than four clusters or SRE proportions less than 0.02 were excluded from the model (24 out of 185 counties); i.e., the model was estimated based on the remaining 161 counties. After the model was fit, the smoothed variance was derived for all 185 counties.

The variance smoothing process started with the county-level survey regression estimates and variances from section 3.2, which had been combined over the PVs, and then smoothed the combined effective sample size. A separate model was fit for each of the six proportions (at or below Level 1, at Level 2, and at Level 3 and above for literacy and numeracy), where the dependent variable was $\ln(neff_k)$, the natural log of the effective sample size for county *k*.

Specifically, $neff_k = \hat{p}_k(1-\hat{p}_k)/\hat{\sigma}_k^2$, where $\hat{p}_k$ was the SRE proportion and $\hat{\sigma}_k^2$ was the SRE variance for county *k*. Then model was specified as

$$\ln(neff_k) = \beta_0 + \beta_1 \ln(C_k) + \beta_2 \ln(B_k) + \epsilon,$$

where $neff_k$ is the effective sample size for county *k*, C_k is the number of clusters in county *k*, B_k is the average cluster size for county *k*, and ϵ is an error term. The natural log was used to satisfy model assumptions based on some model diagnostics. The model was weighted by $C_k - 1$. The exponentiation of the predicted value from this model, $\widetilde{neff}_k$, was used to derive the smoothed variance as $\tilde{\sigma}_k^2 = \hat{p}_k(1-\hat{p}_k)/\widetilde{neff}_k$. Covariances were calculated using formula (3f) and the smoothed variances. In later sections, the variance-covariance matrix after smoothing will be denoted by Σ_{ijk}.

Table 3-3 shows the parameter estimates for the variance smoothing model for proportion at or below Level 1, proportion at Level 2, and proportion at Level 3 and above, respectively, for literacy and numeracy. The parameter estimate (β_1) for $\ln(C_k)$ is almost always about 1 and the parameter estimate (β_2) for $\ln(B_k)$ is a bit below 1, across all three levels for literacy and numeracy.

Table 3-3. Parameter estimates for the variance smoothing model for proportions: 2012/2014/2017

Proficiency domain	Outcome	Parameter	Estimate	Standard error	t value	p value
Literacy	Proportion at or below Level 1	β_0	0.12	0.226	0.51	0.6103
		β_1	0.98	0.072	13.52	<.0001
		β_2	0.77	0.097	7.93	<.0001
	Proportion at Level 2	β_0	-0.25	0.164	-1.54	0.1244
		β_1	0.95	0.053	18.08	<.0001
		β_2	0.84	0.071	11.96	<.0001
	Proportion at Level 3 and above	β_0	0.21	0.204	1.01	0.3133
		β_1	0.95	0.065	14.52	<.0001
		β_2	0.79	0.088	8.98	<.0001
Numeracy	Proportion at or below Level 1	β_0	0.10	0.204	0.48	0.6351
		β_1	0.99	0.065	15.2	<.0001
		β_2	0.77	0.088	8.83	<.0001
	Proportion at Level 2	β_0	-0.46	0.162	-2.87	0.0047
		β_1	0.99	0.052	19.06	<.0001
		β_2	0.90	0.070	12.87	<.0001
	Proportion at Level 3 and above	β_0	0.14	0.187	0.73	0.467
		β_1	0.86	0.060	14.38	<.0001
		β_2	0.91	0.080	11.39	<.0001

SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

3.3.2 Variance Estimation Smoothing Model for Averages

As with proportions, variance smoothing for average proficiency scores needs to take into account the design effect (DEFF) associated with weight variation and clustering within the small areas. Using the combined variances from the SRE process, the variance is smoothed by fitting a weighted least square model as below:

$$\ln(Var_{r_k}) = \beta_0 + \beta_1 \ln(C_k) + \beta_2 \ln(B_k) + \beta_3 \ln(\hat{\sigma}^2_{y_k}) + \epsilon,$$

where Var_{r_k} is the residual variance for each county k, C_k is the number of clusters in each county (SSU), B_k is the average cluster size for each county k, and $\hat{\sigma}^2_{y_k}$ is the estimated population variance of the literacy/numeracy scores among each county k. The model is weighted by $C_k - 1$. The exponentiation of the predicted value from this model would be the smoothed variance. It should be noted that counties with fewer than four clusters were excluded from the smoothing process (24 out of 185 counties), so the model is estimated based on the 161 counties. After the model was fit, the smoothed variance was derived for all 185 counties. In later sections, the smoothed variances are denoted by σ_{ijk}^2. Table 3-4 provides the parameter estimates for the smoothing process for literacy and numeracy average.

Table 3-4. Parameter estimates for the variance smoothing process for county-level variances for literacy and numeracy average: 2012/2014/2017

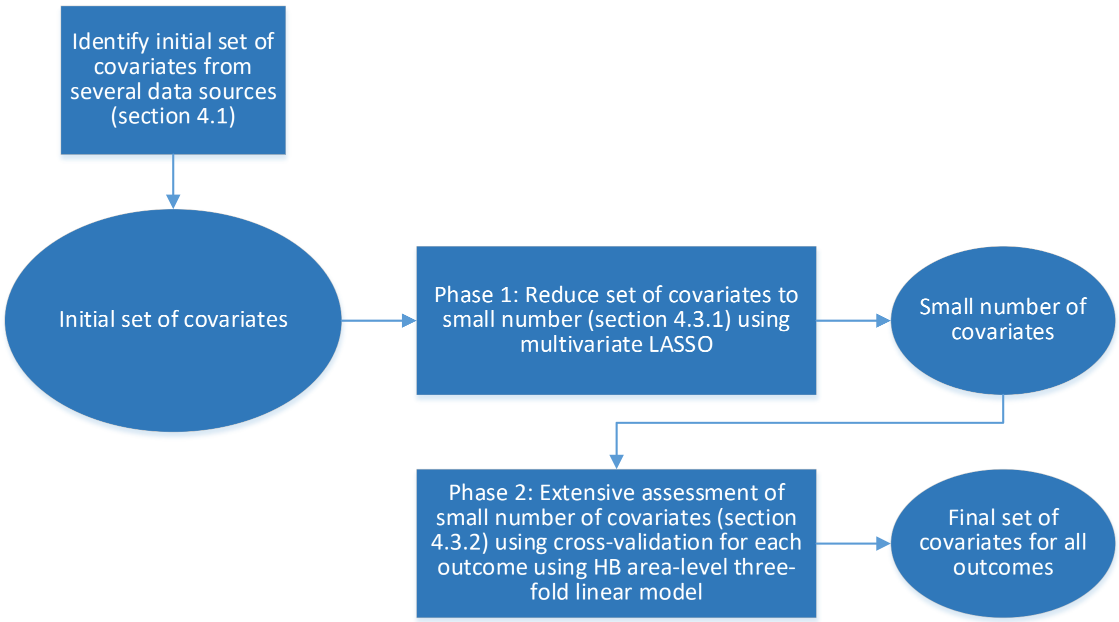
Proficiency domain	Parameter	Estimate	Standard error	*t* value	*p* value
Literacy	β_0	1.1	0.86	1.26	0.209
	β_1	-0.8	0.08	-9.82	<.0001
	β_2	-0.5	0.11	-4.67	<.0001
	β_3	0.7	0.11	6.50	<.0001
Numeracy	β_0	1.0	0.91	1.07	0.286
	β_1	-0.7	0.08	-9.07	<.0001
	β_2	-0.5	0.11	-5.03	<.0001
	β_3	0.7	0.11	6.47	<.0001

SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

4. COVARIATE SELECTION

A critical aspect of small area estimation (SAE) modeling is access to predictor variables that are measured consistently across all counties and states, and that are effective (highly correlated with proficiency) predictors of the estimated proportion of adults at different levels or averages of proficiency in literacy and numeracy. Because there exists a large number of available data sources, and it is impossible to include all the variables in the model, it is important to narrow down the variables to a smaller set, so the final model can be developed based on this initial set of covariates. Section 4.1 provides the covariates that are measured consistently across all counties and states, and are potential effective predictors of the proportion of adults at or below Level 1, at Level 2, and Level 3 and above, or average of literacy and numeracy. The county- and state-level sources from which the potential covariates are selected are shown in section 4.2. Section 4.3 describes the covariate selection process, including two phases (as shown in figure 4-1): phase 1 feeds all state- and county-level covariates identified in section 4.1 into the Least Absolute Shrinkage and Selection Operator (LASSO) model, which selects a small number of covariates that are highly correlated with the outcomes and may be highly predictive in the SAE models, and phase 2, based on the results of phase 1, uses a k-fold cross-validation process to select the set of covariates to be used in the final SAE models. The same set of covariates will be selected and used in all four models for literacy and numeracy proportions and averages. For example, if a covariate was not selected by LASSO as a significant predictor for literacy proportions, we may still want to include it in the final SAE model for literacy proportions if it was selected by LASSO as a significant predictor for averages or numeracy proportions. The inputs to the covariate selection process, and the models, are the county-level survey regression estimates, and, where appropriate, the smoothed variances.

Figure 4-1. Covariate selection process diagram



SOURCE: Author's definition.

4.1 Initial Identification of County and State Covariates

Reliable data sources and variables that are potential covariates of proficiency levels were initially identified, and more than 70 county-level variables across five major variable types were obtained as potential predictors from eight data sources (see details in section 4.2). The major county-level variable types include variables related to demographic characteristics (i.e., race/ethnicity, age, gender, marital status), socioeconomic status (i.e., poverty, income, employment status, occupation), education (i.e., education, English-speaking ability), location (i.e., urbanicity, census division), immigration status (i.e., length of stay for foreign-born people, migration), and other (i.e., journey to work, housing unit tenure/phone service, plumbing facilities, health, tax). In addition, the primary sampling unit (PSU) selection probability was also initially included as a potential county-level covariate to account for the informative sampling design; however, it was not identified as a significant predictor through the covariate selection process described in section 4.3.

In addition to county-level variables, a set of state-level variables was selected to provide additional information, including 24 potential state-level predictors across different variable types from several major data sources (see details in section 4.2). The major state-level variable types included socioeconomic status (i.e., average annual pay, homeownership rate); education (i.e., school enrollment rate, graduation rate, test pass rate, reading/math composite scores); and other area characteristics (i.e., birth rate, fertility rate, infant mortality rate, crime rate, physician availability, federal aid, energy consumption). A listing of all county- and state-level variables considered for modeling is given in appendix A. The listing is sorted by major variable type, and provides details about the source, year, and level (county level or state level) of each variable.

These variable types were chosen because they were found to be related to the adult literacy skills in previous studies (Rampey et al. 2016; Goodman et al. 2013; Kirsch et al. 2002; Greenberg et al. 2001; Coley 1996; Weiss, Hart, and Pust 1991), and were available for all the counties in our sample. To ensure values of variables are most relevant to the Program for the International Assessment of Adult Competencies (PIAAC) study, we obtained variables collected within the time frame of the PIAAC study. If the variable value was based on a single year of data, we used values from 2015 whenever possible (2015 is the middle time point of the PIAAC study), and if not, the most recent data were used. If the variable was from multiple years, we used years closest to the PIAAC study years (i.e., 2013–2017). Sometimes the same variables were selected from multiple data sources. For example, there were two county-level median household income variables selected, one from the American Community Survey (ACS) dataset, the other from the U.S. Census Bureau's Small Area Income and Poverty Estimates (SAIPE) dataset.[11] Different datasets usually have different sample designs and could incur various sampling and nonsampling error (Vaish 2017). Therefore, we gathered variables from multiple sources and attempted to find the most suitable variables to fit the SAE model.

4.2 Initial Set of County and State Covariates Sources

The selected data sources have reliable data publicly available for all counties (or all states) and usually publish the updated data regularly (i.e., annually). In terms of using Big Data as possible covariates, Marchetti et al. (2015) state that Big Data derived from the digital crumbs that humans leave behind in their daily activities are providing more accurate proxies of social life, and using Big Data together with SAE techniques can provide more accurate estimation. In fact, researchers started to utilize mobile usage data to predict illiteracy in developing countries (Sundsoy 2016), or use the number of bookstores, newspaper circulation, library system, or periodical publishing resources to predict literacy levels of the U.S. cities (Miller 2016). However, we did not identify any clear Big Data source (such as social media and interactive platforms, or data publicly available on the Web) that was useful and added value beyond the data sources that had been already considered.

[11] Available at: https://www.census.gov/programs-surveys/saipe/technical-documentation/methodology/counties-states/county-level.html.

The following subsections provide brief descriptions of the data sources and the variables chosen from each source. We begin with sources for county-level variables. More details about the variables are given in appendix A.

4.2.1 Initial Set of Sources for County-Level Covariates

Census Bureau's ACS. The Census Bureau's ACS is an ongoing survey that provides up-to-date estimates for a wide range of topics including socioeconomic, demographic, and housing characteristics of the U.S. population. The 5-year estimates (2013–2017) represent data collected over 5 years for all geographies down to the block-group level (over 578,000 geographic areas). The PIAAC variable pool from ACS includes number of families in poverty, median household income, population sizes with different education levels, population sizes with English-speaking ability, population in rural/urban areas, race/ethnicity, length of stay for foreign-born people, age categories, gender, employment status, occupation, census division, housing unit tenure and phone service, plumbing facilities, marital status, and migration status.

Census Bureau's SAIPE Program. The Census Bureau, with support from other federal agencies, has created the SAIPE program to provide current small area estimates of selected income and poverty statistics. The PIAAC variable pool from SAIPE includes proportion of families in poverty and median household income.

Bureau of Economic Analysis (BEA). The BEA prepares estimates of personal income for local areas (counties, metropolitan areas, and the BEA economic areas). The PIAAC variable pool from BEA includes personal income.

U.S. Department of Agriculture (USDA). The USDA Economic Research Service provides codes that classify each county according to metro and nonmetro classifications. The 2013 Rural-Urban Continuum Codes form a classification scheme that distinguishes metropolitan counties by the population size of their metro area, and nonmetropolitan counties by degree of urbanization and adjacency to a metro area. The official Office of Management and Budget (OMB) metro and nonmetro categories have been subdivided into three metro and six nonmetro categories. Each U.S. county is assigned one of the nine codes. The PIAAC variable pool from USDA includes proportions of metro/nonmetro counties.

Bureau of Labor Statistics (BLS). The Local Area Unemployment Statistics (LAUS) program produces monthly and annual employment, unemployment, and labor force data for census regions and divisions, states, counties, metropolitan areas, and some cities, by place of residence. The PIAAC variable pool from BLS includes the unemployment rate.

Centers for Disease Control and Prevention's (CDC) Division of Diabetes Translation (DDT). The CDC collects and provides updated statistics about diabetes in the United States through the U.S. diabetes surveillance system. The PIAAC variable pool from DDT includes proportions of diagnosed diabetes and obesity.

Centers for Medicare & Medicaid Services (CMS). The CMS developed a geographic variation public use file about the utilization and quality of health care services for the Medicare fee-for-service population. The PIAAC variable pool from CMS includes the proportion of population eligible for Medicaid.

The Statistics of Income Data (SOI). SOI bases its county income data on the addresses reported on the individual income tax returns filed with the Internal Revenue Service. The PIAAC variable pool from SOI includes the number of tax returns, returns with unemployment compensation, and returns with taxable Social Security benefits, as well as adjusted gross personal income, personal unemployment compensation amount, and personal taxable Social Security benefit amount.

4.2.2 Initial Set of Sources for State-Level Covariates

In addition to county-level variables, a set of state-level variables is also selected to provide additional information, not covered by county-level variables. Several variable sources were considered at the state level, as described below.

Bureau of Labor Statistics (BLS). Besides the BLS LAUS program mentioned above, state-level data were considered from the Current Employment Statistics Program, which surveys more than 160,000 businesses and government agencies each month. The Employment and Wages annual averages were also included in the selection process. The PIAAC variable pool from the BLS includes average personal annual income.

Adult Education Data. The Office of Career, Technical, and Adult Education (OCTAE) collects data on adult education program enrollments from each state. Data for 2014–2015 from the National Reporting System (NRS) for Adult Education and Literacy was considered for the small area models. The PIAAC variable pool from the OCTAE are adult basic/secondary education enrollment and English as a Second Language enrollment.

The Integrated Postsecondary Education Data System (IPEDS). This National Center for Education Statistics (NCES) program collects data through a system of surveys from primary providers of postsecondary education. The PIAAC variable pool from the IPEDS includes the graduation rate, instructor salary, average financial aid, and annual college cost.

National Assessment of Educational Progress (NAEP). The NAEP survey is the largest nationally representative and continuing assessment of what our nation's students know and can do in various subject areas. Assessments are conducted periodically in mathematics, reading, science, writing, the arts, civics, economics, geography, U.S. history, and technology and engineering literacy based on representative samples of students at grades 4, 8, and 12 for the main assessments. The PIAAC variable pool from NAEP includes average 4th- and 8th-grade reading/mathematics composite scale scores, while grade 12 data are not available at the state level.

Other Census Bureau Programs. Besides the ACS, other state demographic data from the Census Bureau were collected from Population Estimates and from data on housing vacancies and home ownership from the Housing Vacancy Survey.

Other Sources. State-level data from other sources were obtained, including National Highway Safety Traffic Administration's *Traffic Safety Facts*, National Center for Health Statistics' *Vital Statistics of the United States*, the American Medical Association's *Physician Characteristics and Distribution in the U.S.*, the Federal Bureau of Investigation's *Crime in the United States*, the Energy Information Administration's *State Energy Data Report*, the GED Testing Service's *Annual Statistical Report on the GED Test*, and the *National Vital Statistics Reports.*

4.3 Covariates Selection Process

A key step in model development involves selecting a smaller set of covariates from a large set of potential covariates. As mentioned above, for PIAAC SAE, more than 70 variables in the county-level and more than 20 variables in the state-level variable pools were identified as potential covariates. Each variable was examined against the outcome to identify if transformation was needed, and found that log-transformed income variables (i.e., median household income) have a more linear relationship with the outcome; thus, the three income variables were log transformed.

The process of selecting covariates was conducted in two phases. In the first phase, all the county- and state-level variables were considered as fixed effects and the number of variables was reduced through the variable selection methods. The implemented variable selection method (1) used a correlation matrix among the covariates themselves to identify highly correlated variables that led to dropping one variable in the highly correlated pair to avoid multicollinearity; (2) used the LASSO (Tibshirani 1996) method to select several sets of covariates for each of the four survey regression estimation (SRE) outcome models. The purpose of the screening (first step) based on the correlation was to avoid multicollinearity. The pairs of covariates with high correlation among themselves were examined, and it was decided whether to drop any of the covariates to avoid multicollinearity. In this case, the decision to drop covariates was based on whether they were correlated with other covariates, how many other covariates, and how high the correlation was. We also ran the LASSO regression using all the different sets of covariates (1. Include all the covariates; 2. Exclude those with 0.9 or higher correlation; 3. Exclude those with 0.8 or higher correlation; 4. Exclude those with 0.7 or higher correlation) and different lambda values, which resulted in a final set of covariates that was selected for all the models. Once the list of covariates was reduced to several sets of covariates, the second phase evaluated the different sets of covariates using a cross-validation process (as recommended by the U.S. PIAAC International SAE Experts) adding the random effect estimation to arrive at the final list of covariates. This final list of covariates was used in modeling all the different SRE outcomes (i.e., literacy/numeracy proportion/average models).

4.3.1 Phase 1—Covariates Reduction

This section describes the method used for the covariates reduction process with two steps: (1) a correlation matrix among the potential covariates themselves, as well as the covariates against the SRE outcomes; (2) a LASSO method (i.e., multivariate LASSO) using the county- and state-level covariates to predict the SRE outcomes. Details are provided below.

Step 1. The correlation estimation is a natural first step of the overall process. It was carried out by computing the Pearson-correlation matrix between each pair (n=3,142 counties in the United States) of the potential county-level covariates, each pair (n=51, including 50 states and the District of Columbia) of the potential state-level covariates, as well as the correlation (n=185 counties with PIAAC sample) between the covariates and each of the eight outcomes (SRE proportion at or below Level 1, proportion at Level 2, proportion at or above Level 3, and average proficiency score for both literacy and numeracy) through SAS. Covariates with high correlations with the SRE outcomes turned out to be the education-related variables (i.e., $|\rho|$=0.7 for proportion of population with lower than high school education vs. proportion at or below Level 1 literacy), poverty-related variables (i.e., $|\rho|$=0.6 for proportion of population lower than poverty threshold versus proportion at or below Level 1 literacy), employment-related variables (i.e., |ρ|=0.6 for proportion of population not in labor force vs. proportion at or below Level 1 literacy), and health-related variables (i.e., |ρ|=0.5 for proportion of population having no health insurance vs. proportion at or below Level 1 literacy). Covariates with high pairwise correlations (i.e., |ρ|>0.7) with other covariates were treated as with "high multicollinearity." To avoid the potential effect of high multicollinearity on the model performance and prediction, one covariate from each pair of covariates with high multicollinearity was manually eliminated from the potential covariates pool, based on the correlations with the other covariates. In general, covariates with higher correlations with other covariates, or with lower correlations with the SRE outcomes, were generally eliminated first. In the cases where two highly correlated covariates were correlated by definition, and were found to have key impact on the SRE outcomes (i.e., proportion of population less than high school, proportion of population more than high school), both variables were kept for the following covariate reduction process.

It should be noted that direct estimates are subject to sampling error, and the sample sizes that contribute to each county direct estimate vary widely; therefore, the correlation coefficients presented in this section are biased and attenuated. As pointed out in Lahiri and Suntornchost (2015), the true population correlations are higher. The correlation estimates could be improved if the sampling error is taken into account, as described in Lahiri and Suntornchost (2015).

Step 2. LASSO estimation was carried out in R using the *glmnet* package (Friedman, Hastie, and Tibshirani 2010) to conduct the covariate selection. LASSO (Tibshirani 1996) is a method that applies shrinkage factors to regression coefficients, and thus can more efficiently perform stable covariate selection. The procedure can select a few covariates that are related to the dependent variable from a large amount of possible covariates. LASSO-based methods use "penalized regression" models that impose constraints on the estimated coefficients that tend to shrink the magnitude of the regression coefficients, often eliminating the covariates entirely by shrinking their coefficients to zero. Therefore, nonzero coefficients are estimated for true covariates, whereas the coefficients for irrelevant variables are zeroed out.

The final covariate reduction process was based on applying the LASSO model with standardized covariates and LASSO penalty. To select covariates for the proportion estimates (proportion at or below Level 1, proportion at Level 2, proportion at or above Level 3), we used multivariate LASSO with the option "family = "mgaussian" in *glmnet* to predict both SRE outcomes (proportion at or below Level 1, proportion at or above Level 3) at the same time. The second SRE outcome was not included as it could be derived from the other two SRE outcomes. This multitask learning method is useful when there are a number of correlated responses. When a variable is selected, a coefficient will be fit for each response, and a 'group-lasso' penalty was applied on the coefficients for each response. It will result in the selection of the same predictors across all responses. For average proficiency scores, we used univariate LASSO to do the variable selection.

Because the proportion at or below Level 1 and proportion at or above Level 3 were modeled together in the same model (see section 5.2.1), there were a total of four SRE outcome models being analyzed: literacy/numeracy proportion model, literacy/numeracy average model). For each of the four SRE outcome models, we adjusted the lambda (the parameter that controls the overall strength of the penalty) using various values close in magnitude to the lambda that minimizes the mean cross-validated error (0.02 and 0.03 for the proportion model, and 2 and 3 for the average model) to arrive at two sets of covariates for each of the four models. We expected each of these sets contained less or equal to 10 covariates and there were some variation between the sets. The lambda adjustment and coefficients of the predictors were requested through the *cv.glmnet* function, from which two lambda values were applied and two lists of selected covariates were generated, one is more parsimonious and with fewer covariates and the other is more generous and with more covariates.

Covariates with nonzero coefficients from the LASSO model were considered as potential covariates for the phase 1 lists of covariates to be included in the phase 2 cross validation process. Table 4-1 presents the list of selected phase 1 covariates, with the source, year, description and label. Table 4-2 presents the selected covariates with the marker "X" identifying the selected variables for each SRE outcome models (with two lambda options). It should be noted that for the numeracy proportion model, both lambda options (0.02 & 0.03) resulted in the same set of selected covariates.

Table 4-1. List of phase 1 select covariates, including their label, source, and year

Source	Year	Description	Label
American Community Survey	2013–2017	Percentage of population age 25 and over with less than high school education (no high school diploma)	Education—LH
		Percentage of population age 25 and over with more than high school education (including some college, no degree)	Education—MH
		Percentage of population below 100 percent of the poverty line	Poverty
		Percentage of Black or African American population	Black
		Percentage of Hispanic population	Hispanic
		Percentage of civilian noninstitutionalized population who has no health insurance coverage	Health insurance
		Percentage of population age 16 and over with service occupations	Service occupations
		Percentage of foreign-born people who entered the United States after year 2010 among the population born outside the United States	Enter U.S. 2010
		Percentage of population born outside of the United States	Foreign born
		Percentage population 16 and over who did not work at home who spend more than 60 minutes to travel to work	Journey to work
Bureau of Labor Statistics	2015	Unemployment rate	Unemployment rate
Division of Diabetes Translation	2013	Percentage of diabetes diagnosed	Diabetes rate
National Vital Statistics Reports	2015	Birth rate per 1000 women	Birth rate
The Integrated Postsecondary Education Data System	2014–2015	Average amount of grant and scholarship aid received	Grant/scholarship received

SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

Table 4-2. Predictor variables selected in phase 1, by the outcome model and LASSO lambda option

	Literacy				Numeracy			
	Proportion model		Average model		Proportion model		Average model	
Variable	λ=0.02	λ=0.03	λ=2	λ=3	λ=0.02	λ=0.03	λ=2	λ=3
Education—LH	X	X	X	X	X	X	X	X
Education—MH	X	X	X	X	X	X	X	X
Poverty	X	X	X	X	X	X	X	X
Black	X		X		X	X	X	X
Hispanic							X	
Health insurance	X		X	X	X	X	X	X
Service occupations			X	X			X	X
Enter U.S. 2010	X		X					
Foreign born	X							
Journey to work			X					
Unemployment rate			X				X	X
Diabetes rate					X	X		
Birth rate	X							
Grant/scholarship received	X		X				X	

SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

Tables A-3 and A-4 in appendix A provide the listings of the county- and state-level selected variables with the correlation estimates, random forest importance score, and LASSO estimates, sorted descending by the correlations, respectively, for each SRE outcome.

4.3.2 Phase 2—Cross Validation

The SAE models (discussed in chapter 5) were used to make predictions for the nonsampled counties (the counties that have no PIAAC sample or have too few sampled cases to be usable). Once the set of covariates was reduced to a manageable number, the cross-validation approach was applied to validate the model using the area-level HB linear three-fold model, as well as determine the best set of predictors through trying different combinations, for all eight outcomes. A cross-validation analysis evaluates the prediction power of the model as compared to other models using alternative sets of covariates selected from the LASSO models through k-fold cross validation (Fushiki 2011).

The k-fold cross validation was implemented in the following steps to select the best set of covariates for the bivariate model of literacy proportions, which models two proportions (P1, proportion at or below Level 1, and P3, proportion at or above Level 3) jointly.

- Sorted the 184[12] sampled counties from the largest to the smallest sample sizes and divided them into groups of 10 counties, with the last group having only 4 counties. There were 19 groups in total.

- For each group of 10 counties, the counties were randomly assigned to 10 subsets, with each subset containing 1 county from the group. For the group with 4 counties, the counties were randomly assigned to four subsets. At the end of step 2, each subset contained 18 or 19 counties with varying sample sizes.

- Excluding the counties in subset 1, the counties in the remaining 9 subsets were used to fit the bivariate SAE model (the model specification is discussed in chapter 5) for each given set of covariates and made predictions for the group of counties that were deleted (the prediction method is discussed in chapter 5).

- Repeated step 3 by excluding subsets 2 through 10, one at a time. At the end of this process, the predicted proportions at or below Level 1, at Level 2, and at or above Level 3 were calculated for all the counties.

- Compared the predicted proportions against the direct estimates for all 184 counties and only the counties with large sample sizes (sample size greater than 100). Calculated the sum of squared differences.

The smaller the sum of squared differences, the better the set of covariates predicted the proportions for the counties that were excluded from modeling. For literacy proportions, five sets of covariates (all county level) were used to fit the models and to compare the predicted proportions against the direct estimates. The results are summarized in table 4-3.

[12] One county was excluded from modeling because it has negative SRE estimate for literacy proportion at or below Level 1.

Table 4-3. Covariates used in cross validation for literacy proportions and results of summed squared differences between predicted proportions and direct estimates: 2012/2014/2017

	Scenarios				
Variable	1	2	3	4	5
Education—LH	X	X	X	X	X
Education—MH	X	X	X	X	X
Poverty	X	X	X	X	X
Black		X	X	X	X
Enter U.S. 2010		X	X		
Health insurance		X		X	X
Birth rate		X			
Grant/scholarship received		X			
Foreign born		X			
Hispanic			X	X	X
Service occupations					X
Sum of squared differences between predicted proportions and direct estimates over 44 counties with sample size at least 100					
P1	0.109	0.078	0.081	0.076	0.076
P2	0.136	0.137	0.144	0.141	0.143
P3	0.212	0.155	0.186	0.170	0.183

SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

For the cross validation analysis, scenarios 1 and 2 were chosen from the LASSO models with λ=0.03 and λ=0.02, respectively. Scenario 3 used the five predictors adopted by the Hierarchical Bayes model in the National Assessment of Adult Literacy (NAAL) study to predict the proportion of adults lacking basic prose literacy skills, and added the percentage of Hispanic as a predictor, which is highly correlated with proportion at or below Level 1. Compared to scenario 3, scenario 4 added another predictor, proportion of people with no health insurance coverage, which was shown to be significant in the LASSO models for predicting proportions and averages for both literacy and numeracy. Scenario 5 added an extra predictor, proportion in service occupation, to the set of covariates used on scenario 4 because this variable was shown to be a significant predictor in the LASSO models for predicting averages for both literacy and numeracy.

The results in table 4-3 show that scenarios 2, 4, and 5 have similar performance and their sum of squared differences between model predictions and direct estimates are smaller for all three proportions than those from scenarios 1 and 3. Therefore, the variables in table 4-4 were considered as the base variables for the other three models for literacy averages, numeracy averages, and numeracy proportions. The goal was to include the same terms in each model to help retain the associations between the resulting eight outcomes. That being said, the covariate/cross-validation selection process looked at various alternative models versus keeping the same variables in each. The results in tables C-1 through C-3 show that there were no alternative models that would justify a different set of covariates. Combining these results with the other cross validation results for literacy average and numeracy proportions and average (the results can be found in appendix C), a decision was made to use the seven county-level covariates from the 2013–2017 ACS data, as shown in table 4-4, in all four models fitted for proportions and averages for literacy and numeracy. Table 4-5 shows the correlation coefficients among these covariates. The seven covariates are highly correlated with the proportions and averages. For example, the adjusted R-square is 0.58 for the linear regression of literacy proportions at or below Level 1 on the seven covariates.

Table 4-4. List of covariates for the final small area models

Covariates	Label
Percentage of population age 25 and over with less than high school education	Education—LH
Percentage of population age 25 and over with more than high school education	Education—MH
Percentage of population below 100 percent of the poverty line	Poverty
Percentage of Black or African American population	Black
Percentage of Hispanic population	Hispanic
Percentage of civilian noninstitutionalized population who has no health insurance coverage	Health insurance
Percentage of population age 16 and over with service occupations	Service occupations

SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

Table 4-5. Correlation coefficients among covariates for the final small area model: 2012/2014/2017

Variable	Education—MH	Poverty	Black	Hispanic	Health insurance	Service occupation
Education—LH	-0.76	0.64	0.34	0.42	0.58	0.21
Education—MH		-0.53	-0.20	-0.04	-0.38	-0.13
Poverty			0.47	0.08	0.47	0.37
Black				-0.11	0.19	0.15
Hispanic					0.40	0.15
Health insurance						0.19

SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

5. MODEL DEVELOPMENT

This chapter highlights some important aspects of critical items that were considered in the development of Program for the International Assessment of Adult Competencies (PIAAC) small area estimation (SAE) models.

- **SAE Models Incorporating Informative Sampling.** An important point made during the U.S. PIAAC Summit of International SAE Experts was that models need to account for the variance impact from complex samples, and account for differential weighting in direct estimates. The sample of states and counties are not simple random samples; therefore, the sample design is informative. One approach to addressing informative sampling is to include the probability of selection of the primary sampling units (PSUs) in the model. However, the probability of selection of PSUs was included in the covariate selection process (described in chapter 4) but was not an important factor to include in the small area model. Also, weighting adjustments for nonresponse can reduce bias to the extent that the weighting variables are related to the proficiency scores. Being able to handle informative sampling needs to be considered and addressed because, otherwise, the process will lead toward biased estimates.

- **Accounting for Sources of Error.** Another key discussion point made during the U.S. PIAAC Summit of International SAE Experts was that models need to account for all important sources of variability so that the reported estimated error reflects the true level of precision. For PIAAC, these sources of error include the following:

  - Sampling error results from probability sampling and that different results would occur for repeated samples.

  - Model error results from estimation of model parameters, such as area-level random effects. This type of error accounts for different results occurring for different runs of the modeling process due to its random mechanism in fitting the models. Hierarchical Bayes (HB) methods typically account for the noise contributions attributed to estimating model parameters (beta coefficients, and random effects variance-covariance parameters).

  - Measurement error occurs when using covariates that are subject to sampling or nonsampling error. The model can be as strong as its covariates' precision levels, and its strength of association with the small area estimate of interest. The covariates may have inaccurate measurements with possible systematic bias. Some models in the literature may rely on estimates in low levels of geography (e.g., full cross-tabulation of county-level variables from the American Community Survey [ACS], or tract-level data); however, these estimates are subject to substantial sampling error or measurement error. In order to allow for the propagation of the measurement error into the small area estimates, an approach such as introduced by Ybarra and Lohr (2008) could be incorporated; however, it may not be able to extend to this many variables. Therefore, we have not accounted for this source of measurement error. Related to the sampling error in county-level proportions from the ACS 5-year estimates, Bell et al. (2019) conclude that in the presence of measurement error (covariate sampling error) the model is misspecified if not accounted for, so it will misstate the prediction mean square error (MSE) except for counties with sampling error for the covariates equal to the average sampling error across the counties. In general, the resulting prediction error is an overestimate when the county's sampling error is below average, and more severely underestimates when above average. Because small counties could have less precise ACS data, we have notes on the state and county indirect estimates website to caution users that the covariates used in the model or predictions for counties with population less than 1,500 (2 percent of counties) may have higher associated uncertainty.

- Prediction error results from making estimates from the final model for areas without sample cases.
- Imputation error results from the generation of plausible values (PVs) and that different results would occur for replications of the imputation process. The PVs themselves come from a model, and that uncertainty has been accounted for in the SAE process.

This section begins with a brief description of a simulation study designed to evaluate a selected subset of various models in section 5.1, and continues with technical details on the final SAE models in section 5.2. A summary of the model fitting process is provided in section 5.3, including a discussion of software considered and model estimation for proportions and average scores. Section 5.4 includes a description of how the predictions are computed, and section 5.5 provides a discussion of the precision measures. Finally, section 5.6 addresses the user's interest in simultaneous inference. In summary, the final models produced indirect estimates for eight outcomes, at or below Level 1, in Level 2, at or above Level 3, and averages for both literacy and numeracy. In addition, credible intervals were produced for each indirect estimate for states and counties.

5.1 Summary of Simulation Results

A simulation study (presented in appendix B) was conducted to evaluate the performance of possible small area estimators under conditions similar to those faced by the PIAAC application. By regarding a large sample from ACS as a true population and then drawing subsamples from it, the predictions from a small area estimator based on the subsamples were compared to the actual values in the full sample. A strength of the simulation approach is that it tests the small area estimators against actual data, rather than against artificial data created with the small area estimator in mind.

The simulation study evaluated the HB approach implemented for National Assessment of Adult Literacy (NAAL) as well as several variants of it. One of the findings is that using survey regression estimators (SRE) and smoothed estimates of the variance can improve the performance of the small area estimates.

As a result, both SRE and variance smoothing were applied to the PIAAC county-level direct estimates before they were used in the SAE models. The comparisons of different SAE models showed some advantage to the bivariate models in predicting proportions over univariate models. The bivariate models are for two proportions (proportion of low education and proportion of high education) jointly whereas the univariate models only one proportion at a time, independently. It is then likely that the bivariate models provide more closely coordinated estimates of proportions concurrently than if they each were modeled in a univariate model. Moreover, the bivariate models support interval estimates for the sum of two proportions, equivalent in the PIAAC application to obtaining interval estimates for Level 2 (P2 = 1-P1-P3), where P1 represents the proportion at or below Level 1 and P3 represents the proportion Level 3 and above. Another finding is incorporating census division as a random effect together with state- and county-level random effects in the model may improve the results. Two software products, STAN and JAGS, were used in the simulation and produced essentially equivalent results.

5.2 Final SAE Models

The final SAE models for proportions and averages for literacy and numeracy are described in subsections 5.2.1 and 5.2.2, respectively.

5.2.1 Area-Level Bivariate HB Linear Three-Fold Model for Proportions

The progression of the previous research and simulation study has led to developing an area level bivariate HB linear three-fold model for estimating PIAAC proportions of literacy and numeracy proficiency at the county level.

- Area-level—the model is fitted at the area level with the areas defined by counties, the input data being the sets of survey regression estimates and their associated variance estimates (smoothed), available at the county level.
- Bivariate—two proportions (Level 1 and below and Level 3 and above) are modeled jointly whereas the third proportion (Level 2) is derived by subtracting the proportions of Level 1 and below and Level 3 and above from 1.
- Linear—the linking model assumes linear relationship between the proportions and the predictors.
- Three-fold—the model accounts for random effects at three nested levels defined by the county, census division, and state.

- HB model—the model is written using a hierarchical form (a sampling level for the direct estimates of proportions and a linking level for the relationship between the target proportions and the covariates), prior distributions are adopted for the model parameters, and the Bayesian approach is used for inference. Various sources of error are accounted for using the hierarchical model specification: smoothed sampling variances, and random effects at county, division and state levels. The use of Bayes methods over frequentist methods provides a straightforward framework for constructing summaries for the indirect estimates, including credible intervals, and for functions of the model parameters, such as the Level 2 proportion (defined as 1 minus the sum of the proportions at or below Level 1 and at or above Level 3). The use of HB methods over empirical Bayes methods provides a framework in which prior distributions for all the model parameters are adopted without using information from the observed data.

For each Markov Chain Monte Carlo (MCMC) sample, the results can be combined to estimate the proportion at Level 2, and provide credible intervals for all point estimates. The model takes advantage of the covariance between the two set of proportions, i.e., proportions at or below Level 1 and proportions at Level 3 and above, which may result in reduced MSE of model estimates. Another feature of the model is the inclusion of three levels of random effects: county, state, and census division. The benefits of the three-fold model are that (1) benchmarking the estimates may not be necessary as estimates are controlled through the random effects (a consensus among the U.S. PIAAC International SAE Experts), (2) estimates for states without sample will not be synthetic because all census divisions have PIAAC sample, and (3) associations of counties within states, and states within divisions will have some impact because the same random effect is applied to those areas.

The model employs the traditional SAE structure, including a sampling model and a linking model, using matrix form notation to account for multiple domains, as follows:

$$\begin{aligned} P_{ijk} &\sim N(\theta_{ijk}, \Sigma_{ijk}) \\ \theta_{ijk} &\sim X'_{ijk}\beta + c_{ijk} + v_{ij} + d_i, \end{aligned}$$

where i is an index for the division, j is an index for the state, k is an index for the county, P_{ijk} is a jointly normally[13] distributed bivariate vector of survey regression estimates for proportions at or below Level 1 and at or above Level 3, with associated estimated variance-covariance matrix Σ_{ijk}, X'_{ijk} is a vector of covariates, β is a matrix of coefficients, and c_{ijk}, v_{ij}, d_i are county-level, state-level, and division-level random effects, respectively.[14] The estimated variance-covariance matrices Σ_{ijk} are the result of smoothing functions of the variances for the survey regression estimates, and treated as fixed and known

[13] Although the proportions are strictly between 0 and 1, their distributions can be approximated by normal distributions since Σ_c is small and proportions are rarely close to 0 and 1. In case the predicted proportions from the models are outside the interval of [0,1], they are truncated at 0 and 1.

[14] The county estimates are assumed to be independent of each other. For PIAAC, sampling occurred independently within PSUs. Among the 184 counties used in the modeling, 116 of them are from PSUs constructed from single PSUs, while the other 68 counties fall into 31 PSUs. The correlations between counties in the same PSUs are small and therefore ignored in all the SAE models in this section.

in the HB model. Independent priors are assumed for the regression coefficients and the random effects. Specifically, it is assumed $\beta \sim N(0,100)$, where the normal distribution specification uses the mean of 0 and the standard deviation of 10. It is also assumed that the random effects are mutually independent, following bivariate normal distributions,

$$\begin{aligned} c_{ijk} &\sim N(0, \Sigma_c) \\ v_{ij} &\sim N(0, \Sigma_v) \\ d_i &\sim N(0, \Sigma_d) \end{aligned}$$

where Σ_c, Σ_v, and Σ_d are 2 by 2 variance-covariance matrices. For example, in $\Sigma_c = \begin{pmatrix} \Sigma_{c,1,1} & \Sigma_{c,1,2} \\ \Sigma_{c,2,1} & \Sigma_{c,2,2} \end{pmatrix}$, the elements on the diagnol $\Sigma_{c,1,1}$ and $\Sigma_{c,2,2}$ are the variances of the random effects for P1 and P3, respectively, at the county level, and the elements off the diagonal $\Sigma_{c,2,1}$ and $\Sigma_{c,1,2}$ are identical and denote the covariance of the random effects for P1 and P3. The variance-covariance matrices Σ_c, Σ_v, and Σ_d can be decomposed as follows:

$$\begin{aligned} \Sigma_c &= S_c \Omega_c S_c \\ \Sigma_d &= S_d \Omega_d S_d \\ \Sigma_v &= S_v \Omega_v S_v \end{aligned}$$

where S_c, S_d, and S_v are diagonal matrices with standard deviations along the diagonal, and Ω_c, Ω_d, and Ω_v are correlation matrices (with diagonal entries being equal to 1). In the model for literacy proportions, the prior distribution adopted for the standard deviation parameters (diagonal entries in S_c, S_d, and S_v) is Cauchy, with a location (median) hyper-parameter of 0 and a scale (half the interquartile range) hyper-parameter of 5 ; the support of the distribution was restricted to the positive real line. An $LKJ_{corr}(1)$ prior (Lewandowski, Kurowicka, and Joe 2009) is adopted as the prior distribution for the correlation matrices Ω_c, Ω_d, and Ω_v. In the model for numeracy proportions, $LKJ_{corr\_cholesky}(1)$ is adopted as the prior distribution for the Cholesky factors of correlation (lower triangular) matrices L_c, L_d, and L_v, where $\Omega_c = L_c L_c^T, \Omega_d = L_d L_d^T$, and $\Omega_v = L_v L_v^T$, respectively. The idea behind the LKJ prior is based on the decomposition of the variance-covariance matrix, with priors adopted for its components. Some details are provided in STAN's guide and reference manual, available at http://blackwell.math.yorku.ca/MATH6635/2016/stan-reference-2.12.0.pdf (STAN is the software used for PIAAC); see, for example, page 63 and page 72. More details on the specifications are provided in Lewandowski, Kurowicka and Joe (2009), whose authors' first initials are used to define 'LKJ,' too. Also, Alvarez, Niemi, and Simpson (2014) present a recent simulation study, on different priors for the variance-covariance matrices.

5.2.2 Modeling Averages—Area-Level Univariate HB Linear Three-Fold Model

An area-level univariate HB linear three-fold model was used for estimating PIAAC averages in domains. Similar to the HB models for proportions, the models for averages include three levels of random effects: county, state, and census division. The model is specified as follows:

$$\begin{aligned} Y_{ijk} &\sim N(\theta_{ijk}, \sigma_{ijk}^2) \\ \theta_{ijk} &\sim X_{ijk}'\beta + c_{ijk} + v_{ij} + d_i, \end{aligned}$$

where Y_{ijk} is the survey regression estimate of average literacy or numeracy scores at the county level, normally distributed with associated estimated variance σ_{ijk}^2, X_{ijk}' is a vector of covariates, β is a vector of coefficients, and c_{ijk}, v_{ij}, d_i are county-level, state-level, and division-level random effects, respectively. The estimated variances σ_{ijk}^2 are the result of smoothing functions of the variances for the survey regression estimates, and treated as fixed and known in the HB model. Independent priors are assumed for the regression coefficients and the random effects. Specifically, it is assumed $\beta \sim N(0,1000)$, where the normal distribution specification uses the mean and the standard deviation. It is also assumed that the random effects are mutually independent, following normal distributions,

$$\begin{aligned} c_{ijk} &\sim N(0, \sigma_c^2) \\ v_{ij} &\sim N(0, \sigma_v^2) \\ d_i &\sim N\left(0, \sigma_d^2\right). \end{aligned}$$

The variances of the random effects σ_c^2, σ_v^2, and σ_d^2 are assumed to follow a uniform prior distribution over a wide range, 0 to 1,000. The choice of vague (almost noninformative) priors for the model parameters ensures that little information about the values of the parameters is provided to the model, and hence the data (likelihood) have the major role in the posterior distribution.

5.3 Model Fitting

Two bivariate HB models and two univariate HB models were fitted for estimating proportions and averages for literacy and numeracy, respectively, based on the data from 184 counties in the PIAAC sample. One county was excluded due to its small sample size of 2 and negative survey regression estimate for proportion at or below Level 1 for literacy. To ensure reproducibility, the R and STAN (where STAN interfaces with R) starting points used in the generation of sequences of random numbers (seeds) were set equal to constants.

5.3.1 Software

The software chosen for the PIAAC state and county indirect estimates work was driven by model choice, which was driven mainly from the literature review. The review helped steer the decision toward a highly complex specification for proportions, which resulted in an area-level bivariate HB linear three-fold model. A simulation study, presented in appendix B, shows that bivariate models may work better for estimating correlated statistics (e.g., proportion of population that are at different levels of literacy or numeracy proficiency) while univariate models are much easier to implement. For averages, an area-level HB univariate linear three-fold model was chosen.

The Bayesian inference Using Gibbs Sampling (BUGS), Just Another Gibbs Sampler (JAGS), and STAN (named in honor of Stanislaw Ulam, a pioneer of the underlining Monte Carlo method[15]) are three popular software options for fitting HB models. WinBUGS is an established and stable stand-alone version of the software with no further development. WinBUGS (Lunn et al. 2000) uses the MCMC techniques to obtain estimates of posterior distributions and was used in the NAAL and National Adult Literacy Survey (NALS) SAE process. JAGS is a program for analysis of Bayesian hierarchical models using MCMC simulation, developed by Martyn Plummer. It was designed to remain similar to the BUGS family as well as add improvements. STAN refers both to a programming language and open source statistical software that implements it. STAN interfaces with several languages, including R and Python, as well as through a command line version. For R, the package RSTAN provides this interface. Some studies use INLA, http://www.r-inla.org/ (for example, the paper cited in this report, Chen, Wakefield, and Lumely 2014).

A thorough review and testing of the available software concluded that the STAN programming language would be the most appropriate software for PIAAC, and thus, chosen for the simulation study described in appendix B as well as the development and production phases of the PIAAC SAE process. The R package RSTAN provides the most flexibility in model fitting, prediction and diagnostics for the HB models we chose for the PIAAC SAE study, compared to JAGS and WinBUGS. It works better to handle models with complex structures and various types of priors for model hyper-parameters. *Shinystan* in R makes it very convenient to conduct model diagnostics through visual and numerical summaries.

[15] Available at http://www.stat.columbia.edu/~gelman/research/published/stan_jebs_2.pdf.

5.3.2 Model Estimation

MCMC methods are used to fit HB models, which are the models chosen for PIAAC (section 2.4 provides discussion that led to the use of HB models, including points made during the U.S. Summit of International SAE Experts). The RSTAN software was employed for this purpose. Three independent Markov Chains were processed to facilitate the calculation of Monte Carlo standard errors (see Gelman and Rubin 1992; Rao 2003, p. 229). The procedure started with three sets of initial values for β (while the initial values for other model parameters were randomly generated within RSTAN) corresponding to the three independent MCMC chains and then updated all the parameter values of $\eta = (\theta, \beta, c, v, d, \Sigma_c, \Sigma_v, \Sigma_d)$ for proportions or $\eta = \left(\theta, \beta, c, v, d, \sigma_c^2, \sigma_v^2, \sigma_d^2\right)$ for averages in each chain.

One set of the initial values of β were created from running weighted linear regressions of the proportions or averages on the set of seven covariates, with the county sample size being the weight. The results are shown in table 5-1. The other two sets of initial values were derived by adding or subtracting a constant from the first set of the initial values.

Table 5-1. Initial parameter values of β for literacy and numeracy proportions and averages: 2012/2014/2017

Parameter	Literacy P1	Literacy P3	Numeracy P1	Numeracy P3	Literacy average	Numeracy average
Intercept	0.13	0.15	0.30	-0.03	250.28	231.93
Education—LH	0.59	-0.02	0.40	0.27	-46.93	-22.03
Education—MH	-0.16	0.75	-0.34	0.85	67.00	81.71
Poverty	0.16	-0.12	0.37	-0.13	-2.94	-30.13
Black	0.15	-0.16	0.23	-0.17	-20.34	-28.94
Health insurance	-0.17	-0.21	-0.01	-0.31	-4.66	-20.64
Hispanic	0.20	-0.11	0.19	-0.10	-28.35	-32.35
Service occupations	0.30	-0.32	0.30	-0.35	-69.23	-72.84

SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

Given a set of initial values, each run was then processed separately. For the first iteration in a run, the value of the first component of $\eta = (\theta, \beta, c, v, d, \Sigma_c, \Sigma_v, \Sigma_d)$, i.e., θ, was updated, then the next component, β, was updated using the updated value of the first component θ and the initial values of the other components, and then the third component, c, was updated using the updated values of the first two components and the initial values of the remaining components, and so on. The run's second iteration started with the updated values of all eight components of η and repeated the process. The process was repeated 5,000 times, which is where convergence was determined to have been reached. The iterations up to this point (called the warm-up period) were discarded.

After that point, for proportions models, 30,000 further iterations were produced for each of the three runs. Since the results from neighboring iterations after warm-up are correlated, they were "thinned" by taking a systematic sample of 1 in every 20; for averages models, 15,000 further iterations were produced for each of the three runs and were "thinned" by taking 1 in every 10.[16] Thus, over the three runs, a total of 4,500 iterations remained. These 4,500 final iterations (referred to as MCMC samples) then simulated the posterior distributions of all the parameters in η. The means of the parameter estimates across the 4,500 MCMC samples are the HB estimates of the parameters. The results of the final HB models are shown in tables 5-2 and 5-3 for the parameters β and the variances of the random effects. The signs of the regression coefficients are as expected in general. As an example for illustration, Education—LH is a significant predictor with a positive sign in the models for literacy and numeracy proportions P1. As another example, Black is a significant predictor with a positive sign in the models for literacy and numeracy proportions P1; and it is a significant predictor with a negative sign in the models for literacy and numeracy proportions P3 and averages. We note that Health insurance and service occupations are not significant (the 95 percent credible interval includes zero) in all the models. However, health insurance is marginally significant in the model for numeracy proportions P3 and service occupations is marginally significant in the models for literacy and numeracy averages.

The posterior means of the intraclass correlations (ICCs) were computed for the four models for both literacy and numeracy. The ICCs are defined as:

$$
\begin{aligned}
(county\ ICC) \quad ICC_{c,ijk} &= \Sigma_c\left(\Sigma_c + \Sigma_{ijk}\right)^{-1}, \\
(state\ ICC) \quad ICC_{v,ijk} &= \Sigma_v\left(\Sigma_v + \Sigma_{ijk}\right)^{-1}, \\
(division\ ICC) \quad ICC_{d,ijk} &= \Sigma_d\left(\Sigma_d + \Sigma_{ijk}\right)^{-1},
\end{aligned}
$$

for the bivariate models for proportions, and as

$$
\begin{aligned}
(county\ ICC) \quad ICC_{c,ijk} &= \frac{\sigma_c^2}{\sigma_c^2 + \sigma_{ijk}^2}, \\
(state\ ICC) \quad ICC_{v,ijk} &= \frac{\sigma_v^2}{\sigma_v^2 + \sigma_{ijk}^2}, \\
(division\ ICC) \quad ICC_{d,ijk} &= \frac{\sigma_d^2}{\sigma_d^2 + \sigma_{ijk}^2},
\end{aligned}
$$

for the univariate models for averages. A summary of ICCs is as follows:

[16] Thinning the sample reduces the autocorrelation between iterations. The "thinning" factor was chosen to ensure that the autocorrelation function computed from the "thinned" sample is close to zero.

For literacy averages:

- County ranges from 0.04 to 0.47, with a mean of 0.23.
- State ranges from 0.02 to 0.23, with a mean of 0.10.
- Division ranges from 0.02 to 0.24, with a mean of 0.11.

For literacy proportions at or below Level 1:

- County ranges from 0.00 to 0.08, with a mean of 0.02.
- State ranges from 0.00 to 0.11, with a mean of 0.03.
- Division ranges from 0.00 to 0.08, with a mean of 0.03.

For literacy proportions at or above Level 3:

- County ranges from 0.00 to 0.16, with a mean of 0.04.
- State ranges from 0.00 to 0.11, with a mean of 0.03.
- Division ranges from 0.00 to 0.22, with a mean of 0.06.

For numeracy averages:

- County ranges from 0.04 to 0.45, with a mean of 0.22.
- State ranges from 0.02 to 0.22, with a mean of 0.10.
- Division ranges from 0.02 to 0.20, with a mean of 0.09.

For numeracy proportions at or below Level 1:

- County ranges from 0.00 to 0.10, with a mean of 0.03.
- State ranges from 0.00 to 0.11, with a mean of 0.03.
- Division ranges from 0.00 to 0.10, with a mean of 0.03.

For numeracy proportions at or above Level 3:

- County ranges from 0.01 to 0.27, with a mean of 0.08.
- State ranges from 0.00 to 0.16, with a mean of 0.04.
- Division ranges from 0.00 to 0.17, with a mean of 0.05.

In general, the county ICCs tend to be larger than the state and division ICCs. Note that for the univariate models for averages three scalar ICCs were defined; however, for the bivariate models for proportions three ICC matrices (of dimension 2 x 2) were defined.

Table 5-2. Regression coefficients and components of the variance-covariance matrices of random effects for the final HB models: For literacy and numeracy proportions: 2012/2014/2017

Model	Parameters	HB mean	HB standard deviation	Median	95 percent credible interval Lower bound	Upper bound
Literacy proportions P1	Intercept	0.12	0.09	0.12	-0.05	0.29
	Education—LH	0.67	0.23	0.67	0.22	1.11
	Education—MH	-0.13	0.09	-0.13	-0.31	0.06
	Poverty	0.20	0.15	0.20	-0.10	0.50
	Black	0.16	0.05	0.16	0.07	0.25
	Health insurance	-0.07	0.16	-0.08	-0.38	0.25
	Hispanic	0.19	0.06	0.19	0.07	0.30
	Service occupations	0.10	0.18	0.10	-0.26	0.46
Literacy proportions P3	Intercept	0.09	0.10	0.09	-0.10	0.28
	Education—LH	-0.02	0.25	-0.02	-0.54	0.45
	Education—MH	0.81	0.11	0.81	0.59	1.02
	Poverty	-0.14	0.18	-0.14	-0.49	0.21
	Black	-0.17	0.05	-0.17	-0.27	-0.08
	Health insurance	-0.15	0.20	-0.15	-0.52	0.23
	Hispanic	-0.12	0.06	-0.12	-0.25	0.01
	Service occupations	-0.19	0.21	-0.19	-0.60	0.22
Literacy proportions	$\Sigma_{c,1,1}$	0.00010	0.00013	0.00005	0.00000	0.00045
	$\Sigma_{c,1,2}$, $\Sigma_{c,2,1}$	-0.00007	0.00013	-0.00002	-0.00045	0.00005
	$\Sigma_{c,2,2}$	0.00023	0.00027	0.00012	0.00000	0.00096
	$\Sigma_{v,1,1}$	0.00013	0.00016	0.00007	0.00000	0.00057
	$\Sigma_{v,1,2}$, $\Sigma_{v,2,1}$	-0.00005	0.00011	-0.00001	-0.00035	0.00007
	$\Sigma_{v,2,2}$	0.00015	0.00020	0.00008	0.00000	0.00073
	$\Sigma_{d,1,1}$	0.00010	0.00018	0.00004	0.00000	0.00055
	$\Sigma_{d,1,2}$, $\Sigma_{d,2,1}$	-0.00001	0.00014	0.00000	-0.00030	0.00021
	$\Sigma_{d,2,2}$	0.00037	0.00072	0.00017	0.00000	0.00186

See notes at end of table.

Table 5-2. Regression coefficients and components of the variance-covariance matrices of random effects for the final HB models: For literacy and numeracy proportions: 2012/2014/2017—Continued

Model	Parameters	HB mean	HB standard deviation	Median	95 percent credible interval: Lower bound	95 percent credible interval: Upper bound
Numeracy P1	Intercept	0.28	0.10	0.28	0.09	0.47
	Education—LH	0.52	0.25	0.52	0.03	1.00
	Education—MH	-0.30	0.11	-0.30	-0.51	-0.10
	Poverty	0.37	0.17	0.37	0.04	0.72
	Black	0.26	0.05	0.26	0.16	0.36
	Health insurance	0.01	0.18	0.01	-0.34	0.36
	Hispanic	0.18	0.07	0.18	0.05	0.31
	Service occupations	0.18	0.22	0.18	-0.25	0.59
Numeracy P3	Intercept	-0.16	0.10	-0.16	-0.35	0.03
	Education—LH	0.53	0.25	0.53	0.06	1.03
	Education—MH	0.97	0.11	0.97	0.76	1.19
	Poverty	-0.13	0.18	-0.13	-0.47	0.22
	Black	-0.17	0.05	-0.17	-0.27	-0.08
	Health insurance	-0.32	0.19	-0.32	-0.69	0.06
	Hispanic	-0.16	0.06	-0.16	-0.29	-0.04
	Service occupations	-0.17	0.21	-0.17	-0.57	0.24
Numeracy proportions	$\Sigma_{c,1,1}$	0.00019	0.00022	0.00011	0.00000	0.00078
	$\Sigma_{c,1,2}, \Sigma_{c,2,1}$	-0.00018	0.00024	-0.00009	-0.00080	0.00005
	$\Sigma_{c,2,2}$	0.00047	0.00042	0.00038	0.00000	0.00149
	$\Sigma_{v,1,1}$	0.00015	0.00020	0.00008	0.00000	0.00072
	$\Sigma_{v,1,2}, \Sigma_{v,2,1}$	-0.00008	0.00015	-0.00002	-0.00050	0.00007
	$\Sigma_{v,2,2}$	0.00023	0.00027	0.00014	0.00000	0.00094
	$\Sigma_{d,1,1}$	0.00014	0.00025	0.00005	0.00000	0.00084
	$\Sigma_{d,1,2}, \Sigma_{d,2,1}$	-0.00005	0.00016	-0.00001	-0.00046	0.00012
	$\Sigma_{d,2,2}$	0.00026	0.00052	0.00011	0.00000	0.00136

SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

Table 5-3. Regression coefficients and variances of random effects for the final HB models: For literacy and numeracy averages: 2012/2014/2017

Model	Parameters	HB mean	HB standard deviation	Median	95 percent credible interval Lower bound	Upper bound
Literacy	Intercept	243.50	11.53	243.61	220.94	265.65
	Education—LH	-29.87	30.44	-30.23	-88.48	29.83
	Education—MH	72.46	12.89	72.22	47.94	97.77
	Poverty	-17.83	20.80	-17.43	-58.66	22.72
	Black	-23.49	5.97	-23.42	-35.57	-11.87
	Health insurance	-14.30	24.32	-13.90	-63.55	31.51
	Hispanic	-28.79	7.94	-28.78	-43.93	-12.58
	Service occupations	-42.78	24.42	-42.95	-90.01	5.63
	σ_c^2	15.32	6.74	14.64	3.99	30.33
	σ_v^2	5.70	5.35	4.25	0.16	19.44
	σ_d^2	6.96	10.52	3.74	0.14	33.44
Numeracy	Intercept	223.99	12.60	224.26	198.98	248.55
	Education—LH	1.36	32.93	1.52	-64.17	67.30
	Education—MH	88.38	13.93	88.39	61.28	116.52
	Poverty	-44.31	22.31	-44.75	-88.42	0.45
	Black	-34.40	6.58	-34.47	-46.93	-21.39
	Health insurance	-25.65	25.79	-25.87	-76.49	24.29
	Hispanic	-36.47	8.58	-36.50	-53.64	-20.25
	Service occupations	-46.98	26.69	-47.04	-99.43	5.46
	σ_c^2	17.71	8.20	16.91	4.33	35.98
	σ_v^2	6.53	5.79	5.03	0.26	21.49
	σ_d^2	6.63	10.32	3.57	0.12	30.79

SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

5.4 Predicted Values

As mentioned above, estimates of the parameters in each of the 4,500 MCMC samples were produced through the RSTAN software for sampled counties. Once the final model was processed and the model parameters estimated, the next step was to estimate the proportions and averages for the sampled counties, nonsampled counties, and for states and nation. The prediction process for sampled and nonsampled counties is described in sections 5.4.1 and 5.4.2, respectively. The process for making state- and national-level estimates is described in section 5.4.3.

5.4.1 Indirect Estimates for Sampled Counties

For the sampled counties, the posterior mean $\hat{\theta}_{ijk}^{HB}$, which is also called the county-level HB model prediction (for proportions or averages), for sampled county k in state j and division i, is produced by the RSTAN software as:

$$\hat{\theta}_{ijk}^{HB} = \frac{\sum_{b=1}^{4,500} \theta_{ijk}^{(b)}}{4,500}, \tag{5a}$$

where the value of $\theta_{ijk}^{(b)}$ for MCMC sample b is obtained from

$$\theta_{ijk}^{(b)} = X'_{ijk}\beta^{(b)} + c_{ijk}^{(b)} + v_{ij}^{(b)} + d_i^{(b)}. \tag{5b}$$

5.4.2 Indirect Estimates for Nonsampled Counties

For the sampled counties, estimates of all the components on the right-hand side of equation (5b) were available. However, for all of the nonsampled counties, the values of $c_{ijk}^{(b)}$ were not available, and for the nonsampled counties in states without a sampled county, values of $v_{ij}^{(b)}$ were not available either. To simulate the MCMC procedure, in cases where a component was not available, it was drawn at random from the appropriate normal distribution. Thus, following Rao (2003), for proportions, $\boldsymbol{c}_{\boldsymbol{ijk}}^{(\boldsymbol{b})}$ was drawn from $N\left(0, \Sigma_c^{(b)}\right)$ and for averages, $\boldsymbol{c}_{\boldsymbol{ijk}}^{(\boldsymbol{b})}$ was drawn from $N\left(0, \sigma_c^{2(b)}\right)$. When necessary, $\boldsymbol{v}_{\boldsymbol{ij}}^{(\boldsymbol{b})}$ was drawn from $N\left(0, \Sigma_v^{(b)}\right)$ for proportions and $\boldsymbol{v}_{\boldsymbol{ij}}^{(\boldsymbol{b})}$ was drawn from $N\left(0, \sigma_v^{2(b)}\right)$ for averages.

For the nonsampled counties in states with one or more sampled counties, the estimated state effect was available from RSTAN. For such counties, the model prediction of $\theta_{ijk}^{(b)}$ was computed from

$$\theta_{ijk}^{(b)} = X'_{ijk}\beta^{(b)} + c_{ijk(RD)}^{(b)} + v_{ij}^{(b)} + d_i^{(b)},$$

where $c_{ijk(RD)}^{(b)}$ is a random draw from $N\left(0, \Sigma_c^{(b)}\right)$ for proportions and from $N\left(0, \sigma_c^{2(b)}\right)$ for averages.

For the nonsampled counties in states with no sampled county, the estimate of $\theta_{ijk}^{(b)}$ was computed from

$$\theta_{ijk}^{(b)} = X_{ijk}'\beta^{(b)} + c_{ijk(RD)}^{(b)} + v_{ij(RD)}^{(b)} + d_i^{(b)},$$

where $v_{ij(RD)}^{(b)}$ is a random draw from $N\left(0, \Sigma_v^{(b)}\right)$ for proportions and from $N\left(0, \sigma_v^{2(b)}\right)$ for averages.

In both cases, once the set of 4,500 values of $\theta_{ijk}^{(b)}$ was obtained, the posterior mean for nonsampled counties was computed using equation (5a).

Due to the linearity of the model, an estimated proportion could be less than zero. This only occurred for the proportion at or above Level 3. Tables D-1 and D-2 in appendix D show one county with a negative value for the proportion at or above Level 3 in literacy, and eight counties with negative proportions at or above Level 3 in numeracy.

5.4.3 Indirect Estimates for States and Nation

The indirect estimates for states and nation were computed as weighted aggregates of small area county estimates, where the weights represent the proportion of the state's, division's, or nation's household population of adults aged 16–74 in each county. The county populations of the household residents 16–74 were obtained from the 2013–2017 ACS data.

Table 5-4 compares the national-level model predictions with the direct estimates for proportions and averages for literacy and numeracy. The two sets of estimates are not significantly different.

Table 5-4. National-level indirect and direct estimates: 2012/2014/2017

Model	Indirect estimate	Posterior standard deviation of indirect estimate	Direct estimate	Standard error of direct estimate
Literacy P1	0.218	0.0048	0.226	0.0050
Literacy P2	0.323	0.0062	0.322	0.0063
Literacy P3	0.458	0.0056	0.452	0.0061
Numeracy P1	0.319	0.0052	0.321	0.0069
Numeracy P2	0.322	0.0062	0.321	0.0069
Numeracy P3	0.360	0.0054	0.359	0.0074
Literacy average	263.5	0.61	263.3	0.44
Numeracy average	249.1	0.67	248.9	0.84

SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

5.5 Measures of Precision for the Indirect Estimates

The primary measure of precision reported for each county-level indirect estimate is its credible interval, described in section 5.5.1. An alternative measure of uncertainty is the coefficient of variation (CV), discussed in section 5.5.2. An assessment of precision of the indirect estimates using both measures is provided in section 5.5.3.

5.5.1 Credible Intervals

A credible interval is a posterior probability interval used in Bayesian statistics for purposes similar to those of a confidence interval in frequentist statistics. A 95 percent credible interval is any interval with a probability under the posterior distribution of 0.95. For example, a statement such as "following the model result, a 95 percent credible interval for θ is 7 percent to 21 percent" means that the posterior probability that θ lies in the interval from 7 percent to 21 percent is 0.95. The 95 percent credible intervals for the county estimates $\hat{\theta}_{ijk}^{HB}$ (or the state estimates $\hat{\theta}_{ij}^{HB}$, the division estimates $\hat{\theta}_{i}^{HB}$, and the national estimate $\hat{\theta}^{HB}$) were computed by calculating the 2.5 percent (lower bound) and 97.5 percent (upper bound) quantiles of $\theta_{ijk}^{(b)}$ (or $\theta_{ij}^{(b)}$, $\theta_{i}^{(b)}$, and $\theta^{(b)}$), respectively, from the 4,500 MCMC samples that simulated the posterior distributions. Since these posterior distributions are skewed, the credible intervals are nonsymmetric around the estimate. Figure 6-15 in Section 6 shows the credible intervals of literacy proportions relative to the survey regression estimate for each sampled county.

Due to the linearity of the model, the credible interval's lower bound of an estimated proportion could be less than zero. This occurred for the proportion at or below Level 1 and the proportion at or above Level 3 mostly for counties without PIAAC sample. Tables D-3, D-4, and D-5 show a list of counties where negative values occur for the credible interval's lower bound for literacy and numeracy, respectively.

5.5.2 Coefficient of Variation

The CV of the HB estimate for county *k* in state *j* and division *i* is computed as

$$CV_{ijk} = \frac{\sqrt{\mathrm{Var}\left(\hat{\theta}_{ijk}^{HB}\right)}}{\hat{\theta}_{ijk}^{HB}},$$

where the posterior variance $\mathrm{Var}\left(\hat{\theta}_{ijk}^{HB}\right)$ is computed as

$$\mathrm{Var}\left(\hat{\theta}_{ijk}^{HB}\right) = \frac{\sum_{b=1}^{4,500}\left(\theta_{ijk}^{(b)} - \hat{\theta}_{ijk}^{HB}\right)^2}{4,500-1}.$$

The CV for states and nation can be computed similarly.

5.5.3 Assessment of Precision Measures

Table 5-5 summarizes the distributions of the widths (the difference between the upper bound and the lower bound) of the credible intervals as well as the CVs for the 3,142 counties and 51 states in the United States, for literacy proportion at or below Level 1.

Table 5-5. Distribution of credible interval widths and coefficients of variation for indirect estimates for literacy proportion at or below Level 1: 2012/2014/2017

Statistic	Percentile 20	Percentile 40	Percentile 60	Percentile 80	Median
County indirect estimates					
95 percent credible interval width (percent)	7.2	7.7	8.3	9.4	8.0
Coefficient of variation (percent)	7.8	9.1	10.9	13.5	10.0
Sampled counties					
95 percent credible interval width (percent)	6.6	7.0	7.4	8.1	7.2
Coefficient of variation (percent)	7.4	8.7	10.0	12.8	9.3
Nonsampled counties					
95 percent credible interval width (percent)	7.2	7.8	8.4	9.5	8.0
Coefficient of variation (percent)	7.8	9.1	10.9	13.6	10.0
State indirect estimates					
95 percent credible interval width (percent)	5.2	5.7	6.4	6.9	6.1
Coefficient of variation (percent)	6.3	7.0	8.6	10.3	8.1

SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

Overall, the state predictions are more precise than the county predictions, and to a less extent, the counties with sample are more precise than counties without PIAAC sample (figure 6-17 in section 6 shows that the indirect estimates of literacy proportions are more precise than the survey regression estimates for counties with PIAAC sample). For example, the median credible interval width for county predictions is 8.0 percent (i.e., percentage points), while the median is 6.1 percent for state predictions.

The table also shows that the median credible interval width is 7.2 percent for counties with PIAAC sample cases and 8.0 percent for counties without PIAAC sample. The CVs for the county-level model predictions are of the order of 10 percent. Estimates with CVs of this magnitude are considered precise. It is important for the users of these county predictions to recognize that for some counties the CVs can be large. While the state predictions are more precise, with a median CV of 8.1 percent, it is still important for users to consider the credible interval along with the model prediction.

5.6 Simultaneous Inference

As mentioned in chapter 1, the PIAAC state and county indirect estimates website is available to make comparisons between areas on eight outcomes. That is, for each of literacy and numeracy, comparisons are conducted on the proportion at or below Level 1, proportion at Level 2, proportion at Level 3 and above, and the average. The areas involved in the comparisons cover the following:

- state-to-nation;
- state-to-state;
- county-to-state; and
- county-to-county.

Often, the user may be interested in conducting multiple comparisons to make simultaneous inferences.

Initial interest was in pairwise comparisons rather than multiple testing. We acknowledge that both multiple comparisons and multiple testing are subareas of simultaneous inference. However, multiple comparisons apply to simultaneous comparisons of values for the same measurement/quantity of interest, and represent a common area in the literature; see, for example the book by Hsu (1996). For example, the comparisons of k treatment means for a one-way model $y_{ij} = \mu_i + \epsilon_{ij}, i = 1, \ldots, k$ and $j = 1, \ldots, n_i$, falls under this category, of multiple comparisons. On the other hand, multiple testing applies to simultaneous tests of values for multiple measurements/quantities of interest. For example, the testing of k measurements (weight, blood pressure, heart rate, etc.) taken on a random group of people, to make inference on the effect of a treatment, falls into the multiple testing category.

Computations are conducted for one-to-one pairwise comparisons allowed between any two states or any two counties (i.e., within or across states). Figure 5-1 is an illustration showing a comparison of two counties from the same states.

Figure 5-1. Illustration of county-to-county comparison: 2012/2014/2017

SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

The final interest is in multiple testing. Particularly to PIAAC, for a given pair of domains, (county1, county2) or (county1, state1) or (state1, state2) or (state1, nation), the interest is in testing simultaneously whether zero is significantly different from the differences between the domains in eight measurements mentioned above. Hence, the set consists of eight tests,

$$\theta_i^{L,p1} - \theta_j^{L,p1} = 0, \theta_i^{L,p2} - \theta_j^{L,p2} = 0, \theta_i^{L,p3} - \theta_j^{L,p3} = 0, \theta_i^{L,mean} - \theta_j^{L,mean} = 0,$$

$$\theta_i^{N,p1} - \theta_j^{N,p1} = 0, \theta_i^{N,p2} - \theta_j^{N,p2} = 0, \theta_i^{N,p3} - \theta_j^{N,p3} = 0, \theta_i^{N,mean} - \theta_j^{N,mean} = 0,$$

where $\theta_k^{measure,characteristic}$ denotes the parameter of interest for domain k, with measure ∈ $\{L, N\}$ being literacy or numeracy, and characteristic ∈ $\{p1, p2, p3, mean\}$ being proportion at or below Level 1, proportion at Level 2, proportion at or above Level 3, or mean. For this, the distributions for the differences are constructed using the MCMC iterates available from the model fit, for the eight quantities: literacy proportions and averages, and numeracy proportions and averages.

For the PIAAC SAE website, we applied a Bonferroni adjustment to the simultaneous $100(1-\alpha)$ credible intervals, and text appears with the results, conditional on chosen areas of interest. The text discusses statistical differences under the Bonferroni adjusted critical significance level, while mentioning notable differences, which are not significant under the Bonferroni test, but would have been significant under a single test at the 0.05 level of significance.

6. MODEL DIAGNOSTICS, SENSITIVITY ASSESSMENT, AND EVALUATION

Large-scale small area estimation (SAE) programs require an extensive model evaluation process to ensure goodness of fit, given the assumptions made at the beginning of the development stage. Moreover, the quality of a large number of small area predictions constructed for domains with no survey data depends on the goodness of fit of the model. For our study, model predictions are constructed for over 90 percent of the 3,143 U.S. counties. Therefore, the U.S. Program for the International Assessment of Adult Competencies (PIAAC) International SAE Experts emphasized the need for thorough and in-depth model evaluations leading to the choice of the final model. Hence, we devote this chapter to a discussion of multiple approaches for evaluating the developed SAE models. Internal and external checks are illustrated in sections 6.1 and 6.2, respectively. In summary, the assessments that are discussed in this section were part of several iterations of adjusting and improving the models that were developed. The results shown in this section are related to the checks conducted on the final models.

6.1. Internal Model Validation

In SAE, inference relies on model assumptions. Therefore, it is critical to check the validity of the assumptions. Moreover, when inference is conducted using Markov Chain Monte Carlo (MCMC) output using multiple chains for Hierarchical Bayes (HB) models, convergence and chain mixing diagnostics should be part of the model development methodology. Particularly, diagnostics may indicate whether the inference is based on a good approximation of the target density, to which the Monte Carlo (MC) chains have converged. If convergence is not achieved, then a larger number of iterations, alternative starting values, or alternative tuning parameters for the algorithm used to fit the models should be considered. In addition, if the chains do not mix well or take a long time to achieve well-mixing, then the model specification needs to be revised. Note that values from the true target density are only obtained when the number of MC samples goes to infinity, which is not the case in practical applications.

Internal model validation consists of checking the model for its accuracy and robustness. Some examples of internal model validation are as follows. Alternative models were fit to the data for the model sensitivity checks and model specification (distributional assumptions) checks. Cruze et al. (2019) conducted model sensitivity tests by changing the prior distribution for the variance of the random effects in a univariate Fay-Herriot HB model, and reported failure of convergence for the choices of Inverse-Gamma adopted for the variance and Uniform adopted for the log of the variance, whereas the Uniform prior adopted for the variance parameter resulted in good convergence. The convergence and mixing diagnostics were conducted for each alternative model fit and model assumptions checks were conducted for the alternative model fits that proves competitive to the prior fits. Validation measures for convergence and mixing of MCMC chains are mentioned in the STAN user guide available at https://mc-stan.org/docs/2_19/stan-users-guide-2_19.pdf. Also, HB studies in the literature report such diagnostics checks. See, for example, a United Kingdom report[17] (visual checks of trace plots and Gelman and Rubin's diagnostics tests) and Erciulescu, Cruze, and Nandram (2019) (trace plots, Gelman and Rubin's multiple potential scale reduction factors, MC effective sample size). A failure of the internal model checks indicates that a revision of the model might be necessary (for example, removing predictors or adopting different prior distributions).

As illustrated in section 5.2, for PIAAC two bivariate HB models were fitted independently to literacy and numeracy proportions, and two univariate HB models were fitted independently to literacy and numeracy averages. The final models presented in section 5.2 are the result of an iterative process for model development and model validation. To assess the four HB models, we developed various internal validation checks including:

- convergence and mixing diagnostics;
- checks on model assumptions using
  - a multicollinearity test using variance inflation factors;
  - residual analysis including
    - absence of obvious nonrandom patterns when checked against the fitted values; and
    - absence of outliers or other deviations from normality.
  - posterior predictive checks using different test statistics:
    - indicator function, comparing the survey regression estimates and the corresponding predictions simulated from the posterior predictive distribution for the county-level quantities (proportions or averages);
    - order statistics function, comparing the order of the survey regression estimates and the order of the corresponding predictions simulated from the posterior predictive distribution for the county-level quantities (proportions or averages);

[17] The United Kingdom report is available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/36077/12-1318-2011-skills-for-life-small-area-estimation-technical.pdf.

- difference between the model predictions and the corresponding predictions simulated from the posterior predictive distribution for the county-level quantities (proportions or averages); and

- difference and scaled (by the variance of the model prediction) difference between the survey regression estimates and the corresponding predictions simulated from the posterior predictive distribution for the county-level quantities (proportions or averages).

■ model sensitivity checks using

- initial values specification for nuisance parameters and random effects;

- hyperparameters specification for prior distribution parameters for the variance components; and

- changes in prior distributions for the variance components.

■ model specification using

- univariate versus bivariate model specification for the proportions;

- different software (stan) sampling algorithm (hamiltonian monte carlo/no-u-turn samplers) parameters (max_treedepth, adapt_delta, random seeds);

- different distributions for the sampling model; and

- different link function for the linkage model.

Internal model validation was conducted based on the set of 184 counties to which the models were fitted. For all the fits of the models compared against the final models, estimation and prediction are implemented in RSTAN, using the default parameters for the sampling algorithms and three MC chains. Most of the models were fit using 20,000 samples per MC chain, and thinned every 10 samples, after dropping the first 5,000 samples as burn-in. Hence, inference was conducted using 4,500 samples, from the three chains combined. Selected results are included in this section for the literacy proportions models. The results for numeracy proportions models, and for literacy and numeracy average scores models, are similar to the ones for literacy proportions models, and only briefly mentioned in each of the following subsections.

6.1.1 Convergence and Mixing Diagnostics

Convergence and mixing diagnostics were performed using functions from the R package *coda* and the R function *launch_shinystan* in the library *shinystan*. The set of parameters monitored consists of 897 parameters: β, c, d, v, diagonal entries in S_c, S_d, and S_v, entries in Σ_c, Σ_v, and Σ_d, squares of the diagonal entries in Σ_c, Σ_v, and Σ_d, sums of squares of the diagonal entries in Σ_c, Σ_v, and Σ_d, the entries in the correlation matrices Ω_c, Ω_d, and Ω_v, θ, and the log likelihood. The number of regression parameters β equals 16, corresponding to two sets of 8 parameters (intercept and coefficients associated with the 7 covariates included in the model). The number of county-level random effects parameters c is 368, corresponding to two sets of 184 (number of counties used for modeling) parameters, the number of division-level random effects parameters d is 18, corresponding to two sets of 9 (number of divisions used for modeling) parameters, and the number of state-level random effects parameters v is 88, corresponding to two sets of 44 (number of states used for modeling) parameters. The diagonal matrices S_c, S_d, and S_v are of dimensions 2 by 2, hence there are 6 parameters, corresponding to three sets of two (their diagonal entries). The random-effects variance matrices Σ_c, Σ_v, and Σ_d are of dimensions 2 by 2, hence there are 12 parameters corresponding to three sets of four (their entries), and 6 parameters corresponding to three sets of two (their diagonal entries squared). Two additional variance parameters correspond to the sums of squares of two diagonal entries in the random-effects variance matrices Σ_c, Σ_v, and Σ_d. The correlation matrices Ω_c, Ω_d, and Ω_v are of dimensions 2 by 2, hence there are 12 parameters, corresponding to three sets of four (their entries). The number of county-level parameters θ is 368, corresponding to two sets of 184 (number of counties used for modeling) parameters. There is one additional parameter for the log likelihood. Note that the diagonal entries in the correlation matrices are equal to 1. The sets of parameters monitored in the fit of models with alternative distributional specifications and the sets of parameters monitored in the fit of various univariate models for averages are different from the set presented here, since they depend on the parameterization of the models.

Table 6-1 shows the mean and quartiles of the diagnostic statistics, including effective sample size, Gelman-Rubin $\hat{R}$ statistic (see Gelman and Rubin 1992), MC standard error, autocorrelation, and cross-correlation, across all monitored parameters. The Gelman-Rubin $\hat{R}$ diagnostic is based on a comparison of the between-chain and within-chain variances, and convergence is diagnosed when the output from the multiple, parallel, chains is indistinguishable and the different initial values that were provided to the multiple chains have been 'forgotten.' The results indicate that convergence and mixing of the three chains have been reached. Particularly, after accounting for autocorrelation, none of the monitored parameters has a MC chain sample size less than 5 percent of the total sample size (4,500 samples), or a MC standard error greater than 10 percent of the posterior standard deviation, or an $\hat{R}$ above 1.1. These criteria refer to acceptable margins of error derived based on theory and methods in Gelman et al. (2013) and STAN's guide and reference manual. Autocorrelations within chains and cross-correlation among the monitored parameters are low. Since two of the four components in the matrices Ω_c, Ω_d, and Ω_v, are fixed (equal to 1), their associated posterior standard deviations are zero (for the first diagonal entry) or nearly zero (for the second diagonal entry), their associated MC standard errors are not defined (for the first diagonal entry) or nearly zero (for the second diagonal entry), and leading to three undefined effective sample sizes (corresponding to the first diagonal entry). The cross-correlation is a matrix of dimension 897 by 897 and here we report summaries for all its entries, so the 1's correspond to the diagonal entries in that matrix. Convergence and mixing checks are conducted for the alternative model specifications for proportions, as well as for the various model specification for averages, and the diagnostics for the final models adopted do not indicate lack of convergence and mixing.

Table 6-1. Convergence diagnostics for the MCMC: 2012/2014/2017

Metric	$\hat{R}$	Effective sample size	MC standard error/posterior standard deviation	Auto-correlation Lag1	Auto-correlation Lag5	Cross-correlation
Minimum	0.9993	0.000	0.0000	-0.0520	-0.0439	-0.9554
1st quantile	0.9998	3,934.763	0.0149	0.0030	-0.0077	-0.0190
Median	1.0000	4,303.378	0.0153	0.0189	0.0045	-0.0001
Mean	1.0002	4,073.012	0.0162	0.0312	0.0095	0.0040
3rd quantile	1.0004	4,500.000	0.0160	0.0372	0.0178	0.0189
Maximum	1.0234	5,844.301	0.0982	0.9302	0.7471	1.0000

SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

6.1.2 Checks on Model Assumptions

To check some of the model assumptions, we conducted a collinearity test, residual analysis and a set of posterior predictive checks.

6.1.2.1 Collinearity Test

Collinearity may be detected among a set of covariates using the variance inflation factor (VIF). Each covariate is regressed on all the other covariates, and the resulting coefficients of determination, R_p^2, are used to construct the VIFs, where $p = 1, \ldots, 7$ denotes the covariate. Specifically, for covariate p, $VIF_p = (1 - R_p^2)^{-1}$. Large VIFs, typically larger than 10 (Erciulescu, Cruze, and Nandram [2019] used a similar threshold), indicate collinearity issues. Table 6-2 shows that there is no indication of collinearity among the set of covariates that were used in final SAE models. The variance inflation factors are all less than 10. The results in table 6-2 hold for all the different model specifications adopted for proportions and averages, because the same set of counties and the same set of covariates were used in all.

Table 6-2. Variance inflation factors: 2012/2014/2017

Variables	VIF
Education—LH	7.2240
Education—MH	4.8709
Poverty	4.0601
Black	1.7436
Health insurance	2.0829
Hispanic	2.6459
Service occupations	1.5397

SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

6.1.2.2 Residual Analysis

Naive normality checks using residuals indicate that there are no significant departures from normality, as shown in figures 6-1 and 6-2. Also, there is no significant pattern in the plot of residuals against the fitted values. For these checks, two sets of residuals are constructed, where the first set of residuals is defined as

$$r_{ijk,1}^{Set1} = \frac{P_{ijk,1} - X_{ijk}' \hat{\beta}_1}{\sqrt{\hat{\Sigma}_{c,1,1} + \hat{\Sigma}_{d,1,1} + \hat{\Sigma}_{v,1,1} + \Sigma_{ijk,1,1}}}$$

$$r_{ijk,2}^{Set1} = \frac{P_{ijk,2} - X_{ijk}' \hat{\beta}_2}{\sqrt{\hat{\Sigma}_{c,2,2} + \hat{\Sigma}_{d,2,2} + \hat{\Sigma}_{v,2,2} + \Sigma_{ijk,2,2}}}$$

and the second set of residuals is defined as

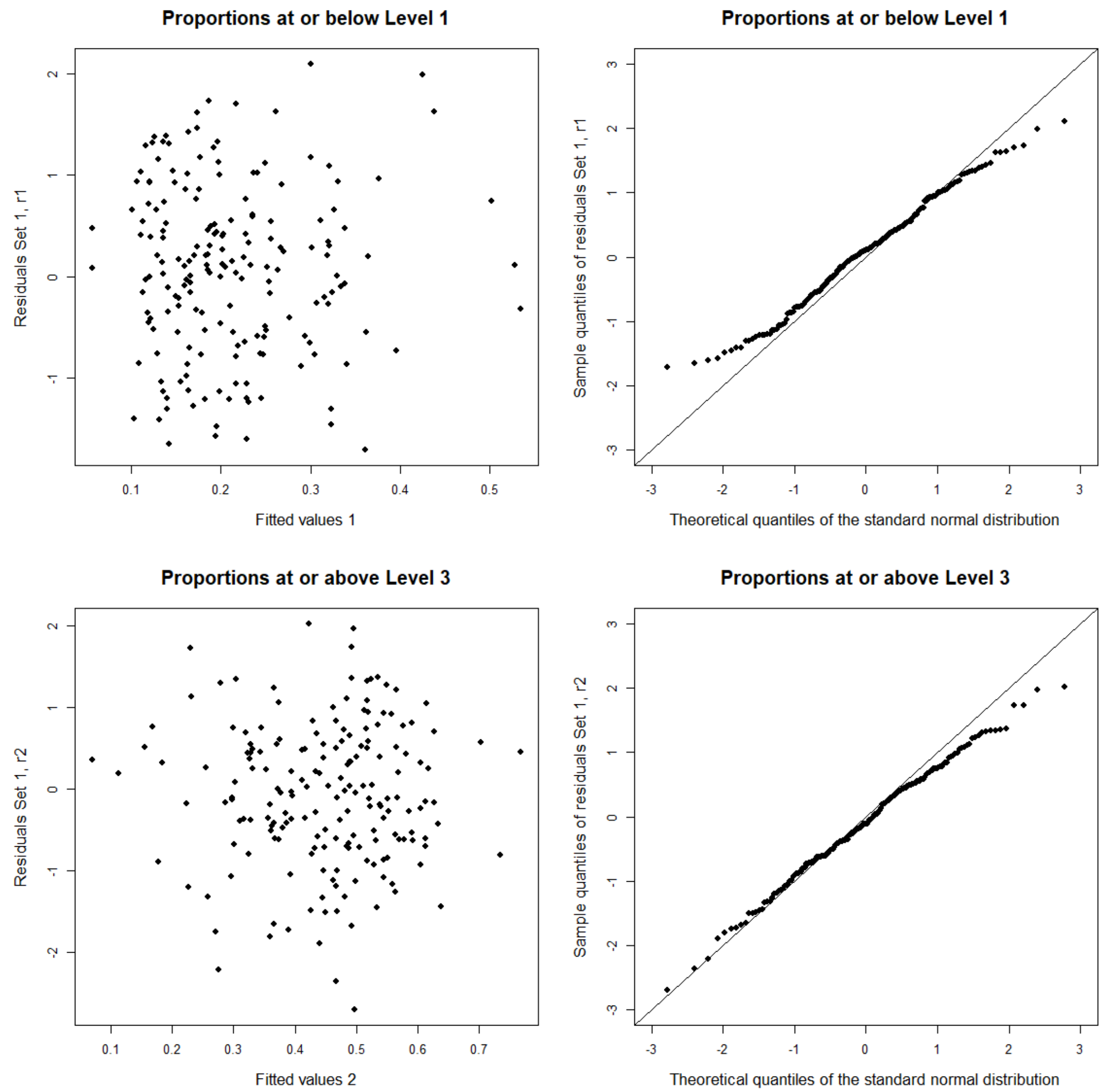
$$r_{ijk,1}^{Set2} = \frac{P_{ijk,1} - \hat{\theta}_{ijk,1}}{\sqrt{\Sigma_{ijk,1,1}}}$$

$$r_{ijk,2}^{Set2} = \frac{P_{ijk,2} - \hat{\theta}_{ijk,2}}{\sqrt{\Sigma_{ijk,2,2}}}$$

The two components in each of the two residual sets correspond to proportions at or below Level 1 ($P_{ijk,1}$) and at or above Level 3 ($P_{ijk,2}$), respectively. Note that the smoothed sampling variances, $\Sigma_{ijk,1,1}$ and $\Sigma_{ijk,2,2}$, are denoted without the hat symbol, because they were treated as fixed and known in the model fit. Also, note that the first set of residuals is inspired by the transformed residuals computed in

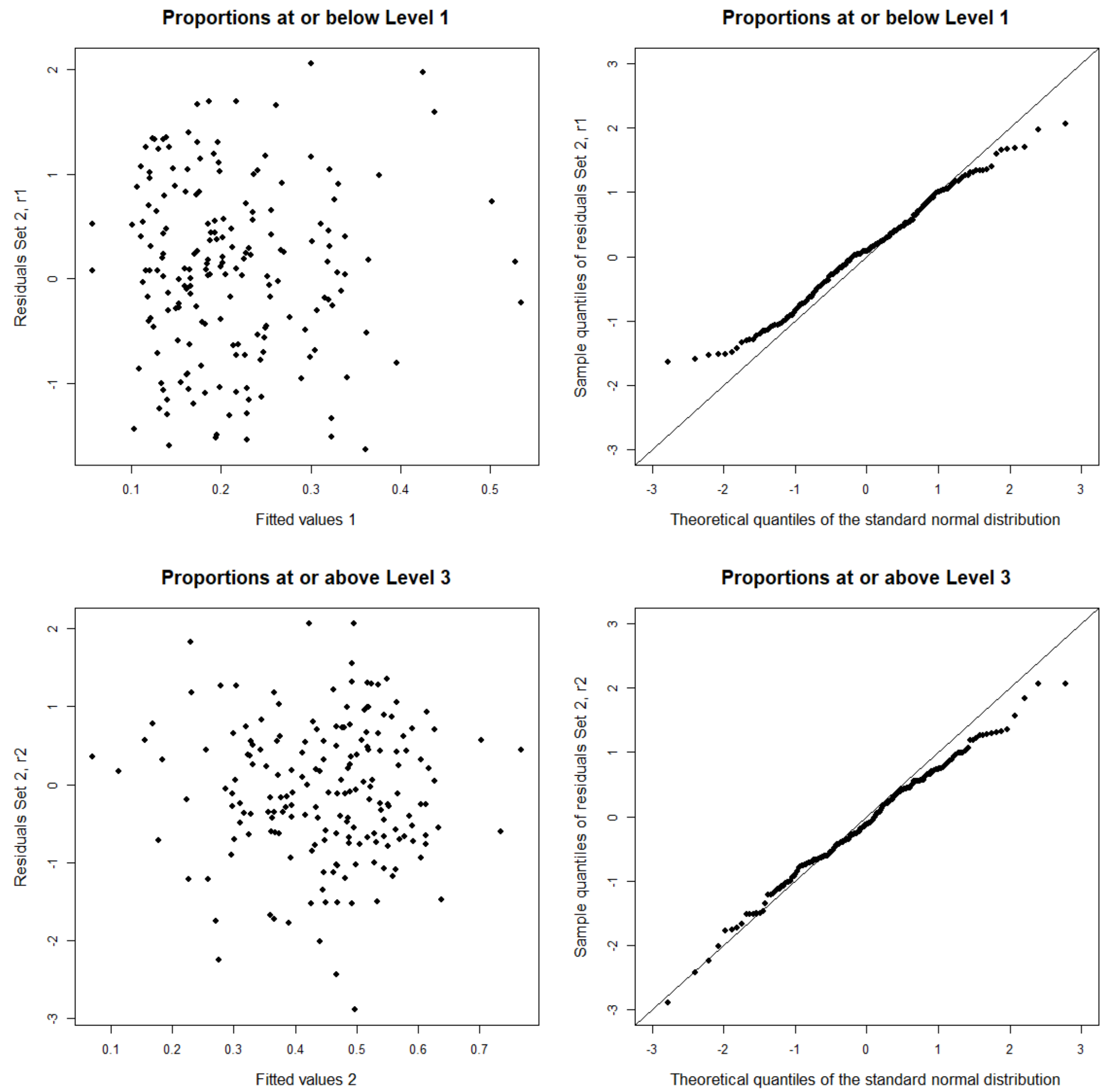
Battese, Harter, and Fuller (1988); the authors constructed unit-level transformed residuals that are approximately independent (the correlation between any two residuals is not exactly zero) and identically distributed with mean zero and variance equal to the common constant across all the units, while we constructed area-level transformed residuals (Set 1) that are approximately independent (the correlation between any two residuals is not exactly zero) and identically distributed with mean zero and variance equal to one.

Figure 6-1. Residual plots for the first set of residuals: 2012/2014/2017



NOTE: The top row corresponds to proportions at or below Level 1 and the bottom row corresponds to proportions at or above Level 3. The plots on the left-hand side illustrate the relationship between the residuals and the fitted values ($X'_{ijk}\hat{\beta}_{1(2)}$). The normal quantile plots are displayed on the right-hand side.
SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

Figure 6-2. Residual plots for the second set of residuals (conditional on the random effects components): 2012/2014/2017



NOTE: The top row corresponds to proportions at or below Level 1 and the bottom row corresponds to proportions at or above Level 3. The plots on the left-hand side illustrate the relationship between the residuals and the fitted values ($X'_{ijk}\hat{\beta}_{1(2)}$). The normal quantile plots are displayed on the right-hand side.
SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

The normal quantile-quantile plots in figure 6-2 indicate slightly heavier tails, because the residuals in the second set of residuals are constructed after conditioning on the random effects. Slight heavy tails in the distributions of the residuals are not a concern of violating the model assumptions because the comparison with the quantiles of a standard normal distribution is not exactly fair: the residuals we constructed are still slightly correlated and they were constructed separately for the two quantities modeled jointly. Typically, a Student *t*-distribution may be a better fit for data showing heavy tails in the distributions of the residuals. As it will be described in section 6.1.4.3, we investigated this option and the estimated degrees of freedom of the assumed Student *t*-distribution was very large indicating a normal distribution is appropriate.

The residuals plots with the fitted values being the *x* axis allow us to investigate the relationship between the residuals and the fitted values; the scatterplots indicate no significant departure from the homogeneous variance assumption for the conditional residuals considered here (not conditional on the random effects) or for the second set of residuals (conditional on the random effects). The two sets of residuals are very similar among each other because the variances of the random effects are very small; additional checks of the model predictions and their components: fitted values and predicted random effects indicated that the model predictions are nearly equal to the fitted values.

The two sets of residuals defined above apply to proportions; the pairs reduce to one single residual for the averages. Initial residual checks for averages indicated departures from normality in one county of sample size two and hence it was removed from the set of final 184 counties to which the models are fit. Residual checks are conducted for the alternative model specifications for proportions, as well as for the various model specification for averages, and the diagnostics for the final models adopted do not indicate substantial lack of fit (most of the residuals values are between -2 and 2, and their sample quantiles are very close to the quantiles of a standard normal distributions, as indicated by the points falling close to the 45 degrees line in the residuals plots).

6.1.2.3 Posterior Predictive Checks

To assess the adequate fit of the model, we conducted posterior predictive checks. Particularly, for a set of predefined statistics, we compare statistics based on their posterior predictive distribution to their corresponding values obtained using the original sample. The procedure to generate data from the posterior predictive distribution is as follows. Consider the posterior samples for $\beta, c, d, v, \Sigma_c, \Sigma_d$, and Σ_v, denoted by $\beta^{(b)}, c_{ijk}^{(b)}, d_{ij}^{(b)}, v_i^{(b)}, \Sigma_c^{(b)}, \Sigma_d^{(b)}$, and $\Sigma_v^{(b)}$, respectively, for $b = 1, \ldots, B = 4{,}500$. Construct $\theta_{ijk}^{(b)}$ and draw replicates $P_{ijk}^{(b)}$ following the model specified in chapter 5:

$$\begin{aligned} \theta_{ijk}^{(b)} &= X'_{ijk}\beta^{(b)} + c_{ijk}^{(b)} + d_{ij}^{(b)} + v_i^{(b)}, \\ P_{ijk}^{(b)} &\sim N(\theta_{ijk}^{(b)}, \Sigma_{ijk}) \end{aligned}$$

For this application, we consider the following statistics: indicator, order indicator, mean deviation, unscaled residual constructed as deviation from the survey regression estimate, and scaled residual constructed as deviation from the survey regression estimate scaled by the posterior variance of the predictor using simulated values. For each proportion (at or below Level 1 and at or above Level 3), the corresponding posterior predictive quantities are

$$\text{(indicator) } T^{-1}\sum_{t=1}^{T} I\,[P_t^{(b)} > P_t],$$
$$\text{(order) } T^{-1}\sum_{t=1}^{T} I\,[P_{(t)}^{(b)} > P_{(t)}],$$
$$\text{(deviation) } T^{-1}\sum_{t=1}^{T}(\,P_t^{(b)} - \theta_t^{(b)}),$$
$$\text{(unscaled residual) } T^{-1}\sum_{t=1}^{T} P_t^{(b)} - P_t,$$
$$\text{(scaled residual) } ((\mathrm{T}-1)^{-1}(P_t^{(b)} - T^{-1}\sum_{t=1}^{T} P_t^{(b)})^2)^{-1}(T^{-1}\sum_{t=1}^{T} P_t^{(b)} - P_t),$$

where T is the total number of counties (184), the subscript t is used instead of $ijk, 1(2)$, for simplicity, and the subscript (t) is used to denote the order statistics of the sample of county-level proportions, of size T. Posterior predictive *p* values are constructed for the first three statistics. By definition, the posterior predictive *p* value is the proportion of summary statistics calculated with samples generated from the posterior predictive distribution that exceed the corresponding value based on the original sample. The posterior predictive *p* values corresponding to the statistics indicator, order, and mean deviation are defined as $B^{-1}\sum_{b=1}^{B} T^{-1}\sum_{t=1}^{T} I\,[P_t^{(b)} > P_t]$, $B^{-1}\sum_{b=1}^{B} T^{-1}\sum_{t=1}^{T} I\,[P_{(t)}^{(b)} > P_{(t)}]$, and $B^{-1}\sum_{b=1}^{B} I\,[T^{-1}\sum_{t=1}^{T}(\,P_t^{(b)} - \theta_t^{(b)}) \geq 0]$. A *p* value close to 0.5 indicates that the model provides a reasonable fit to the sample data; see, for example, Gelman (2013). For the unscaled and scaled residuals, we report their ranges; values of the scaled residual between -1.96 and 1.96 and a global average of around 0 may be considered to be reasonable.

The resulting posterior predictive *p* values constructed using the definitions in the previous paragraph (averages over the generated values for those statistics we constructed) are 0.474, 0.180 and 0.500, for the indicator, order and mean deviation statistics, respectively. Summary results for the posterior predictive statistics are provided in table 6-3, for literacy proportions at or below Level 1. The posterior predictive values for the ordered county-level proportions below Level 1 indicate that values generated from the posterior predictive distribution tend to be smaller than the sample values; the summaries in the last column of table 6-3 are all above zero. This result may be an effect of very small (and negative) model predictions. The posterior predictive values for the indicator test statistics are close to 0.5, the deviations and the unscaled residuals are close to zero, and the scaled residuals range is within -1.96 to 1.96. Therefore, overall there is no substantial indication for model lack of fit. The posterior predictive checks for the other quantities of interest (literacy proportions at or above Level 3, numeracy proportions, literacy and numeracy averages) do not indicate lack of fit for the models adopted.

Table 6-3. Posterior predictive checks for bivariate HB model: Summaries of posterior predictive statistics for literacy proportions at or below Level 1: 2012/2014/2017

Metric	Deviation	Unscaled residual	Scaled residual	Indicator	Order
Minimum	-0.0042	-0.3729	-1.8234	0.3478	0.0000
1st quantile	-0.0002	-0.0504	-0.4873	0.4511	0.0652
Median	0.0038	-0.0053	-0.0910	0.4728	0.1576
Mean	0.0038	-0.0157	-0.0768	0.4740	0.1801
3rd quantile	0.0078	0.0269	0.4135	0.5000	0.2622
Maximum	0.0118	0.1312	1.3535	0.6033	0.9130

SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

6.1.3 Model Sensitivity Checks

We conducted model sensitivity checks using a different specification for the prior distribution for the variance-covariance matrices, an alternative, classic, prior distribution for the variance-covariance matrices, different initial values specifications for nuisance parameters and random effects, as well as different hyperparameters specifications for the prior distribution for the variance-covariance matrices.

6.1.3.1 Changes in the Specification of the Prior Distribution for the Variance-Covariance Matrices

In chapter 5, we described the model specification with the $LKJ_{corr}(1)$ adopted for the correlation matrices Ω_c, Ω_d, and Ω_v; see Lewandowski, Kurowicka, and Joe (2009). As a sensitivity check, we now consider an alternative model specification, adopting an $LKJ_{corr_{cholesky}}(1)$ prior distribution for the Cholesky factors of the correlation matrix. For this, we decompose the correlation matrices into lower triangular matrices L_c, L_d, and L_v, where $\Omega_c = L_c L_c^T, \Omega_d = L_d L_d^T$, and $\Omega_v = L_v L_v^T$, respectively. As a consequence, the set of parameters monitored now consists of 897 parameters: $\beta, c, d, v,$ diagonal entries in S_c, S_d, and S_v, on original scale and squared, the four entries in the Cholesky factors L_c, L_d, and L_v, θ, and the log likelihood. Note that the entries above the diagonal in the Cholesky factors are equal to 0.

The results in table 6-4 indicate that convergence and mixing of the three chains have been reached. However, note that the mean and median effective sample sizes are smaller than the corresponding values for the final model fit. Therefore, based on the final model, for most of the parameters, we have a larger number of what would have constituted a set of independent draws from the posterior distribution containing the same information as our correlated 4,500 MCMC samples, providing more confidence for making inference using these sets of MCMC samples.

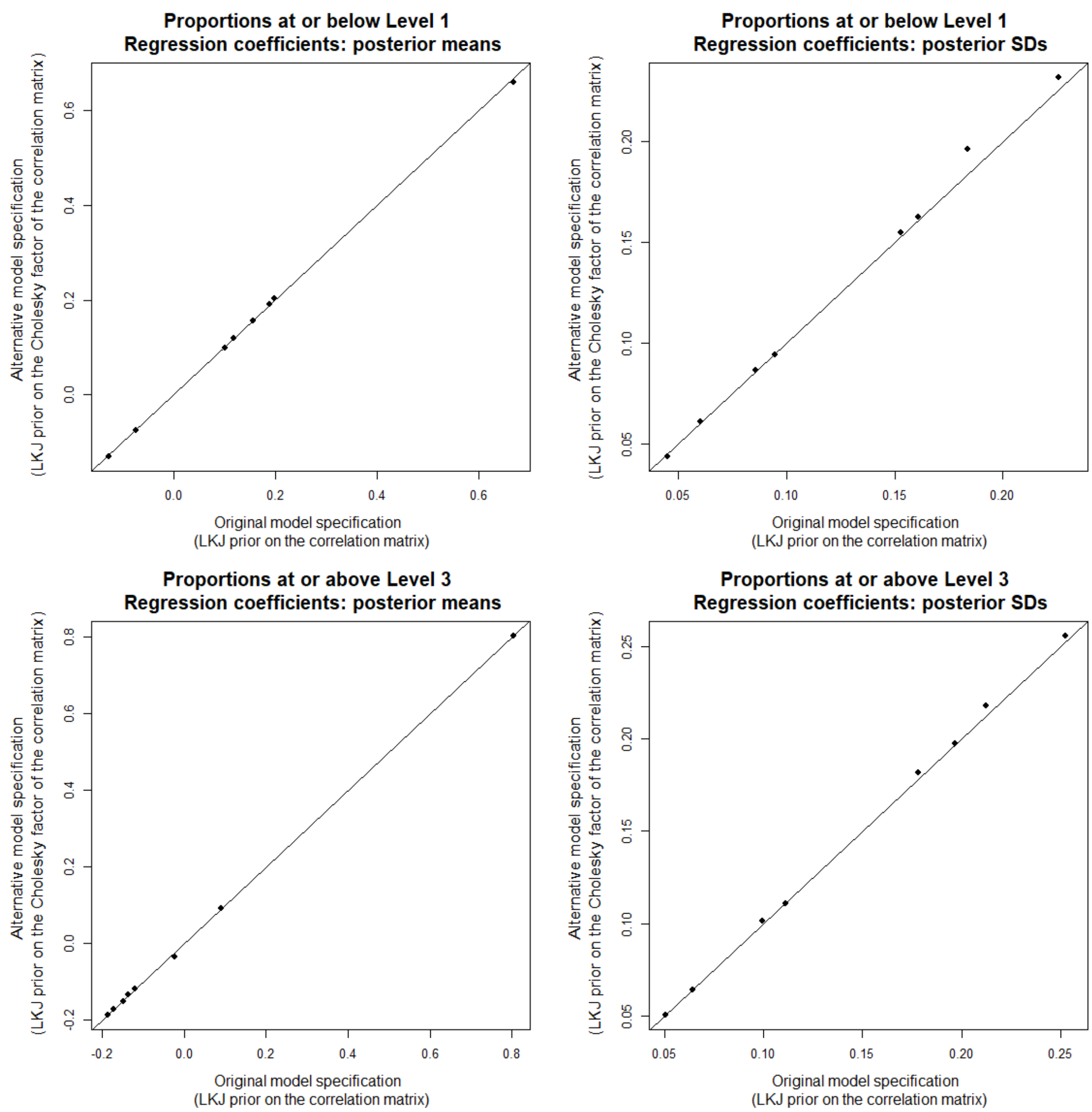
Table 6-4. Convergence diagnostics for the MCMC using alternative specification for the variance-covariance matrices (LKJ priors for the Cholesky factors of the decomposed matrices): 2012/2014/2017

Metric	$\hat{R}$	Effective sample size	MC standard error/posterior standard deviation	Auto-correlation Lag1	Auto-correlation Lag5	Cross-correlation
Minimum	0.9993	0.000	0.0000	-0.0401	-0.0498	-0.9514
1st quantile	0.9998	3,570.176	0.0151	0.0087	-0.0049	-0.0191
Median	1.0002	4,100.827	0.0157	0.0266	0.0085	-0.0003
Mean	1.0006	3,811.016	0.0171	0.0471	0.0162	0.0038
3rd quantile	1.0008	4,415.321	0.0170	0.0511	0.0240	0.0185
Maximum	1.0436	6,005.506	0.1036	0.9557	0.8105	1.0000

SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

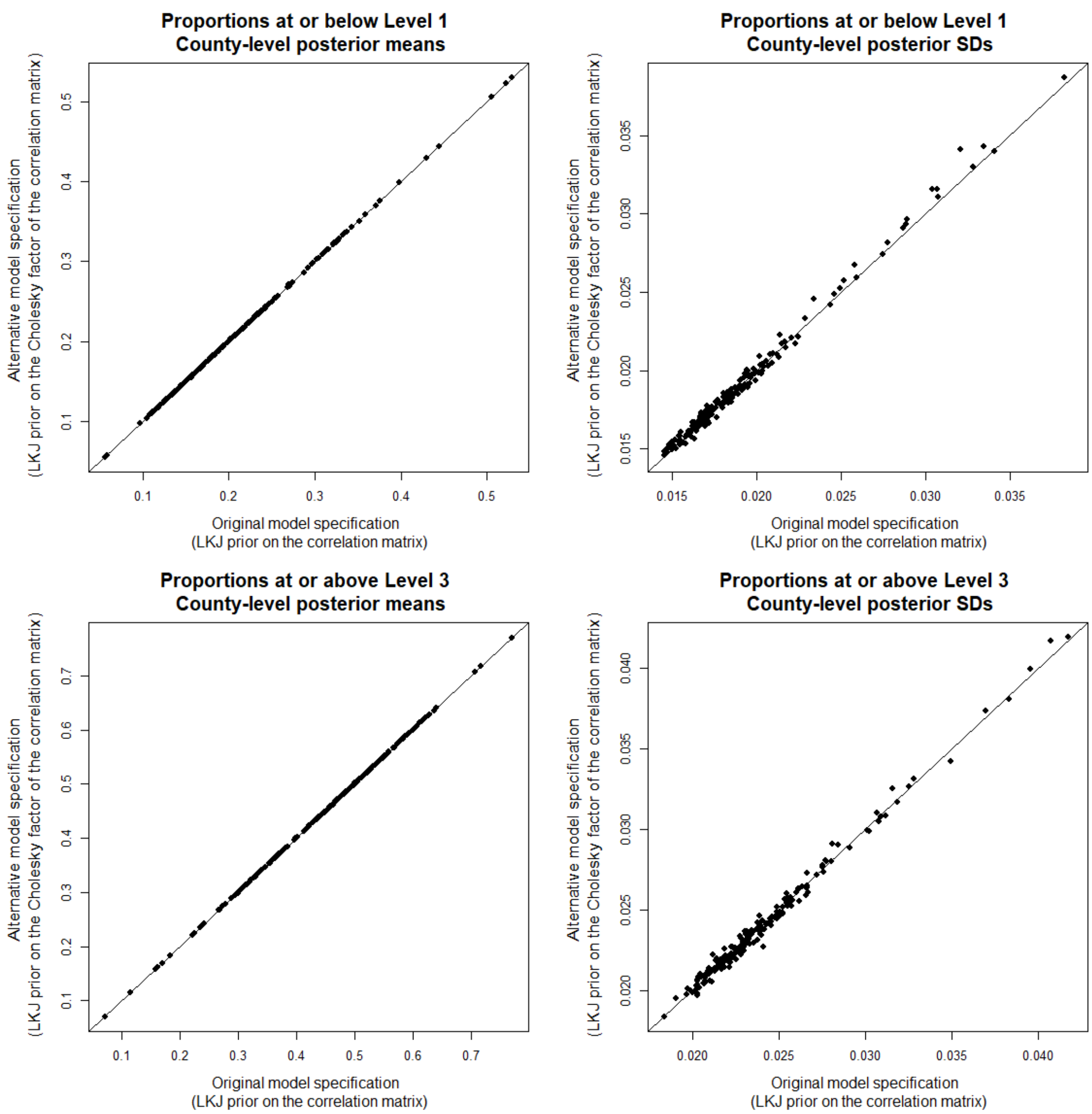
Posterior summaries for the regression coefficients and county-level literacy proportions are similar for the different model specifications. Figures 6-3 and 6-4 below facilitate these comparisons. Hence, the model specification is not sensitive to the LKJ-type prior distribution adopted for the variance-covariance matrices.

Figure 6-3. Posterior means and standard deviations for the regression coefficients under HB models with LKJ prior on the correlation matrix versus LKJ prior on the Cholesky factor of the correlation matrix: 2012/2014/2017



NOTE: The top row corresponds to proportions at or below Level 1 and the bottom row corresponds to proportions at or above Level 3. The plots on the left-hand side illustrate the relationship between the posterior means. The relationship between the posterior standard deviations are displayed on the right-hand side.
SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

Figure 6-4. Posterior means and standard deviations for county-level literacy proportions under HB models with LKJ prior on the correlation matrix versus LKJ prior on the Cholesky factor of the correlation matrix: 2012/2014/2017



NOTE: The top row corresponds to proportions at or below Level 1 and the bottom row corresponds to proportions at or above Level 3. The plots on the left-hand side illustrate the relationship between the posterior means. The relationship between the posterior standard deviations are displayed on the right-hand side.

SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

6.1.3.2 Changes in the Prior Distribution for the Variance-Covariance Matrices

The inverse-Wishart (IW) distribution has been a common choice of prior distribution adopted for the variance-covariance matrices in multivariate HB models, due to its conjugacy properties (i.e., the posterior distribution of the variance-covariance matrices is in the same probability distribution family as the prior distributions, in this case both the prior and the posterior are IW distributions). For example, Erciulescu, Berg, Cecere, and Ghosh (2019) adopted this distribution for the variance-covariance matrix of random effects in a unit-level bivariate HB model. However, with the growth in computational approaches and development of available software for fitting multivariate HB models, conjugacy is no longer needed to facilitate computations. Also, note that the IW prior implies marginal inverse-Gamma priors, while the LKJ implies marginal uniform priors on the interval [-1,1], on the correlations. For the univariate case, Gelman (2006) has a discussion on the issues related to adopting inverse-Gamma priors versus using half-Cauchy priors or uniform priors for the variance components. Generalizing to the use of IW priors versus LKJ priors, similar issues may persist; see a discussion in Alvarez, Niemi, and Simpson (2014). Nevertheless, for sensitivity analysis we consider the IW prior distribution and fit an HB model assuming the following:

$$\begin{aligned} \Sigma_c &\sim IW(2, S_c) \\ \Sigma_d &\sim IW(2, S_d) \\ \Sigma_v &\sim IW(2, S_v). \end{aligned}$$

One choice of initial value for the hyperparameters S_c, S_d, S_v is a variance-covariance matrix inspired by the estimated variances in the simulation studies (see appendix B),

$$S = \begin{pmatrix} 0.001156 & 0.000098 \\ 0.000098 & 0.001156 \end{pmatrix}.$$

For this alternative fit, we also provide initial values for the random effects parameters, again inspired by the predicted random effects in the simulation studies. Convergence and mixing diagnostics for this alternative model, with IW prior specification for the variance-covariance matrices, do not indicate lack of fit; see the results in table 6-5.

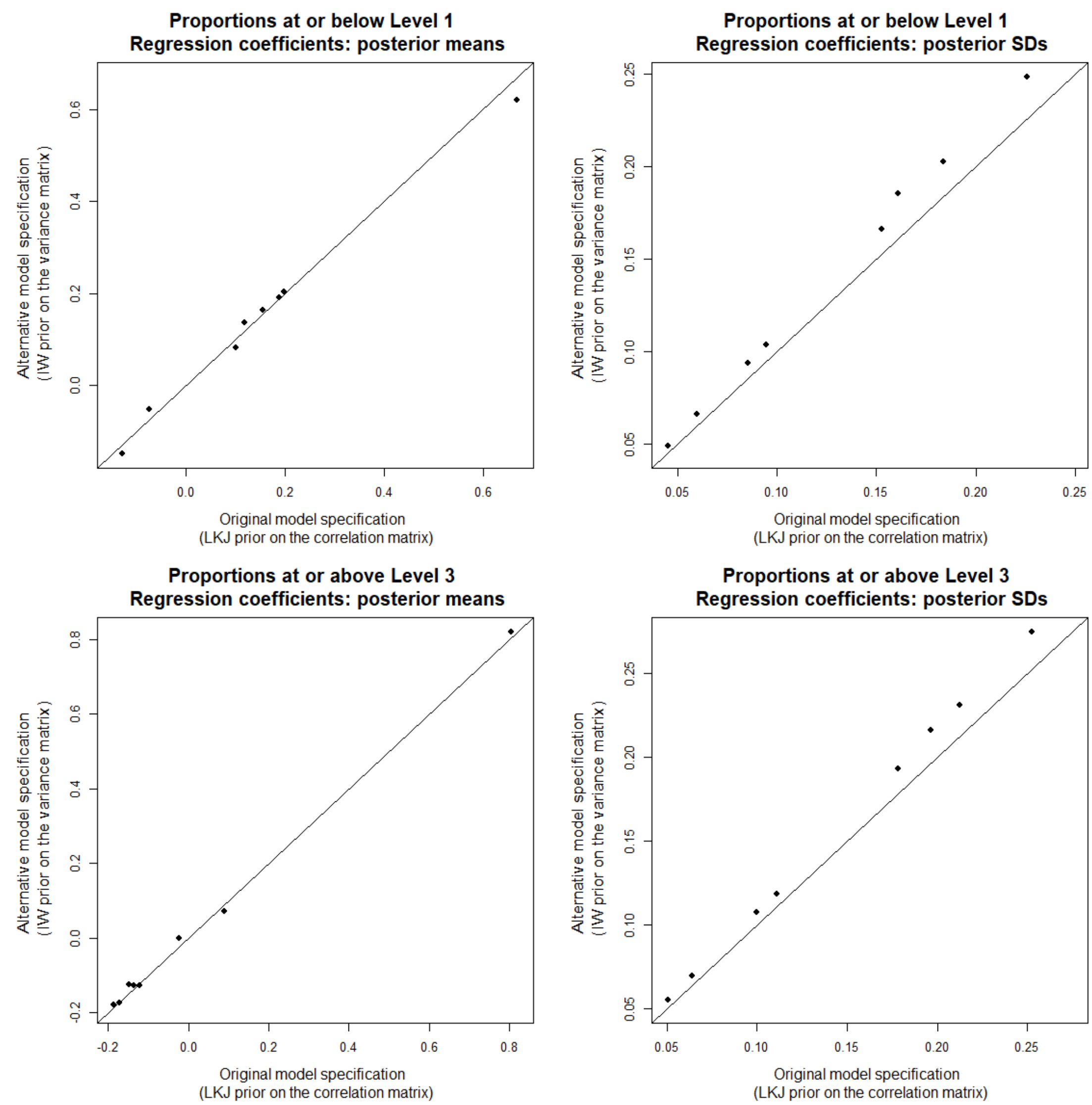
Table 6-5. Convergence diagnostics for the MCMC using alternative specification for the variance-covariance matrices (IW priors for the matrices): 2012/2014/2017

Metric	$\hat{R}$	Effective sample size	MC standard error/posterior standard deviation	Auto-correlation Lag1	Auto-correlation Lag5	Cross-correlation
Minimum	0.9994	1,536.470	0.0135	-0.0447	-0.0453	-0.9539
1st quantile	0.9997	4,346.758	0.0149	-0.0042	-0.0087	-0.0146
Median	0.9999	4,500.000	0.0149	0.0068	0.0020	0.0000
Mean	1.0000	4,437.164	0.0151	0.0110	0.0013	0.0026
3rd quantile	1.0003	4,504.472	0.0152	0.0172	0.0112	0.0144
Maximum	1.0033	5,615.795	0.0255	0.4522	0.0564	1.0000

SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

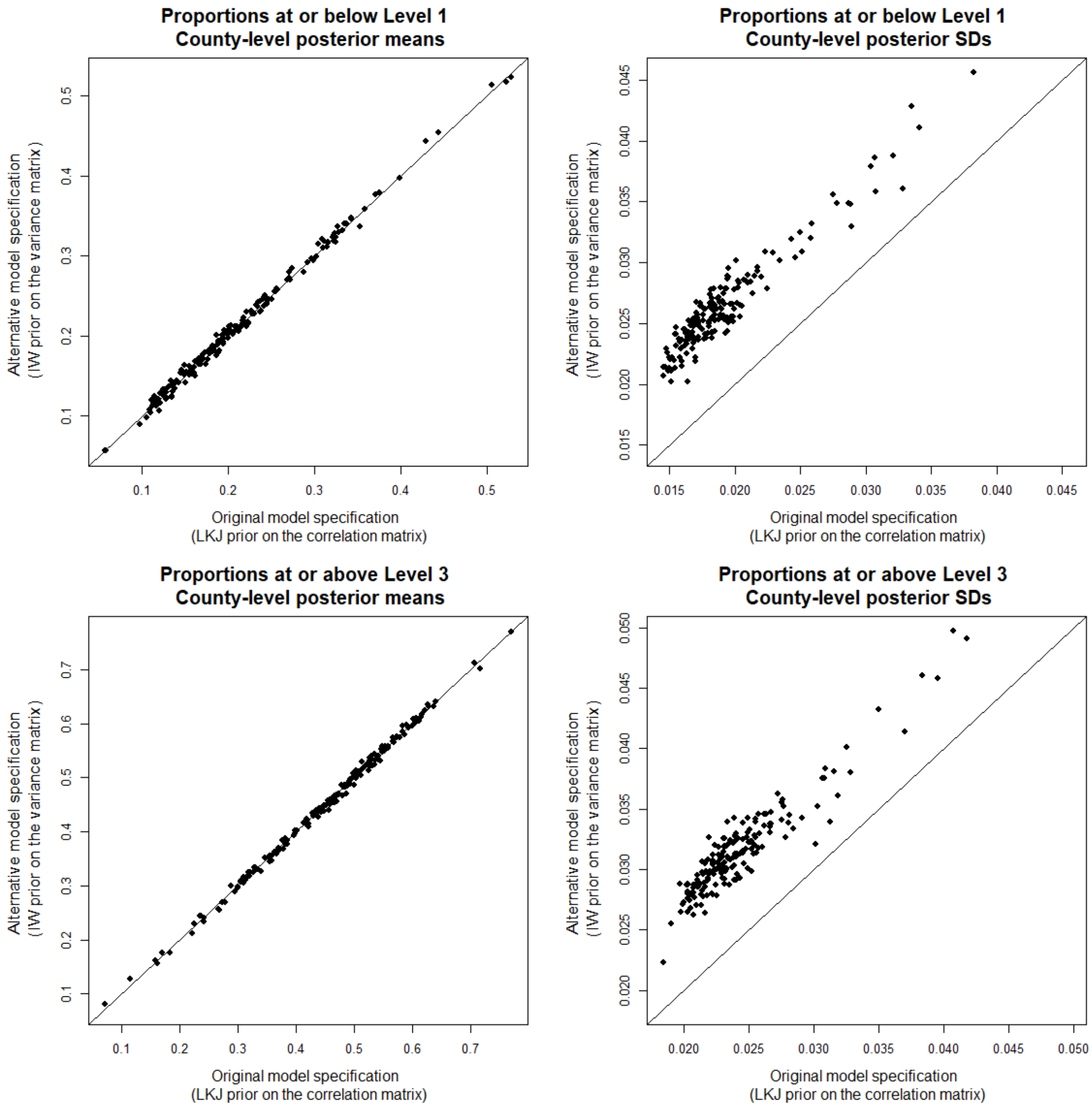
Regression coefficients posterior means and county-level literacy proportions posterior means are similar for the different model specifications. However, as expected, the posterior standard deviations of these parameters are different. Figures 6-5 and 6-6 below facilitate these comparisons. Hence, the model specification is sensitive to the prior distribution adopted for the variance-covariance matrices. For the final models, priors adopted were suggested by the hierarchical Bayes modeling literature, less sensitive to specifications of hyperparameters or initial values, and well-performing in the simulation studies conducted.

Figure 6-5. Posterior means and standard deviations for the regression coefficients under HB models with LKJ prior on the correlation matrix versus IW prior on the variance matrix: 2012/2014/2017



NOTE: The top row corresponds to proportions at or below Level 1 and the bottom row corresponds to proportions at or above Level 3. The plots on the left-hand side illustrate the relationship between the posterior means. The relationship between the posterior standard deviations are displayed on the right-hand side.
SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

Figure 6-6. Posterior means and standard deviations for county-level literacy proportions under HB models with LKJ prior on the correlation matrix versus IW prior on the variance matrix: 2012/2014/2017



NOTE: The top row corresponds to proportions at or below Level 1 and the bottom row corresponds to proportions at or above Level 3. The plots on the left-hand side illustrate the relationship between the posterior means. The relationship between the posterior standard deviations are displayed on the right-hand side.
SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

6.1.3.3 Changes in Initial Values

In the previous model fit (section 6.1.3.2), somewhat arbitrary (previously constructed using simulation results under one of the scenarios considered in appendix B) initial values were provided for the variance parameters and for the random effects parameters. As an alternative model fit, we consider the HB model with IW prior on the variance matrix and use only initial values for the coefficients β, letting the program generate initial values for the other parameters. This initial values scenario coincides with the one adopted for the final model, as presented in chapter 5.

Convergence and mixing diagnostics for this alternative model, with IW prior specification for the variance-covariance matrices and no initial values for the variance components and random effects, do not indicate lack of fit; see the results in table 6-6.

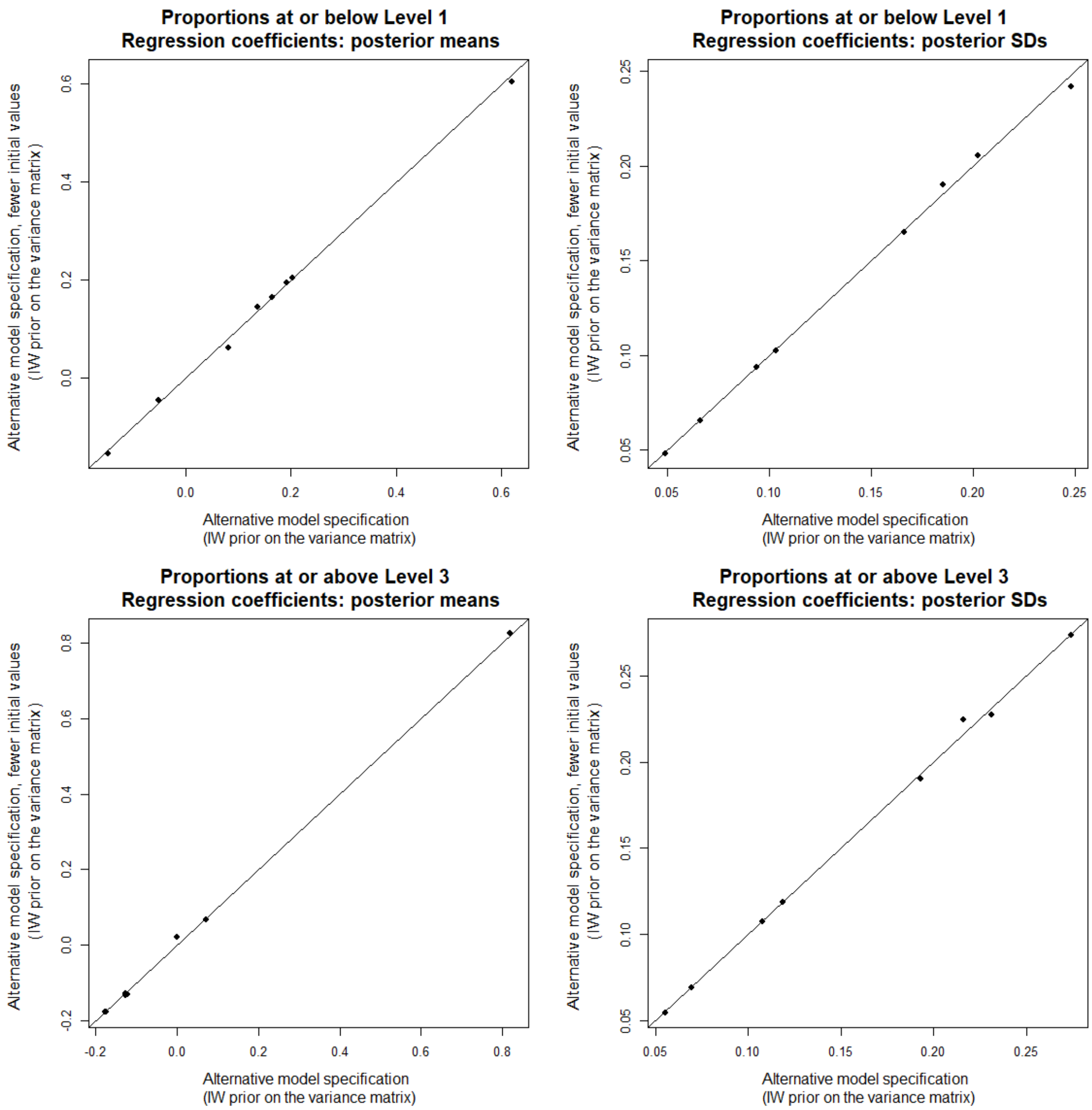
Table 6-6. Convergence diagnostics for the MCMC using alternative specification for the variance-covariance matrices and fewer initial values: 2012/2014/2017

Metric	$\hat{R}$	Effective sample size	MC standard error/posterior standard deviation	Auto-correlation Lag1	Auto-correlation Lag5	Cross-correlation
Minimum	0.9994	1,066.753	0.0135	-0.0356	-0.0449	-0.9557
1st quantile	0.9997	4,306.362	0.0149	-0.0047	-0.0123	-0.0154
Median	0.9999	4,500.000	0.0149	0.0067	-0.0018	-0.0001
Mean	1.0000	4,419.223	0.0152	0.0118	-0.0004	0.0027
3rd quantile	1.0002	4,543.911	0.0153	0.0182	0.0097	0.0150
Maximum	1.0021	5,912.353	0.0307	0.5801	0.1103	1.0000

SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

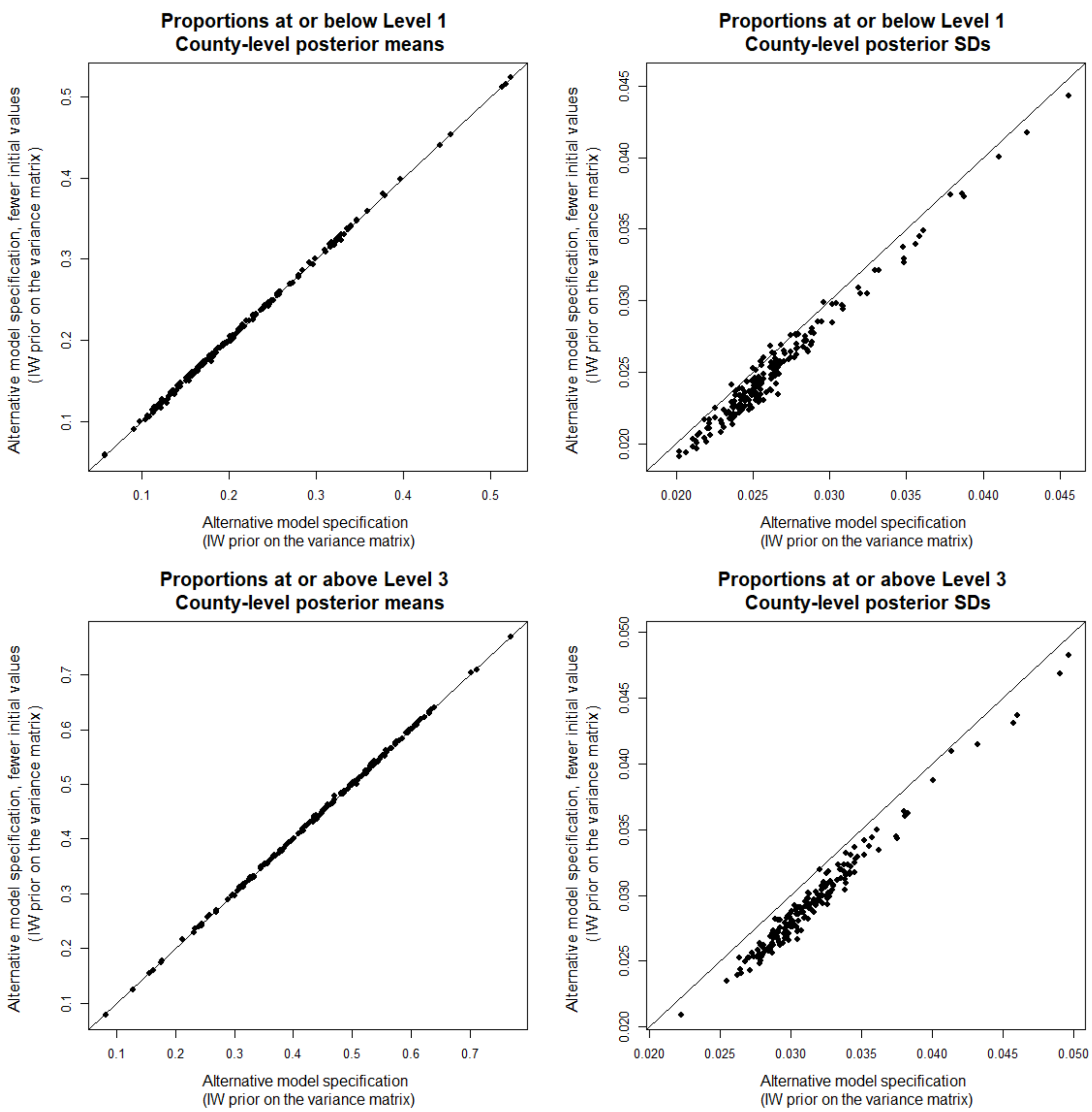
Regression coefficients posterior means and county-level literacy proportions posterior means are similar for the different model specifications. However, the posterior standard deviations of the county-level literacy proportions are different. Figures 6-7 and 6-8 below facilitate these comparisons. Hence, the model specification with IW prior adopted for the variance matrix is also sensitive to the set of initial values provided to the model parameters.

Figure 6-7. Posterior means and standard deviations for the regression coefficients under HB models with IW prior on the variance matrix and different sets of initial values: 2012/2014/2017



NOTE: The top row corresponds to proportions at or below Level 1 and the bottom row corresponds to proportions at or above Level 3. The plots on the left-hand side illustrate the relationship between the posterior means. The relationship between the posterior standard deviations are displayed on the right-hand side.
SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

Figure 6-8. Posterior means and standard deviations for county-level literacy proportions under HB models with IW prior on the variance matrix and different sets of initial values: 2012/2014/2017



NOTE: The top row corresponds to proportions at or below Level 1 and the bottom row corresponds to proportions at or above Level 3. The plots on the left-hand side illustrate the relationship between the posterior means. The relationship between the posterior standard deviations are displayed on the right-hand side.

SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

6.1.3.4 Changes in Hyperparameters Values

In the previous model fit (sections 6.1.3.2 and 6.1.3.3), somewhat arbitrary (previously constructed using simulation results under one of the scenarios considered in appendix B) hyperparameters were provided for the priors on the variance parameters. As an alternative model fit, we consider the HB model with IW prior on the variance matrix, using only initial values for the coefficients β, letting the program generate initial values for the other parameters, and we replace the hyperparameters S_c, S_d, S_v by S,

$$S = 10^{-3}\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix} = \begin{pmatrix}0.001 & 0\\ 0 & 0.001\end{pmatrix}.$$

Convergence and mixing diagnostics for this alternative model, with IW prior specification for the variance-covariance matrices, no initial values for the variance components and random effects, and noninformative choice of hyperparameters for the IW priors, do not indicate lack of fit; see the results in table 6-7.

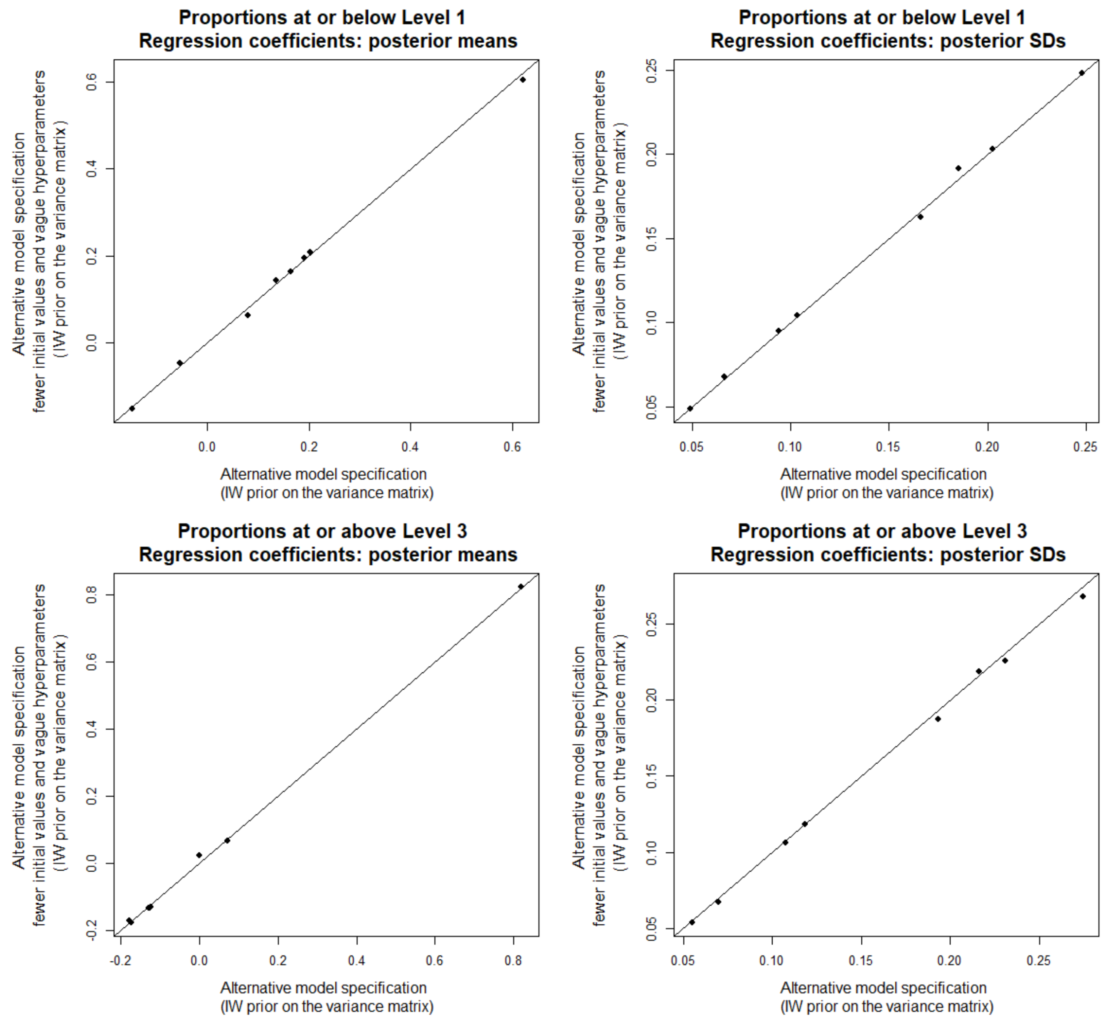
Table 6-7. Convergence diagnostics for the MCMC using alternative specification for the variance-covariance matrices and fewer initial values and noninformative choice of hyperparameters: 2012/2014/2017

Metric	$\hat{R}$	Effective sample size	MC standard error/posterior standard deviation	Auto-correlation Lag1	Auto-correlation Lag5	Cross-correlation
Minimum	0.9994	1,110.672	0.0138	-0.0380	-0.0398	-0.9559
1st quantile	0.9997	4,332.085	0.0149	-0.0069	-0.0113	-0.0152
Median	0.9999	4,500.000	0.0149	0.0052	-0.0010	-0.0001
Mean	1.0001	4,422.562	0.0152	0.0103	-0.0009	0.0027
3rd quantile	1.0003	4,540.658	0.0152	0.0170	0.0088	0.0148
Maximum	1.0032	5,551.681	0.0301	0.5748	0.0923	1.0000

SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

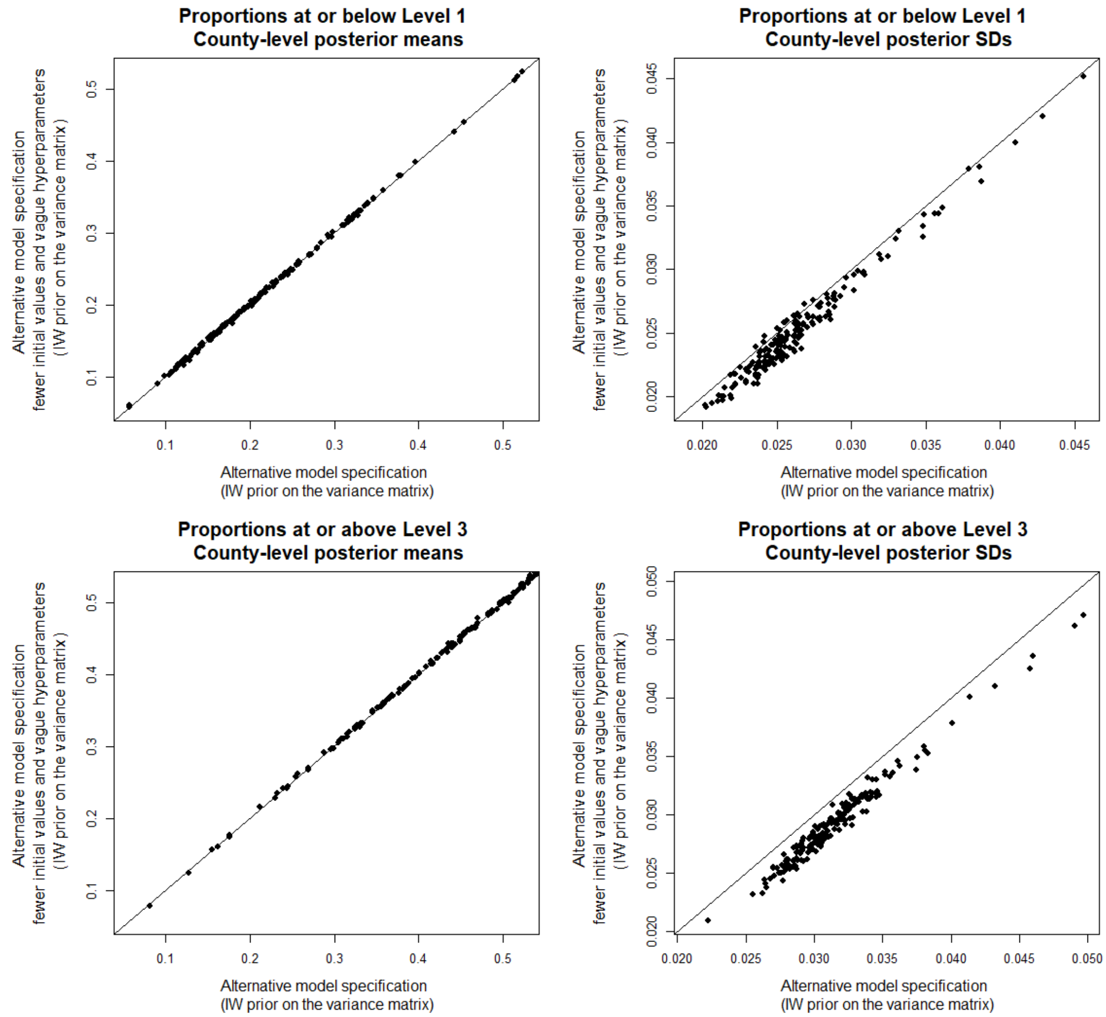
Regression coefficients posterior means and county-level literacy proportions posterior means are similar for the different model specifications. However, the posterior standard deviations of the county-level literacy proportions are different. Figures 6-9 and 6-10 below facilitate these comparisons. Hence, the model specification with IW prior adopted for the variance matrix is also sensitive to the set of hyperparameter values provided to the model parameters.

Figure 6-9. Posterior means and standard deviations for the regression coefficients under HB models with IW prior on the variance matrix and different sets of hyperparameters: 2012/2014/2017



NOTE: The top row corresponds to proportions at or below Level 1 and the bottom row corresponds to proportions at or above Level 3. The plots on the left-hand side illustrate the relationship between the posterior means. The relationship between the posterior standard deviations are displayed on the right-hand side.
SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

Figure 6-10. Posterior means and standard deviations for county-level literacy proportions under HB models with IW prior on the variance matrix and different sets of hyperparameters: 2012/2014/2017



NOTE: The top row corresponds to proportions at or below Level 1 and the bottom row corresponds to proportions at or above Level 3. The plots on the left-hand side illustrate the relationship between the posterior means. The relationship between the posterior standard deviations are displayed on the right-hand side.
SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

The methods in section 6.1.3 apply to the bivariate models for the literacy proportions. As illustrated in subsection 6.1.3.2, the model that uses IW distribution as the prior distribution for the random effects variance-covariance matrices is sensitive to changes in initial values provided to the model parameters and to the hyperparameters provided to the model parameters, so it was not adopted as the prior distribution for the final model. The prior distribution adopted for the final model random effects variance-covariance matrices is the LKJ, as suggested in STAN's guide and reference manual, and recommended in recent literature studies, i.e., Alvarez, Niemi, and Simpson (2014). Similar methods were applied, and comparable results were obtained for the numeracy proportions and for the univariate models for the averages. Because the models for averages are univariate, models using uniform and inverse-gamma prior distributions for the variance parameters are compared using different choices of initial values. Similar conclusions to the ones presented in this section hold for numeracy model evaluation. From the model evaluation for averages, it was observed that the models are not sensitive to the initial values, but are sensitive to the choice of prior distribution (leading to different posterior variances for the model parameters). Following the discussion in Gelman (2006) and the default choice in STAN, we adopted uniform priors on the random effects variances in the models for averages, on a wide range, 0 to 1,000, as shown to have better performance when compared to the classical inverse-Gamma priors for the variances, and because the number of counties we are modeling is larger than 3, the model would not be sensitive to the choice as upper bound on the uniform distribution range, in this case, 1,000. Alternative choices include half-Cauchy priors, again suggested in Gelman (2006) and STAN's guide and reference manual, and matching priors, as adopted, for example, by Datta, Rao, and Smith (2005) and Nandram, Erciulescu, and Cruze (2019). In the simulation studies we include results on the performance of the posterior variance under different prior choices, with respect to tracking the design MSE.

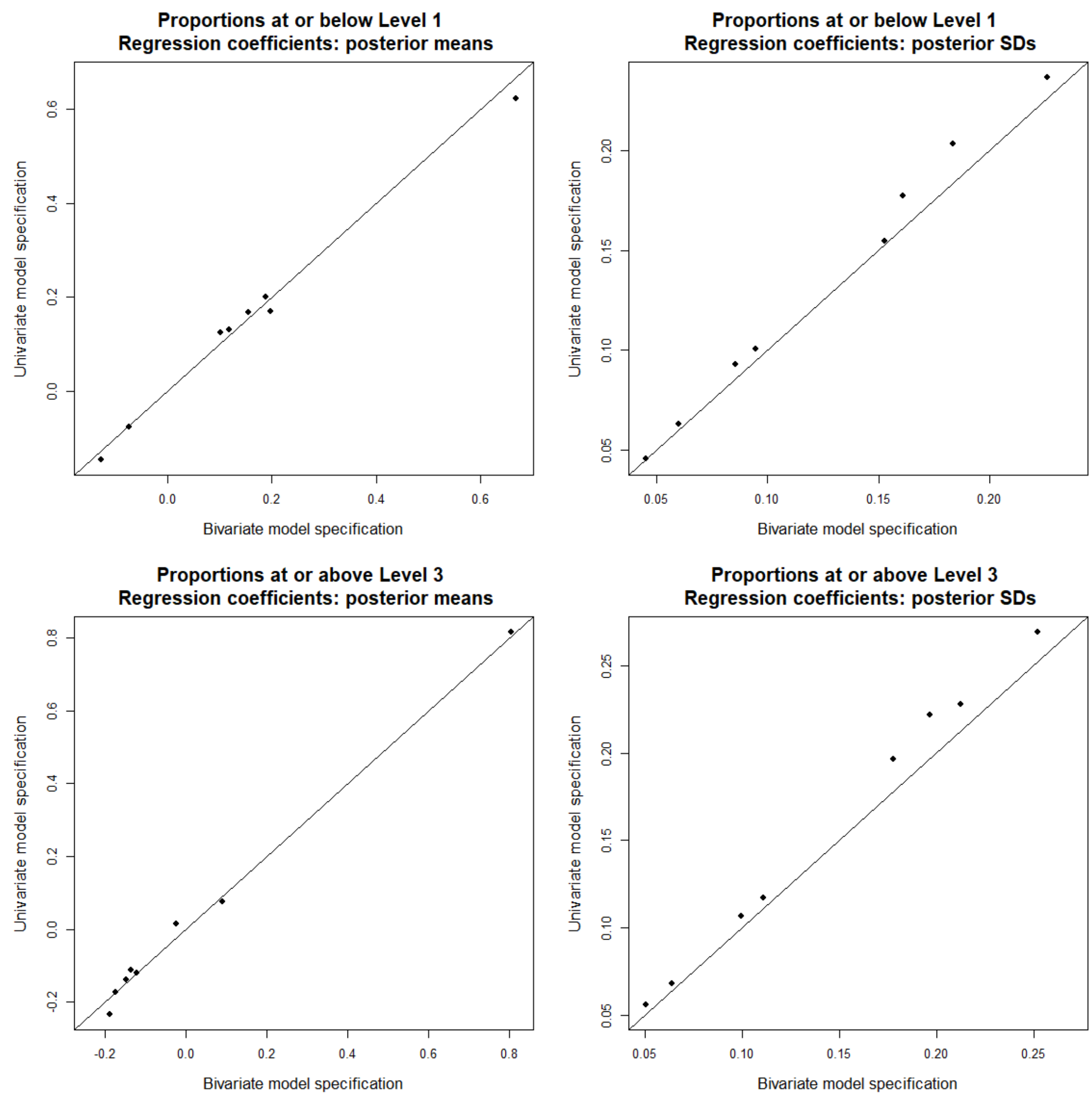
6.1.4 Changes in the Model Specification

We illustrate additional results, based on changes in the model specification. For this, univariate models were compared against the bivariate models adopted for the two proportions, different sets of software (STAN) parameters were used to fit the HB models, an alternative distribution (to the normal distribution) was assumed for the sampling model and an alternative link function (to the linear) was assumed for the linkage model.

6.1.4.1 Univariate Versus Bivariate Models for Literacy Proportions

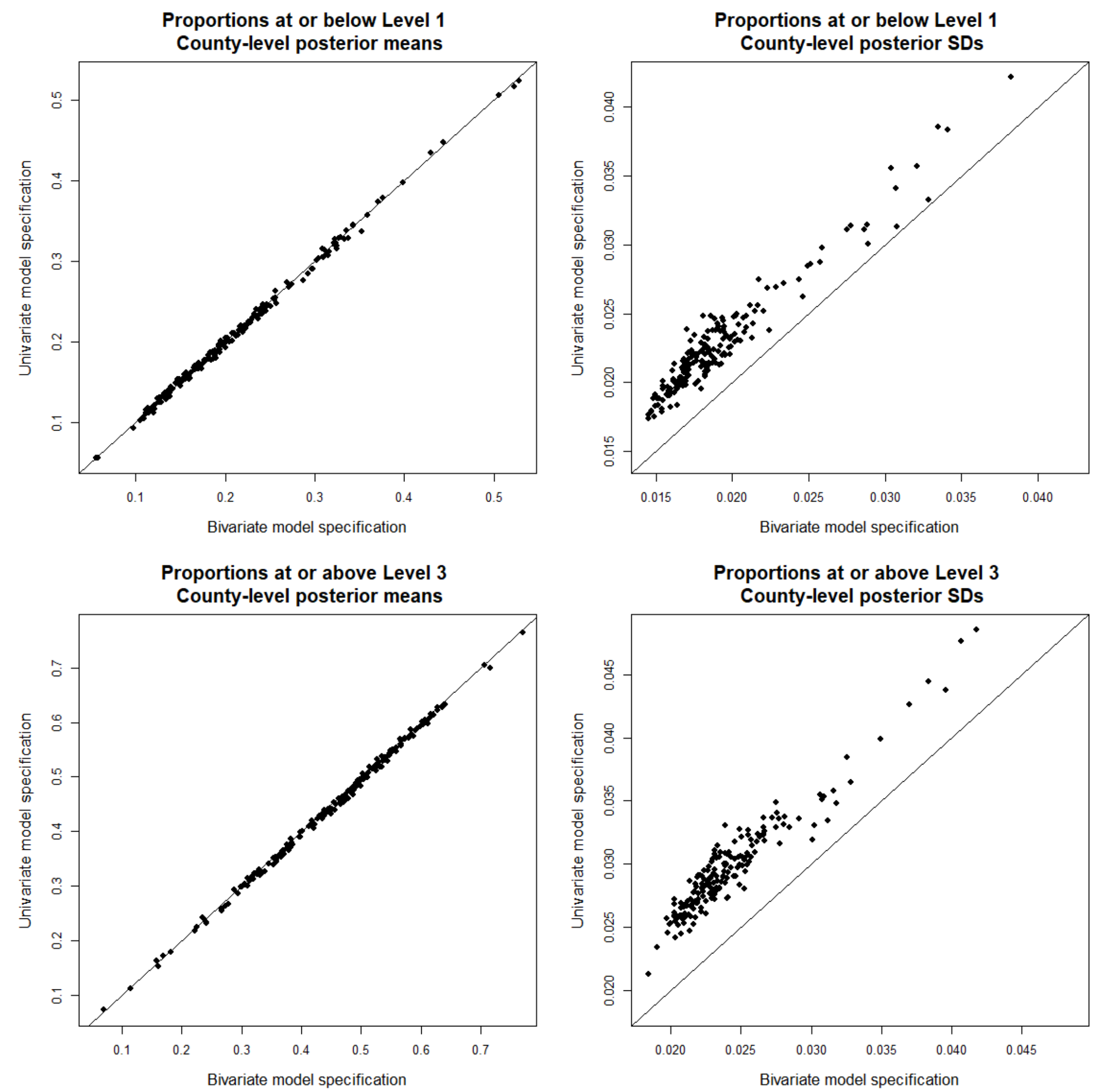
We modeled the proportions using univariate models with uniform priors on the variances, on a wide range, 0 to 1,000, and compared against the bivariate model with LKJ on the correlation matrix (specification one in the report) and half-Cauchy on the standard deviations. Figures 6-11 and 6-12 illustrate the results for these comparisons for regression coefficients and for county-level literacy proportions under univariate and bivariate HB models, respectively. As expected, the posterior variances are reduced when the proportions are modeled using a bivariate model. However, more work could be done to assess whether the two modeling approaches were reasonable fits to the data and that the corresponding MCMC chains converged and mixed properly. Also, additional comparisons may be conducted using univariate models with inverse-gamma priors on the variances and compared against the bivariate model with IW prior on the variance-covariance matrix, as well as univariate models with half-Cauchy priors on the variances and compared against the bivariate model with LKJ-type priors.

Figure 6-11. Posterior means and standard deviations for regression coefficients under univariate and bivariate HB models: 2012/2014/2017



NOTE: The top row corresponds to proportions at or below Level 1 and the bottom row corresponds to proportions at or above Level 3. The plots on the left-hand side illustrate the relationship between the posterior means. The relationship between the posterior standard deviations are displayed on the right-hand side.
SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

Figure 6-12. Posterior means and standard deviations for county-level literacy proportions under univariate and bivariate HB models: 2012/2014/2017



NOTE: The top row corresponds to proportions at or below Level 1 and the bottom row corresponds to proportions at or above Level 3. The plots on the left-hand side illustrate the relationship between the posterior means. The relationship between the posterior standard deviations are displayed on the right-hand side.
SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

6.1.4.2 Tuning Parameters in the Hamiltonian Monte Carlo and No-U-Turn Samplers Algorithms

As a consequence of internal model validation results, the length of the MC chains used for the final bivariate models, for proportions, was increased to 30,000 (after dropping the first 5,000 samples), and the parameters 'max_treedepth' and 'adapt_delta' of the Hamiltonian Monte Carlo (HMC) and No-U-Turn Samplers (NUTS) algorithms, were increased above the default software values, to 13 and 0.99, respectively. The final univariate models, for averages, were fit using 20,000 MC samples (after dropping the first 5,000 samples) and the default HMC/NUTS parameters, 'max_treedepth' equal to 10 and 'adapt_delta' equal to 0.8.

For example, table 6-8 shows the mean and quartiles of the diagnostic statistics for the fit of the final bivariate model for literacy proportions, using only 20,000 samples (after dropping the first 5,000) and the default parameters 'max_treedepth' and 'adapt_delta.' Note the smaller effective sample sizes and larger associated autocorrelations, when compared to the results in table 6-1

Table 6-8. Convergence diagnostics for the MCMC, using a smaller number of MC samples and default software parameters for the sampling algorithms: 2012/2014/2017

Metric	$\hat{R}$	Effective sample size	MC standard error/posterior standard deviation	Auto-correlation Lag1	Auto-correlation Lag5	Cross-correlation
Minimum	0.9993	0.000	0.0000	-0.0201	-0.0213	-0.9547
1st quantile	1.0013	2,071.948	0.0243	0.0736	0.0441	-0.0424
Median	1.0044	2,549.760	0.0352	0.1102	0.0771	0.0001
Mean	1.0085	2,476.899	0.0413	0.1283	0.0905	0.0039
3rd quantile	1.0114	2,933.286	0.0532	0.1567	0.1190	0.0438
Maximum	1.1168	8,146.805	0.1560	0.9638	0.8475	1.0000

SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

6.1.4.3 Relaxed Normality Assumptions in the Bivariate HB Models for Literacy Proportions

Alternative distributional and relational assumptions were made: for the sampling model a Student *t*-distribution is investigated, and for the linking model a logit link function is investigated. Unlike the specification used for the final model, where normal distribution is assumed at both the sampling level and the linking level (hence a matched model), these alternative specifications are referred to unmatched models. The results for the unmatched model using Student *t*-distribution at the sampling level are similar to the results based on the matched normal-normal model. The results are as expected, given that the posterior means for the degrees of freedom parameter for the Student *t*-distribution is very large (approximately 50,000). The second unmatched model using logit link resulted in different model predictions, however, and model validation was left for future development. However, this is not a major concern because the models proposed were developed as the result of substantial simulation studies (presented in appendix B) and iterative steps with extensive internal validation checks (presented in section 6.1). In addition, the model estimates are further validated using external checks presented in the next section, 6.2.

6.2 External Model Validation

External checks generally consist of comparing aggregations of model estimates to direct estimates for larger geographic areas for which reliable direct estimates are available, or to external control totals from other surveys or from administrative data. The external validation is illustrated by various graphs in section 6.2.1 and a table that compares the model and direct estimates in section 6.2.2.

6.2.1 Model Validation Graphs

The model validation conducted in this section is mainly graphical, as influenced by Khan et al. (2018). In general, the following graphs were conducted for the purpose of evaluation.

- Histograms of differences between survey regression estimates and indirect estimates. The main objective of this plot is to take a first look at the results through reviewing the distribution of the differences. The graph can also indicate outliers that would need more investigation, especially to check the sample sizes in those small areas.

- Bubble plots of survey regression estimates versus indirect estimates, with the sizes of bubbles being related to the sample sizes. One would expect to see from this plot that the large bubbles are close to the diagonal line, given that the direct estimates based on large sample sizes are fairly precise. The outliers would more likely be denoted by small bubbles indicating small sample sizes.

- Shrinkage plots with arrows showing the direction from survey regression estimates to indirect estimates, by sample size. The main purpose of this plot is to show how the model impacts the estimates. There should be some shrinkage; that is, the predictions are pulled toward the average, if the predictions are more dependent on the model than the direct estimates. The longer arrows show larger impact from the model, which should occur for areas with smaller sample sizes than others.

- Interval coverage plots, showing the credible interval of the survey regression and indirect estimates, by sample size. These plots help show whether or not the resulting credible interval from the model covers the survey regression estimate. Again, the focus is solely on the areas with the largest samples. If the interval does not cover the survey regression estimates in the areas with largest sample sizes, it may indicate that the modeling process can be improved.

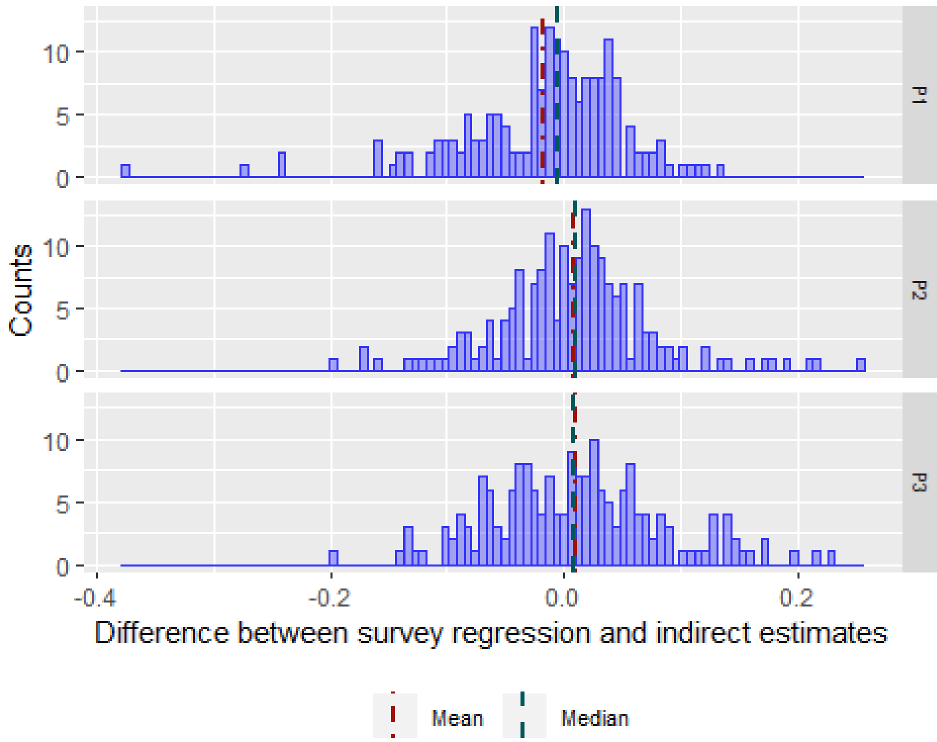
- Variance plots showing the resulting smoothed variances and model variances, by sample size. While the aforementioned graphs are used to review the point estimates, this plot shows the impact on precision. This review is a common aspect of SAE evaluations, and as an example, Bijlsma et al. (2017) also reviewed the decrease in standard errors from the SAE approach. Erciulescu, Cruze, and Nandram (2017) reviewed the outcomes to ensure (1) the negative correlation between coefficient of variance (CV) and sample size, and (2) the impact of the SAE approach on the CV. Mohadjer et al. (2011) reviewed the credible interval widths and CVs (direct estimates and model predictions) for sampled counties and nonsampled counties to ensure the impact of direct estimates on the model predictions' variances for counties with sample. In general, the posterior variances associated with the model predictions should be smaller than the smoothed sampling variances associated with survey regression estimates. If not, it may be due to weak covariates used in the models.

The evaluation results discussed below are based on the model predictions for literacy proportions. Other results from the evaluation on the models for literacy averages, and numeracy proportions and averages are available in appendix C.

6.2.1.1 Histograms of Differences

The differences between survey regression estimates and indirect estimates are shown in the histograms in figure 6-13 for literacy proportions. The means and medians of the differences are around zero. The majority of the differences are within 20 percentage points. The outliers in the plots show that a few model predictions deviate from the survey regression estimates by about 20–40 percentage points. These large deviations are not a concern because they are mostly observed for counties with small sample sizes, for which the direct survey estimates are less reliable than the corresponding estimates for counties with large sample sizes.

Figure 6-13. Literacy proportion—Histograms of differences between survey regression estimates and indirect estimates: 2012/2014/2017

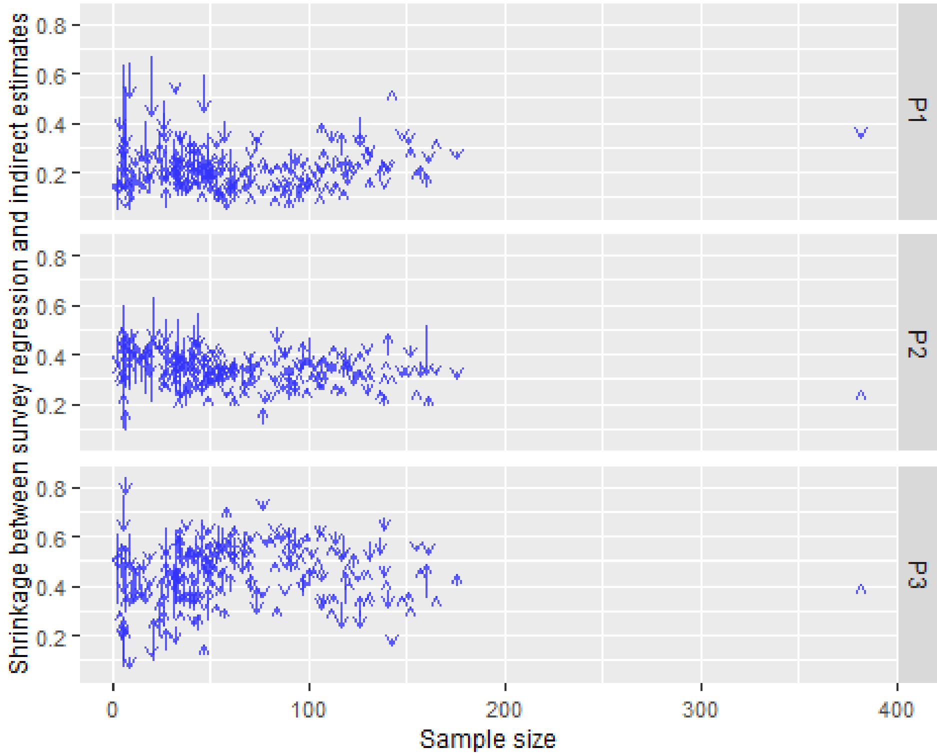


SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

6.2.1.2 Shrinkage Plots

Shrinkage towards the means can be observed in figure 6-14. As expected, the shrinkages are more significant in areas with smaller sample sizes than those in areas with larger sample sizes. The model predictions and the survey regression estimates become much more similar when the sample sizes are above 100. One county with sample size around 160 shows larger shrinkage in proportions at Level 2 and proportion at or above Level 3.

Figure 6-14. Literacy proportion—Shrinkage plots of point estimates, by sample size: 2012/2014/2017

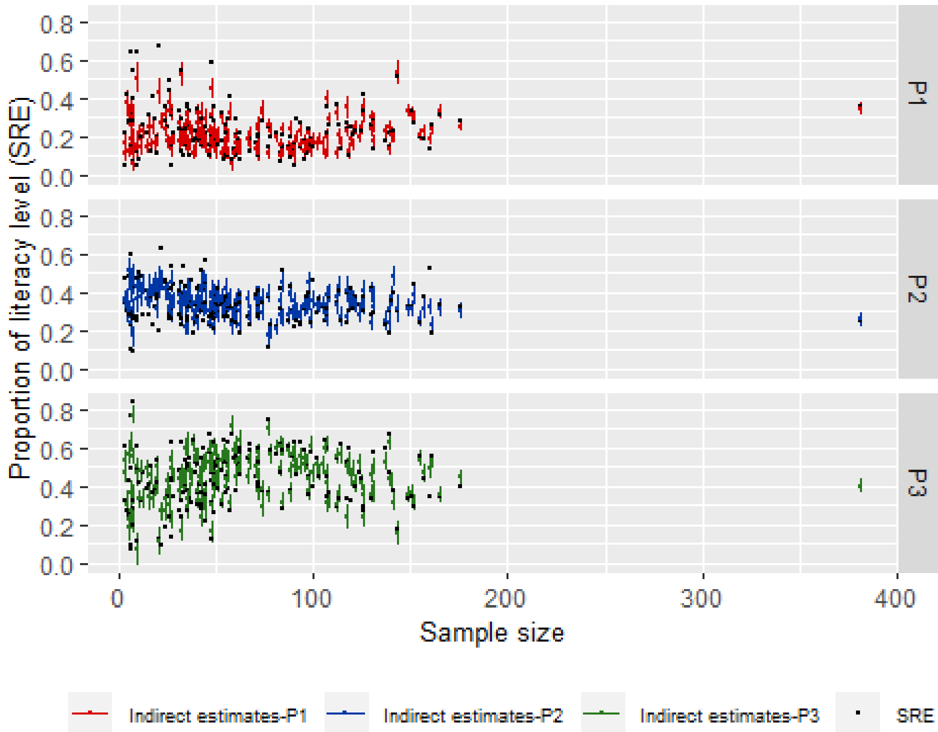


SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

6.2.1.3 Interval Coverage Plots

The interval coverage plots in figure 6-15 show that, for a majority of the small areas, the credible intervals generated from the models cover the survey regression estimates, especially for areas with large sample sizes. When the sample sizes are less than 50, sometimes the credible intervals from the models do not cover the survey regression estimates. This is as expected because the survey regression estimates contribute less to the indirect estimates if the survey regression estimates are derived from samples of smaller sizes (i.e., less reliable).

Figure 6-15. Literacy proportion—Indication of coverage by credible interval: 2012/2014/2017

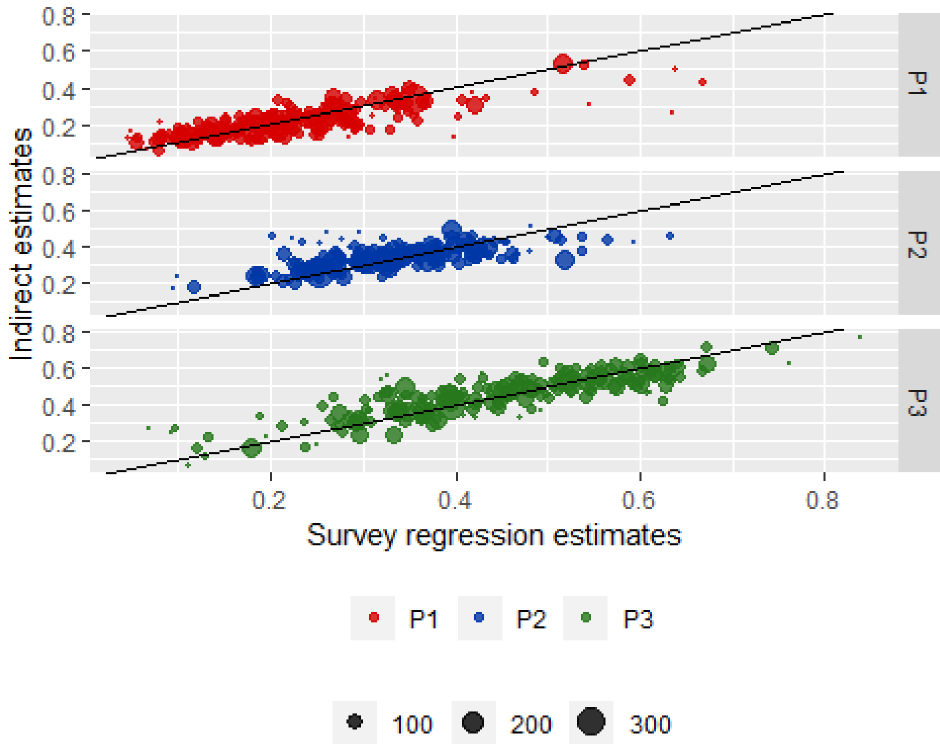


SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

6.2.1.4 Bubble Plots of Survey Regression Estimates and Indirect Estimates

Figure 6-16 shows the survey regression estimates versus indirect estimates. The majority of the points are around the 45-degree line, indicating that the model predictions are close to the survey regression estimates. Larger bubbles (i.e., counties with larger sample sizes) have closer estimates than smaller bubbles. Some of the small bubbles, with the sizes of bubbles being proportional to the sample sizes in the small areas, are farther away from the 45-degree lines. Similar as above, this is as expected due to higher sampling errors for the survey regression estimates constructed using samples with small sizes. The proportion at or below Level 1 (P1) and proportion Level 3 and above (P3) have closer estimates than proportion at Level 2 (P2), which could be a result of using P1 and P3 in the model fit and estimation. Similar results are shown for each outcome in appendix C.

Figure 6-16. Literacy proportion—Comparison between survey regression estimates and indirect estimates: 2012/2014/2017

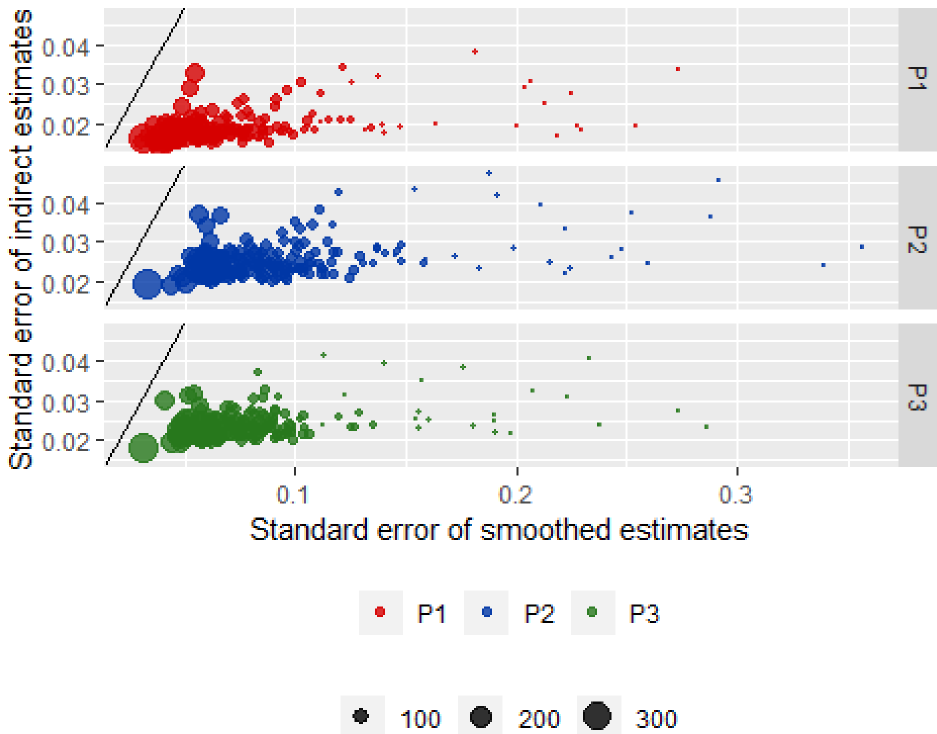


SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

6.2.1.5 Smoothed and Small Area Model Variances

Figure 6-17 shows the smoothed standard errors of the survey regression estimates and the posterior standard deviations from the small area model. For these plots, while in general the resulting model predictions are similar to the survey regression estimates, keep in mind that the standard errors of proportions depend on the sizes of the estimated proportions. Therefore if the model proportion is different from the survey regression proportion, the variance will in theory be different. The plot shows that the model produces smaller posterior standard deviations than the smoothed standard errors, especially for areas of very small sample sizes.

Figure 6-17. Literacy proportion—Comparison between model standard errors and smoothed standard errors: 2012/2014/2017



SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

6.2.2 Comparison of Aggregates of Model Predictions and Direct Estimates

Another method for the external model validation is to compare the model predictions with the corresponding direct estimates (the direct estimates referred in this section is the actual direct estimates, not the survey regression estimates) at various aggregate geographical levels for which the direct estimates are reasonably reliable.

The county-level model predictions were aggregated to a number of domains using county-level characteristics, including various three-level categorizations of county variables relating to variables used in the model: race/ethnicity, education attainment, poverty, as well as covariates not used in the model, such as Beale codes (Rural-Urban Continuum Codes), population size, language, foreign-born status, and mortality rate. The direct estimates were computed from the unit-level sample data for the same domains. The direct estimates and model predictions of the literacy/numeracy proficiency outcomes are comparable, and table 6-9 below shows the literacy proportion at or below Level 1. Similar tables for other outcomes are presented in appendix C. In general, the relative differences are within 10 percent. The side-by-side comparisons in table 6-9 should be viewed with caution because the direct estimates cannot be thought of as the truth, rather the comparison is the difference between two estimates. The PIAAC sample was not designed for estimates within classes of counties, so the representativeness is in question. Table 5-4 provides the comparison at the national level.

Table 6-9. Comparison of aggregated county-level indirect and direct estimates for Literacy P1, by subgroup: 2012/2014/2017

Subgroup (source: Year)	Small area estimate: Number of counties	Small area estimate: Weighted estimate	Direct estimate: Sample size	Direct estimate: Estimate	Direct estimate: Standard error	Percentage point difference	Relative difference (percent)
Variables used in the model:							
Percentage of population age 25+: with education less than high school (ACS: 2013–2017)							
<9.28	862	13.8	4,099	14.6	0.91	-0.8	-5.4
9.28–13.82	938	20.5	4,168	22.1	0.89	-1.7	-7.5
≥13.82	1,342	30.5	4,063	31.6	1.36	-1.1	-3.5
Percentage of population age 25+: with education more than high school (ACS: 2013–2017)							
<56.57	2,107	26.8	4,079	29.8	1.37	-3.0	-9.9
56.57–64.27	652	22.2	4,097	21.4	1.17	0.8	3.8
≥64.27	383	16.5	4,154	17.7	1.14	-1.2	-6.8
Percentage of population below 100 percent poverty line (ACS: 2013–2017)							
<11.99	919	15.5	4,102	16.1	0.98	-0.7	-4.1
11.99–16.78	984	21.3	4,097	22.9	1.05	-1.6	-7.0
≥16.78	1,239	28.7	4,131	29.6	1.31	-0.8	-2.8
Percentage of Blacks (ACS: 2013–2017)							
<4.35	1,937	19.2	4,042	20.9	1.23	-1.7	-8.2
4.35–12.91	523	21.3	4,154	21.6	1.04	-0.3	-1.5
≥12.91	682	24.6	4,134	25.7	1.28	-1.1	-4.3
Percentage of Hispanics (ACS: 2013–2017)							
<5.16	1,860	18.5	4,093	19.1	1.19	-0.6	-3.0
5.16–16.27	810	18.3	4,117	19.3	1.04	-1.0	-5.4
≥16.27	472	27.0	4,120	28.4	1.18	-1.5	-5.2
Percentage of civilian noninstitutionalized population with no health insurance coverage (ACS: 2013–2017)							
<8.34	1,057	16.2	4,091	16.6	0.93	-0.4	-2.5
8.34–12.32	1,001	22.7	4,131	23.8	0.99	-1.1	-4.6
≥12.32	1,084	27.8	4,108	27.6	1.40	0.2	0.9
Percentage of population age 16+: service occupation (ACS: 2013–2017)							
<16.37	1,054	16.7	4,025	18.1	0.92	-1.4	-7.9
16.37–18.60	901	21.5	4,150	21.1	1.10	0.4	1.9
≥18.60	1,187	26.2	4,155	28.0	1.25	-1.8	-6.4

See notes at end of table.

Table 6-9. Comparison of aggregated county-level indirect and direct estimates for Literacy P1, by subgroup: 2012/2014/2017—Continued

Subgroup (source: Year)	Small area estimate: Number of counties	Small area estimate: Weighted estimate	Direct estimate: Sample size	Direct estimate: Estimate	Direct estimate: Standard error	Percentage point difference	Relative difference (percent)
Variables not used in the model:							
Total population (ACS: 2013–2017)							
<164,110	2,754	20.9	4,088	22.2	1.31	-1.3	-5.7
164,110–837,288	319	19.3	4,076	18.9	1.25	0.4	2.0
≥837,288	69	25.1	4,166	26.1	1.05	-1.1	-4.1
Census region (ACS: 2013–2017)							
Northwest	217	20.3	2,140	19.9	1.41	0.4	1.9
Midwest	1,055	17.5	2,995	18.1	1.09	-0.6	-3.2
South	1,422	23.8	5,058	25.2	1.09	-1.5	-5.8
West	448	23.7	2,137	24.4	1.01	-0.7	-2.8
Beale codes (USDA: 2013)							
Counties in metro area of 1 million population or more (1)	432	22.1	6,636	22.2	0.77	-0.1	-0.6
Counties in metro areas of less than 1 million population (2,3)	733	21.0	3,616	22.1	1.18	-1.1	-4.8
Nonmetro counties (4–9)	1,976	22.4	2,078	24.8	1.93	-2.4	-9.5
Percentage of population 1+ moved from abroad in the past year (ACS: 2013–2017)							
<0.358	2,207	20.5	4,010	21.5	1.20	-1.0	-4.5
0.358–0.8207	665	21.9	4,276	22.1	1.06	-0.2	-0.9
≥0.8207	270	22.9	4,044	24.0	1.26	-1.1	-4.4
Percentage of population age 5+: speak other language and speak English not at all or not well (ACS: 2013–2017)							
<14.54	1,581	17.4	4,104	16.9	1.13	0.5	2.9
14.54–21.24	787	20.1	4,129	21.7	0.96	-1.7	-7.6
≥21.24	774	28.0	4,097	29.2	1.06	-1.2	-4.0

See notes at end of table.

Table 6-9. Comparison of aggregated county-level indirect and direct estimates for Literacy P1, by subgroup: 2012/2014/2017—Continued

Subgroup (source: Year)	Small area estimate: Number of counties	Small area estimate: Weighted estimate	Direct estimate: Sample size	Direct estimate: Estimate	Direct estimate: Standard error	Percentage point difference	Relative difference (percent)
Percentage of foreign-born people who entered U.S. after year 2010 (ACS: 2013–2017)							
<12.08	1,668	23.4	4,170	24.2	1.19	-0.8	-3.4
12.08–17.86	666	22.1	3,977	24.7	1.18	-2.6	-10.5
≥17.86	808	18.5	4,183	18.6	0.89	0.0	-0.3
Percentage of population age 16+ and didn't work at home: less than 30 minutes to work (ACS: 2013–2017)							
<56.15	526	23.5	4,103	23.1	1.19	0.4	1.9
56.15–68.69	989	21.1	4,021	22.4	1.02	-1.4	-6.1
≥68.69	1,627	21.0	4,206	22.1	1.23	-1.1	-5.1
Percentage of population receiving SNAP/Food stamps (ACS: 2013–2017)							
<9.5	883	17.9	4,040	19.0	1.02	-1.0	-5.5
9.5–13.658	839	20.4	4,139	21.1	1.07	-0.7	-3.3
≥13.658	1,420	26.3	4,151	28.7	1.16	-2.5	-8.6
Median household income-ACS (ACS: 2013–2017)							
<52,017	2,021	25.1	4,070	27.4	1.27	-2.2	-8.1
52,017–62,293	707	23.4	4,120	23.7	1.17	-0.3	-1.2
≥62,293	414	16.9	4,140	17.8	1.15	-0.9	-5.1
Percentage of diagnosed diabetes (DDT: 2013)							
<8.7	451	19.4	3,935	20.2	1.39	-0.8	-3.9
8.7–10.5	785	22.9	4,137	23.4	1.18	-0.5	-2.1
≥10.5	1,906	22.8	4,258	24.3	1.31	-1.4	-5.9
Average amount of grant and scholarship aid received (IPEDS: 2014–2015)							
<7,030	1,226	23.2	4,041	24.5	1.26	-1.3	-5.4
7,030–7,996	949	22.9	4,134	24.5	0.98	-1.6	-6.4
≥7,996	967	19.2	4,155	18.6	0.94	0.6	3.0

See notes at end of table.

Table 6-9. Comparison of aggregated county-level indirect and direct estimates for Literacy P1, by subgroup: 2012/2014/2017—Continued

Subgroup (source: Year)	Small area estimate: Number of counties	Small area estimate: Weighted estimate	Direct estimate: Sample size	Direct estimate: Estimate	Direct estimate: Standard error	Percentage point difference	Relative difference (percent)
Graduation rate of postsecondary institutes (IPEDS: 2014–2015)							
<50	1,122	20.5	3,386	19.1	1.10	1.4	7.4
50–57	1,551	21.0	4,738	23.2	1.21	-2.1	-9.3
≥57	469	23.6	4,206	24.4	1.08	-0.9	-3.5
Infant mortality rate per 1,000 live births (NCHS: 2013)							
<5.68	767	22.1	4,136	22.6	0.99	-0.6	-2.5
5.68–6.69	1,139	22.7	4,026	24.4	1.28	-1.8	-7.2
≥6.69	1,236	20.6	4,168	20.7	1.15	0.0	-0.2

SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

7. SUMMARY

Demand for reliable estimates through small area estimation (SAE) has increased in the past decades, and over the same time, there have been significant enhancements made in SAE methodology. As a result, the National Center for Education Statistics (NCES) reviewed a range of possible methods for SAE to meet the need for state- and county-level estimates of adult skills. At the conclusion of the review process and after extensive evaluation, a statistical modeling approach was selected and used to produce indirect estimates, which are available to the public on the Skills Map website at http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2020047.

The statistical modeling approach produces four different state- and county-level estimates for adult literacy and numeracy proficiencies: an average score (on the Program for the International Assessment of Adult Competencies [PIAAC] scale of 0–500) and the proportion of adults at or below Level 1, at Level 2, and at or above Level 3. The indirect estimates relied on the pooled 2012/2014/2017 PIAAC data as well as the American Community Survey (2013–2017) data. The modeling depends on (a) PIAAC's direct survey estimates, (b) Hierarchical Bayes (HB) linear three-fold models (bivariate for proportions, univariate for averages), and (c) seven covariates relating to educational attainment, poverty, race/ethnicity, health insurance coverage, and occupation (service industry). The covariates were selected through an exhaustive search process, and the model was subjected to rigorous diagnostic checks before predictions were made for all 3,142 counties. A variety of methods was used to evaluate the fit of the HB models to the county estimates, including various methods of internal model validation as well as external model validation. The checks showed that the final models used were insensitive to different model assumptions and the measures indicated good fits to the data.

The statistical modeling approach that was selected computes the indirect estimates for states as weighted aggregates of the indirect county estimates, where the weights represent the proportion of the state's household population of adults ages 16 to 74 in each county. Overall, the state-level estimates are more precise than the county-level estimates. Specifically, coefficients of variation (CVs) for the county-level estimates for the proportion at or below Level 1 in literacy are generally of the order of 10 percent, while the state predictions have a median CV of 8.1 percent. Estimates with CVs of this magnitude are considered to be precise, i.e., at a high confidence level. To a lesser extent, estimates for states and counties from which some persons were sampled in the PIAAC 2012/2014/2017 combined household sample are more precise than estimates for states or counties that had no persons sampled. With the positive diagnostics and evaluation results, the precision levels of the indirect estimates should give the data users confidence in using these model-based estimates.

REFERENCES

Alvarez, I., Niemi, J., and Simpson, M. (2014). Bayesian Inference for a Covariance Matrix. *Conference on Applied Statistics in Agriculture*. Retrieved December 24, 2019, from https://doi.org/10.4148/2475-7772.1004.

Battese, G.E., Harter, R.M., and Fuller, W.A. (1988). An Error-Components Model for Prediction of County Crop Areas Using Survey and Satellite Data. *Journal of the American Statistical Association*, *83*: 28–36.

Bauder, M., Luery, D., and Szelepka, S. (2016). Small Area Estimation of Health Insurance Coverage in 2010–2014. [Working paper]. Washington, DC: U.S. Census Bureau.

Bell, W., Chung, H., Datta, G., and Franco, C. (2019). Measurement Error in Small Area Estimation: Functional Versus Structural Versus Naïve Models. *Survey Methodology*. *45*(1): 61–80.

Benavent, R., and Morales, D. (2016). Multivariate Fay–Herriot Models For Small Area Estimation. *Computational Statistics and Data Analysis*, *94*: 372–390.

Bijlsma, I., van den Brakel, J., van der Velden, R., and Allen, J. (2017). Estimating Literacy Levels at a Detailed Regional Level: An Application Using Dutch Data. *ROA Research Memorandum ROA-RM-2017/6*. Maastricht, Netherlands: Research Centre for Education and the Labour Market.

Chen, C., Wakefield, J., and Lumley, T. (2014). The Use of Sampling Weights in Bayesian Hierarchical Models for Small Area Estimation. Retrieved December 24, 2019, from *Spatial and Spatio-temporal Epidemiology*, *11*: 33–43.

Citro, C.F., and Kalton, G. (2000). *Small Area Income and Poverty Estimates: Priorities for 2000 and Beyond.* Washington, DC: National Academy Press.

Coley, R.J. (Ed.). (1996). International Adult Literacy. *ETS Policy Notes*, *7*(1).

Cruze, N., Erciulescu, A.L., Nandram, B., Barboza, W.J., and Young, L.J. (2019). Producing Official County-Level Agricultural Estimates in the United States: Needs and Challenges. *Statistical Science*. *34*(2): 301–316. Retrieved December 24, 2019, from https://projecteuclid.org/euclid.ss/1563501643.

Czajka, J., Sukasih, A., and Maccarone, A. (2014). *Small Area Estimation: New Developments and Directions for HHS Data. Final Report.* Prepared for the U.S. Department of Health and Human Services. Washington, DC: Mathematica Policy Research.

Datta, G., Rao, J.N.K., and Smith, D.D. (2005). On Measuring the Variability of Small Area Estimators Under a Basic Area Level Model. *Biometrika*, *92*(1): 183–196. Retrieved December 24, 2019, from https://academic.oup.com/biomet/article/92/1/183/248551.

Erciulescu, A., Berg, E., Cecere, W., and Ghosh, M. (2019). A Bivariate Hierarchical Bayesian Model for Estimating Cropland Cash Rental Rates at the County Level. *Survey Methodology*. Statistics Canada, Catalogue No. 12-001-X, *45*(2). Retrieved December 24, 2019, from https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2019002/article/00002-eng.pdf?st=WTrxRpWi.

Erciulescu, A., Cruze, N., and Nandram, B. (2017, August). *Small Area Estimates for End-Of-Season Agriculture Quantities*. [PowerPoint slides]. Joint Statistical Meetings, *Section on Survey Research Methods*. Baltimore, MD: American Statistical Association. Retrieved December 24, 2019, from https://ww2.amstat.org/meetings/jsm/2017/onlineprogram/AbstractDetails.cfm?abstractid=323125.

Erciulescu, A., Cruze, N., and Nandram, B. (2019). Model-Based County-Level Crop Estimates Incorporating Auxiliary Sources of Information. *Journal of the Royal Statistical Society*, *Series A*, *182*(1): 283–303. doi: 10.1111/rssa.12390. Retrieved December 24, 2019, from https://rss.onlinelibrary.wiley.com/doi/full/10.1111/rssa.12390.

Fay, R.E., and Diallo, M.S. (2012). Small Area Estimation Alternatives for the National Crime Victimization Survey. *Joint Statistical Meetings Proceedings, Section on Survey Research Methods*, 3742–3756. Prepared for the Bureau of Justice Statistics. Washington, DC: Westat. Retrieved December 24, 2019, from https://www.bjs.gov/content/pub/pdf/FayDiallo_JSM2012.pdf.

Fay, R.E., and Diallo, M.S. (2015). *Developmental Estimates of Subnational Crime Rates Based on the National Crime Victimization Survey.* Prepared for the Bureau of Justice Statistics. Washington, DC: Westat. Retrieved December 24, 2019, from https://www.bjs.gov/content/pub/pdf/descrbncvs.pdf.

Fay, R.E., and Herriot, R.A. (1979). Estimates of Income for Small Places: An Application of James-Stein Procedures to Census Data. *Journal of the American Statistical Association*, *74*: 269–277.

Fay, R.E., and Li, J. (2011). Predicting Violent Crime Rates for the 2010 Redesign of the National Crime Victimization Survey (NCVS). *Joint Statistical Meetings Proceedings*, *Section on Survey Research Methods*, 1663–1676. Prepared for the Bureau of Justice Statistics. Washington, DC: Westat. Retrieved December 24, 2019, from https://www.bjs.gov/content/pub/pdf/FayLiJSM2011.pdf.

Fay, R.E., Planty, M., and Diallo, M.S. (2013). Small Area Estimates From the National Crime Victimization Survey. *Joint Statistical Meetings Proceedings*, *Section on Survey Research Methods*, 1544–1557. Prepared for the Bureau of Justice Statistics. Washington, DC: Westat. Retrieved December 24, 2019, from https://www.bjs.gov/content/pub/pdf/jpsm2013.pdf.

Folsom, R., Shah, B., and Vaish, A. (1999). Substance Abuse in States: A Methodological Report on Model-Based Estimates From the 1994–1996 National Household Surveys on Drug Abuse. *Proceedings of the American Statistical Association Section on Survey Research Methods*, 371–375.

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models Via Coordinate Descent. *Journal of Statistical Software*, *33*(1). Retrieved December 24, 2019, from https://www.jstatsoft.org/article/view/v033i01/v33i01.pdf.

Fushiki, T. (2011). Estimation of Prediction Error by Using K-Fold Cross-Validation. *Statistics and Computing*, *21*(2): 137–146. Retrieved December 24, 2019, from https://rd.springer.com/article/10.1007/s11222-009-9153-8.

Gelman, A. (2006). Prior Distributions for Variance Parameters in Hierarchical Models (Comment on Article by Browne and Draper). *Bayesian Analysis*, *1*(3): 515–534. doi: 10.1214/06-BA117A. Retrieved December 24, 2019, from https://projecteuclid.org/euclid.ba/1340371048.

Gelman, A. (2013). Two Simple Examples for Understanding Posterior *P* Values Whose Distributions Are Far From Uniform. *Electronic Journal of Statistics 7*: 2595–2602. doi:10.1214/13-EJS854. Retrieved December 24, 2019, from https://projecteuclid.org/euclid.ejs/1382448225.

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehatri, A., and Rubin, D.B. (2013). *Bayesian Data Analysis* (3rd ed.). Boca Raton, Florida: Chapman & Hall/CRC.

Gelman, A., and Rubin, D. (1992). Inference From Iterative Simulation Using Multiple Sequences. *Statistical Science*, *7*(4): 457–472.

Gibson, A., and Hewson, P. (2012). *2011 Skills for Life Survey: Small Area Estimation User Guide*. The United Kingdom Department for Business, Innovation and Skills Research Paper Number 81D. Retrieved December 24, 2019, from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/36076/12-1316-2011-skills-for-life-small-area-estimation-user-guide.pdf.

Goodman, M., Finnegan, R., Mohadjer, L., Krenzke, T., and Hogan, J. (2013). *Literacy, Numeracy, and Problem Solving in Technology-Rich Environments Among U.S. Adults: Results from the Program for the International Assessment of Adult Competencies 2012: First Look* (NCES 2014-008). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved December 24, 2019, from https://nces.ed.gov/pubs2014/2014008.pdf.

Greenberg, E., Macias, R.F., Rhodes, D., and Chan, T. (2001). *English Literacy and Language Minorities in the United States* (NCES 2001-464). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved December 24, 2019, from https://nces.ed.gov/pubs2001/2001464.pdf.

Hájek, J. (1971). Comment on a paper by D. Basu. In V.P. Godambe and D.A. Sprott (Eds.), *Foundations of Statistical Inference*. Toronto: Holt, Rinehart, and Winston, p. 236.

Hogan, J., Thornton, N., Diaz-Hoffmann, L., Mohadjer, L., Krenzke, T., Li, J., Van De Kerckhove, W., Yamamoto, K., and Khorramdel, L. (2016). *U.S. Program for the International Assessment of Adult Competencies (PIAAC) 2012/2014: Main Study and National Supplement Technical Report* (NCES 2016-036REV). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Hsu, J.C. (1996). *Multiple Comparisons: Theory and Methods*. London: Chapman and Hall, ISBN 0-412-98281-1.

Khan, D., Wei, R., He, Y., Shin, H-C, and Malec, D. (2018). *Bayesian State-Level Estimates of Diabetes Prevalence in the United States 2006–2015*. Presented at the Federal Commission of Statistical Methodology conference on March 7, 2018, Washington, DC. Presentation slides retrieved March 24, 2020, from https://nces.ed.gov/FCSM/pdf/C_4Khan_March7_2018FIN.pdf.

Kirsch, I.S., Jungeblut, A., Jenkins, L., and Kolstad, A. (2002). *Adult Literacy in America: A First Look at the Results of the National Adult Literary Survey (NALS)*. U.S. Department of Education. Washington DC: National Center for Education Statistics.

Krenzke, T., Mohadjer, L., Fay, R., Li, J., Rao, J.N.K., Ren, W., and Van de Kerckhove, W. (2018). *PIAAC Small Area Estimation Analysis Plan*. Submitted to National Center for Education Statistics in April 2018.

Lahiri, P., and Suntornchost, J. (2015). Variable Selection for Linear Mixed Models With Applications in Small Area Estimation. *Sankhya B: The Indian Journal of Statistics*, *77*: 312. doi: 10.1007/s13571-015-0096-0.

Lewandowski, D., Kurowicka, D., and Joe, H. (2009). Generating Random Correlation Matrices Based on Vines and Extended Onion Method. *Journal of Multivariate Analysis*, *100*(9): 1989–2001. Retrieved December 24, 2019, from 10.1016/j.jmva.2009.04.008.

Lunn, D.J., Thomas, A., Best, N., and Spiegelhalter, D. (2000). WinBUGS—a Bayesian Modeling Framework: Concepts, Structure, and Extensibility. *Statistics and Computing*, *10*(4): 325–337.

Marchetti, S., Giusti, C., Pratesi, M., Salvati, N., Giannotti, F., Pedreschi, D., Rinzivillo, S., Pappalardo, L., and Gabrielli, L. (2015). Small Area Model-Based Estimators Using Big Data Sources. *Journal of Official Statistics*, *31*(2): 263–281.

Miller, J.W. (2016). America's Most Literate Cities, 2016. Retrieved December 24, 2019, from http://web.ccsu.edu/americasmostliteratecities/2016/.

Mislevy, R.J. (1991). Randomization-Based Inference About Latent Variables From Complex Samples. *Psychometrika*, *56*(2), 177–196.

Mohadjer, L., Kalton, G., Krenzke, T., Liu, B., Van de Kerckhove, W., Li, L., Sherman, D., Dillman, J., Rao, J.N.K., and White, S. (2009). *National Assessment of Adult Literacy Indirect County and State Estimates of the Percentage of Adults at the Lowest Literacy Level for 1992 and 2003* (NCES 2009-482). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Mohadjer, L., Rao, J.N.K., Liu, B., Krenzke, T., and Van de Kerckhove, W. (2011). Hierarchical Bayes Small Area Estimates of Adult Literacy Using Unmatched Sampling and Linking Models. *Journal of the Indian Society of Agricultural Statistics*, *66*(1): 1–9.

Nandram, B., Erciulescu, A., and Cruze, N. (2019). Bayesian Benchmarking of the Fay-Herriot Model Using Random Deletion. *Survey Methodology*. Statistics Canada, Catalogue No. 12-001-X, *45*(2). Retrieved December 24, 2019, from https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2019002/article/00004-eng.pdf?st=Ypes1CWP.

Organization for Economic Cooperation and Development (OECD). (2016). *Technical Report of the Survey of Adult Skills (PIAAC)* (2nd ed.). Paris, France: OECD. Retrieved December 24, 2019, from https://www.oecd.org/skills/piaac/PIAAC_Technical_Report_2nd_Edition_Full_Report.pdf.

Pfefferman, D. (2013). New Important Developments in Small Area Estimation. *Statistical Science*, *28*: 40–68.

Pfeffermann, D., Terryn, B., and Moura, F.A.S. (2008). Small Area Estimation Under a Two-Part Random Effects Model With Application to Estimation of Literacy in Developing Countries. *Survey Methodology, 34*(2): 235–249.

Raghunathan, T.E., Xie, D., Schenker, N., Parsons, V.L., Davis, W.W., Dodd, K.W., and Feuer, E.J. (2007). Combining Information From Two Surveys to Estimate County-Level Prevalence Rates of Cancer Risk Factors and Screening. *Journal of the American Statistical Association, 102*: 474–486.

Rampey, B.D., Finnegan, R., Goodman, M., Mohadjer, L., Krenzke, T., Hogan, J., and Provasnik, S. (2016). *Skills of U.S. Unemployed, Young, and Older Adults in Sharper Focus: Results From the Program for the International Assessment of Adult Competencies (PIAAC) 2012/2014: First Look* (NCES 2016-039rev). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Rao, J.N.K. (2003). *Small Area Estimation* (Wiley Series in Survey Methodology). New York: Wiley.

Rao, J.N.K., and Molina, I. (2015). *Small Area Estimation* (Wiley Series in Survey Methodology) (2nd ed.). Hoboken, New Jersey: Wiley.

Rubin, D. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley.

Särndal, C.E., and Hidiroglou, M. (1989). Small Domain Estimation: A Conditional Analysis. *Journal of the American Statistical Association*, *84*: 266–275. doi 10.1080/01621459.1989.10478765.

Särndal, C.E., Swensson, B., and Wretman, J. (1992). *Model Assisted Survey Sampling*. New York: Springer.

Sundsoy, P. (2016). Can Mobile Usage Predict Illiteracy in a Developing Country? arXiv preprint arXiv: 1607.01337. Retrieved December 24, 2019, from https://arxiv.org/ftp/arxiv/papers/1607/1607.01337.pdf.

Tibshirani, R. (1996). Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society B*, *58*: 267–288.

Vaish, A.K. (2017, July). *Small Area Estimation With Data From Multiple Sources*. Presented at the 61st International Statistical Institute World Statistics Congress Satellite Meeting, Paris, France.

Weiss, B.D., Hart, G., and Pust, R.E. (1991). The Relationship Between Literacy and Health. *Journal of Health Care Pool Underserved, 1*(4), 351–63.

Wolter, K.M. (2007). Taylor Series Methods and Generalized Variance Functions. In: *Introduction to Variance Estimation*. Statistics for Social and Behavioral Sciences. New York: Springer.

Yamamoto, K. (2014). *Using PIAAC Data to Produce Regional Estimates*. [Unpublished manuscript]. Princeton, N.J.: Educational Testing Service.

Yamamoto, K., Khorramdel, L., and von Davier, M. (2013). Scaling PIAAC Cognitive Data. In: *Technical Report of the Survey of Adult Skills (PIAAC)* [Prepublication copy], chapter 17, pp. 406–438. Paris: Organisation for Economic Cooperation and Development. Retrieved December 24, 2019, from https://www.oecd.org/site/piaac/_Technical%20Report_17OCT13.pdf.

Ybarra, L.M.R., and Lohr, S.L. (2008). Small Area Estimation When Auxiliary Information is Measured With Error. *Biometrika*, *95*(4): 919–931.

Appendix A

List of Potential Covariates

APPENDIX A

LIST OF POTENTIAL COVARIATES

Appendix A of the report includes tables of data sources and associated variables considered in the model development stage as potential covariates. Detailed tables are provided below.

Table A-1. List of county-level variables, by source and year

County characteristics	Source	Year
Poverty		
Percentage of population below 150 percent poverty line	ACS	2013–2017
Percentage of population receiving SNAP/Food stamps	ACS	2013–2017
Percentage of population below 100 percent poverty line	ACS	2013–2017
Percentage of population in poverty (all ages)	SAIPE	2015
Income		
Median household income—ACS	ACS	2013–2017
Median household income—SAIPE	SAIPE	2015
Per capita personal income	BEA	2015
Education		
Percentage of population aged 25+: with education less than high school (no high school diploma)	ACS	2013–2017
Percentage of population aged 25+: with high school diploma, no college	ACS	2013–2017
Percentage of population aged 25+: with education more than high school (including some college, no degree)	ACS	2013–2017
English-speaking ability for people who speak other language		
Percentage of population aged 5+: speaking other languages and speak English not at all or not well	ACS	2013–2017
Percentage of population aged 5+: speaking other languages	ACS	2013–2017

See notes at end of table.

Table A-1. List of county-level variables, by source and year—Continued

County characteristics	Source	Year
Urban/rural		
Metro or nonmetro counties	ACS	2013–2017
Counties in metro area of 1 million population or more	USDA	2013
Counties in metro areas of less than 1 million population	USDA	2013
Nonmetro counties	USDA	2013
Race/ethnicity		
Percentage of Hispanics	ACS	2013–2017
Percentage of Whites	ACS	2013–2017
Percentage of Blacks	ACS	2013–2017
Percentage of Asians	ACS	2013–2017
Percentage of American Indians and Alaska Natives	ACS	2013–2017
Percentage of Native Hawaiians and Pacific Islanders	ACS	2013–2017
Percentage of Other races	ACS	2013–2017
Foreign-born status		
Percentage of foreign-born people who entered United States after year 2010	ACS	2013–2017
Percentage of foreign-born people who entered United States between years 1990 and 2009	ACS	2013–2017
Percentage of foreign-born people who entered United States after year 1990	ACS	2013–2017
Percentage of foreign-born people who entered United States before year 1990	ACS	2013–2017
Percentage of population born outside of United States	ACS	2013–2017
Age		
Percentage of population 16–54 years old	ACS	2013–2017
Percentage of population 55–64 years old	ACS	2013–2017
Percentage of population 65+ years old	ACS	2013–2017
Gender		
Percentage of male population	ACS	2013–2017

See notes at end of table.

Table A-1. List of county-level variables, by source and year—Continued

County characteristics	Source	Year
Employment status		
Unemployment rate	BLS	2015
Percentage of population aged 20–64: in armed forces	ACS	2013–2017
Percentage of population aged 20–64: in labor force and employed	ACS	2013–2017
Percentage of population aged 20–64: in labor force and unemployed	ACS	2013–2017
Percentage of population aged 20–64: not in labor force	ACS	2013–2017
Occupation		
Percentage of population aged 16+: management/professional occupations	ACS	2013–2017
Percentage of population aged 16+: service occupation	ACS	2013–2017
Percentage of population aged 16+: sales/office occupation	ACS	2013–2017
Percentage of population aged 16+: natural resources/construction/maintenance occupation	ACS	2013–2017
Percentage of population aged 16+: military	ACS	2013–2017
Percentage of population aged 16+: production/transportation/moving occupation	ACS	2013–2017
Census division		
New England	ACS	2013–2017
Middle Atlantic	ACS	2013–2017
East North Central	ACS	2013–2017
West North Central	ACS	2013–2017
South Atlantic	ACS	2013–2017
East South Central	ACS	2013–2017
West South Central	ACS	2013–2017
Mountain	ACS	2013–2017
Pacific	ACS	2013–2017
Journey to work		
Percentage of population aged 16+ and didn't work at home: less than 30 minutes to work	ACS	2013–2017
Percentage of population aged 16+ and didn't work at home: 30–44 minutes to work	ACS	2013–2017
Percentage of population aged 16+ and didn't work at home: 45–59 minutes to work	ACS	2013–2017
Percentage of population aged 16+ and didn't work at home: 60+ minutes to work	ACS	2013–2017

See notes at end of table.

Table A-1. List of county-level variables, by source and year—Continued

County characteristics	Source	Year
Housing unit tenure and phone service		
Percentage of owner-occupied housing unit	ACS	2013–2017
Percentage of renter-occupied housing unit	ACS	2013–2017
Percentage of owner-occupied housing unit with phone service available	ACS	2013–2017
Percentage of renter-occupied housing unit with phone service available	ACS	2013–2017
Percentage of occupied housing unit	ACS	2013–2017
Plumbing facilities		
Percentage of housing unit with plumbing facilities	ACS	2013–2017
Marital status		
Percentage of population 15+: never married	ACS	2013–2017
Percentage of population 15+: married	ACS	2013–2017
Percentage of population 15+: widowed	ACS	2013–2017
Percentage of population 15+: divorced	ACS	2013–2017
Migration		
Percentage of population 1+: in different house in the past year	ACS	2013–2017
Percentage of population 1+: in different county in the past year	ACS	2013–2017
Percentage of population 1+: in different state in the past year	ACS	2013–2017
Percentage of population 1+: moved from abroad in the past year	ACS	2013–2017
Health		
Percentage of civilian noninstitutionalized population with one type of health insurance coverage	ACS	2013–2017
Percentage of civilian noninstitutionalized population with two or more types of health insurance coverage	ACS	2013–2017
Percentage of civilian noninstitutionalized population with no health insurance coverage	ACS	2013–2017
Percentage of diagnosed diabetes	DDT	2013
Percentage of obesity	DDT	2013
Percentage of population eligible for Medicaid	CMS	2015

See notes at end of table.

Table A-1. List of county-level variables, by source and year—Continued

County characteristics	Source	Year
Tax		
Average number of tax returns per person	SOI	2014
Average number of returns with unemployment compensation per person	SOI	2014
Average number of returns with taxable Social Security benefits per person	SOI	2014
Proportion of the amount of unemployment compensation among all tax return amount	SOI	2014
Proportion of the amount of taxable Social Security benefits among all tax return amount	SOI	2014

NOTE: ACS: American Community Survey; SNAP: Supplemental Nutrition Assistance Program; SAIPE: Small Area Income and Poverty Estimates program; BEA: Bureau of Economic Analysis; USDA: U.S. Department of Agriculture; BLS: Bureau of Labor Statistics; DDT: Centers for Disease Control and Prevention's Division of Diabetes Translation; CMS: Centers for Medicare & Medicaid Services; SOI: The Statistics of Income Data.
SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

Table A-2. List of state-level variables, by source and year

State characteristics	Source	Year
Socioeconomic status		
Average annual pay	BLS	2015
Homeownership rate	Housing Vacancies and Home Ownership (CPS/HVS)	2015
Education		
Adult basic education enrollment rate	OCTAE	2015
Adult secondary education enrollment rate	OCTAE	2015
English as a second language enrollment rate	OCTAE	2015
Graduation rate of postsecondary institutes	IPEDS	2014–2015
Average weighted monthly salary for full-time instructional staff	IPEDS	2014–2015
Average amount of grant and scholarship aid received	IPEDS	2014–2015
Annual college cost (tuition and fees)	IPEDS	2014–2015
GED test completion rate	GED Testing Service (GEDTS)	2013
Average 4th-grade reading composite scale scores	NAEP	2015
Average 4th-grade math composite scale scores	NAEP	2015
Average 8th-grade reading composite scale scores	NAEP	2015
Average 8th-grade math composite scale scores	NAEP	2015

See notes at end of table.

Table A-2. List of state-level variables, by source and year—Continued

State characteristics	Source	Year
Other area characteristics		
Infant mortality rate per 1,000 live birth	NCHS, Vital Statistics of the United States, annual; and unpublished data	2013
Women 15–50 years old who had a birth in the past 12 months (per 1,000 15- through 50-year-old women)	ACS	2011–2015
Physicians per 100,000 population	AMA, Chicago, IL, Physician Characteristics and Distribution in the United States, 2014	2015
Violent crime rate per 100,000 population	FBI, Crime in the United States, annual	2015
Federal aid to state and local governments per capita	Census Bureau, Federal Aid to States for Fiscal Year 2010	2010
State government general revenue per capita	Census Bureau; State and Local Government Finance Estimates by State, annual, and unpublished data	2014
Energy consumption per person	EIA, State Energy Data Report, 2014	2014
Traffic fatalities per 100 million vehicle miles	NHTSA, Traffic Safety Facts, annual	2015
Birth rate	National Vital Statistics Reports, 2015	2017
Birth rate for teenagers aged 15–19	National Vital Statistics Reports, 2015	2017

NOTE: BLS: Bureau of Labor Statistics; CPS/HVS: Housing Vacancies and Homeownership; OCTAE: Office of Career, Technical, and Adult Education; IPEDS: Integrated Postsecondary Education Data System; GED: General Educational Development; NAEP: National Assessment of Educational Progress; NCHS: National Center for Health Statistics; ACS: American Community Survey; AMA: American Medical Association; FBI: Federal Bureau of Investigation; EIA: Energy Information Administration; NHTSA: National Highway Traffic Safety Administration.
SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

Table A-3. PIAAC county- and state-level variable correlations with literacy/numeracy proficiency outcomes: 2012/2014/2017

Variable	Literacy P1	Literacy P2	Literacy P3	Literacy average	Numeracy P1	Numeracy P2	Numeracy P3	Numeracy average
County-level								
Percentage of population aged 25+: with education less than high school	0.72	0.22	-0.70	-0.73	0.74	-0.11	-0.63	-0.73
Percentage of population aged 25+: with high school diploma, no college	0.28	0.59	-0.59	-0.44	0.36	0.41	-0.59	-0.44
Percentage of population aged 25+: with education more than high school	-0.56	-0.52	0.77	0.68	-0.63	-0.22	0.73	0.68
Percentage of population below 100 percent poverty line	0.65	0.24	-0.65	-0.67	0.74	-0.10	-0.64	-0.71
Percentage of population receiving SNAP/Food stamps	0.59	0.31	-0.66	-0.64	0.69	0.01	-0.66	-0.68
Percentage of population below 150 percent poverty line	0.67	0.28	-0.70	-0.70	0.75	-0.05	-0.68	-0.73
Percentage of population in poverty (all ages)	0.64	0.23	-0.64	-0.64	0.71	-0.09	-0.62	-0.68
ACS median household income—log transformed	-0.49	-0.42	0.65	0.56	-0.59	-0.13	0.64	0.59
SAIPE median household income	-0.49	-0.42	0.65	0.56	-0.59	-0.13	0.64	0.59
Per capita personal income—log transformed	-0.17	-0.34	0.35	0.23	-0.20	-0.15	0.28	0.21
Percentage of population aged 5+: speak other language and speak English not at all or not well	0.15	-0.15	-0.02	-0.10	0.11	-0.18	0.01	-0.12
Percentage of population aged 5+: speaking other languages	0.24	-0.37	0.05	-0.15	0.14	-0.33	0.07	-0.13
Percentage of Hispanics	0.33	-0.25	-0.10	-0.27	0.26	-0.26	-0.09	-0.26
Percentage of Blacks	0.37	-0.03	-0.27	-0.32	0.46	-0.24	-0.28	-0.39
Percentage of Asians	-0.04	-0.40	0.28	0.13	-0.15	-0.27	0.31	0.16
Percentage of American Indians and Alaska Natives	0.01	-0.04	0.02	-0.03	0.01	0.00	-0.01	-0.05
Percentage of Whites	-0.33	0.27	0.08	0.23	-0.34	0.37	0.09	0.28
Percentage of Native Hawaiians and Pacific Islanders	-0.04	-0.13	0.12	0.07	-0.08	-0.05	0.11	0.08
Percentage of Other races	0.20	-0.29	0.03	-0.12	0.14	-0.29	0.05	-0.11
Percentage of foreign-born people who entered United States after year 2010	-0.22	-0.27	0.34	0.31	-0.21	-0.18	0.31	0.26
Percentage of foreign-born people who entered United States between years 1990 and 2009	0.16	-0.19	-0.01	-0.10	0.13	-0.23	0.02	-0.13
Percentage of foreign-born people who entered United States after year 1990	0.00	-0.31	0.19	0.09	-0.01	-0.29	0.20	0.05
Percentage of foreign-born people who entered United States before year 1990	-0.02	-0.02	0.03	0.02	-0.07	0.07	0.02	0.06

See notes at end of table.

Table A-3. PIAAC county- and state-level variable correlations with literacy/numeracy proficiency outcomes: 2012/2014/2017—Continued

Variable	Literacy P1	Literacy P2	Literacy P3	Literacy average	Numeracy P1	Numeracy P2	Numeracy P3	Numeracy average
Percentage of population born outside of United States	0.12	-0.40	0.16	-0.03	0.02	-0.33	0.18	-0.01
Percentage of population 16–54 years old	0.11	-0.20	0.04	-0.05	0.07	-0.17	0.04	-0.05
Percentage of population 55–64 years old	-0.16	0.32	-0.08	0.07	-0.14	0.31	-0.06	0.09
Percentage of population 65+ years old	-0.07	0.36	-0.17	-0.03	-0.03	0.33	-0.17	-0.02
Percentage of male population	0.19	0.00	-0.15	-0.18	0.14	-0.12	-0.06	-0.12
Percentage of population aged 20–64: in armed forces	-0.10	-0.04	0.10	0.11	-0.06	0.01	0.05	0.09
Percentage of population aged 20–64: in labor force and employed	-0.52	-0.41	0.66	0.58	-0.60	-0.12	0.64	0.60
Percentage of population aged 20–64: in labor force and unemployed	0.33	0.05	-0.29	-0.33	0.39	-0.07	-0.33	-0.39
Percentage of population aged 20–64: not in labor force	0.62	0.24	-0.63	-0.63	0.67	-0.07	-0.59	-0.64
Percentage of population aged 16+: management/ professional occupations	-0.38	-0.50	0.61	0.51	-0.44	-0.33	0.62	0.52
Percentage of population aged 16+: service occupation	0.34	0.07	-0.31	-0.37	0.39	-0.07	-0.33	-0.39
Percentage of population aged 16+: sales/office occupation	-0.05	0.17	-0.07	-0.04	0.02	0.24	-0.16	-0.09
Percentage of population aged 16+: natural resources/ construction/maintenance occupation	0.22	0.36	-0.40	-0.30	0.22	0.23	-0.35	-0.28
Percentage of population aged 16+: military	-0.09	-0.02	0.08	0.10	-0.06	0.03	0.04	0.09
Percentage of population aged 16+: production/ transportation/moving occupation	0.29	0.43	-0.50	-0.39	0.32	0.29	-0.48	-0.38
Percentage of population aged 16+ and didn't work at home: less than 30 minutes to work	0.00	-0.02	0.01	0.01	0.02	0.01	-0.03	-0.02
Percentage of population aged 16+ and didn't work at home: 30–44 minutes to work	-0.02	-0.09	0.08	0.05	-0.01	-0.16	0.11	0.05
Percentage of population aged 16+ and didn't work at home: 45–59 minutes to work	-0.07	0.04	0.03	0.06	-0.09	0.00	0.08	0.09
Percentage of population aged 16+ and didn't work at home: 60+ minutes to work	0.07	0.11	-0.12	-0.11	0.02	0.14	-0.10	-0.07
Percentage of owner-occupied housing unit	-0.20	0.32	-0.05	0.08	-0.20	0.38	-0.04	0.12

See notes at end of table.

Table A-3. PIAAC county- and state-level variable correlations with literacy/numeracy proficiency outcomes: 2012/2014/2017—Continued

Variable	Literacy P1	Literacy P2	Literacy P3	Literacy average	Numeracy P1	Numeracy P2	Numeracy P3	Numeracy average
Percentage of renter-occupied housing unit	0.20	-0.32	0.05	-0.08	0.20	-0.38	0.04	-0.12
Percentage of owner-occupied housing unit with phone service available	-0.30	-0.11	0.30	0.29	-0.32	0.04	0.28	0.31
Percentage of renter-occupied housing unit with phone service available	-0.21	-0.05	0.20	0.19	-0.21	0.02	0.18	0.19
Percentage of occupied housing unit	-0.10	-0.26	0.24	0.15	-0.15	-0.14	0.23	0.16
Percentage of housing unit with plumbing facilities	-0.15	-0.07	0.16	0.15	-0.14	0.02	0.12	0.14
Percentage of population 15+: never married	0.24	-0.37	0.05	-0.11	0.23	-0.41	0.03	-0.16
Percentage of population 15+: married	-0.35	0.19	0.15	0.26	-0.40	0.31	0.19	0.33
Percentage of population 15+: widowed	0.35	0.47	-0.57	-0.45	0.41	0.28	-0.57	-0.46
Percentage of population 15+: divorced	0.07	0.34	-0.27	-0.15	0.18	0.23	-0.31	-0.19
Percentage of population 1+: in different house in the past year	-0.10	-0.20	0.20	0.16	-0.05	-0.23	0.19	0.13
Percentage of population 1+: in different county in the past year	0.10	-0.01	-0.07	-0.10	0.11	-0.11	-0.04	-0.08
Percentage of population 1+: in different state in the past year	-0.20	-0.17	0.26	0.27	-0.16	-0.20	0.28	0.25
Percentage of population 1+: moved from abroad in the past year	-0.15	-0.52	0.45	0.32	-0.23	-0.44	0.49	0.33
Percentage of civilian noninstitutionalized population with one type of health insurance coverage	-0.43	-0.20	0.46	0.43	-0.48	0.00	0.46	0.46
Percentage of civilian noninstitutionalized population with two or more types of health insurance coverage	-0.04	0.29	-0.15	-0.02	0.01	0.23	-0.15	-0.01
Percentage of civilian noninstitutionalized population with no health insurance coverage	0.52	0.00	-0.41	-0.48	0.53	-0.17	-0.40	-0.51
Percentage of diagnosed diabetes	0.39	0.45	-0.58	-0.49	0.50	0.19	-0.59	-0.52
Percentage of obesity	0.40	0.38	-0.55	-0.47	0.48	0.17	-0.56	-0.49
Percentage of population eligible for Medicaid	0.54	0.09	-0.48	-0.52	0.55	-0.06	-0.49	-0.54
Average number of tax returns per person	-0.12	-0.40	0.35	0.18	-0.20	-0.22	0.33	0.19

See notes at end of table.

Table A-3. PIAAC county- and state-level variable correlations with literacy/numeracy proficiency outcomes: 2012/2014/2017—Continued

Variable	Literacy P1	Literacy P2	Literacy P3	Literacy average	Numeracy P1	Numeracy P2	Numeracy P3	Numeracy average
Average number of returns with unemployment compensation per person	-0.04	0.00	0.03	0.03	-0.06	0.07	0.01	0.04
Average number of returns with taxable Social Security benefits per person	-0.38	0.28	0.12	0.28	-0.36	0.38	0.10	0.30
Proportion of the amount of unemployment compensation among all tax return amount	0.09	0.10	-0.14	-0.12	0.09	0.09	-0.14	-0.12
Proportion of the amount of taxable Social Security benefits among all tax return amount	-0.09	0.42	-0.19	-0.02	-0.03	0.39	-0.21	-0.03
Unemployment rate	0.48	0.15	-0.46	-0.48	0.54	-0.09	-0.45	-0.51
Counties in metro area of 1 million population or more	-0.12	-0.23	0.24	0.17	-0.15	-0.12	0.22	0.16
Counties in metro areas of less than 1 million population	-0.07	-0.01	0.06	0.06	-0.07	0.05	0.04	0.05
Nonmetro counties	0.22	0.28	-0.35	-0.25	0.26	0.08	-0.29	-0.24
New England	-0.16	-0.02	0.14	0.15	-0.16	0.02	0.14	0.16
Middle Atlantic	-0.07	-0.01	0.06	0.06	-0.08	0.05	0.04	0.06
East North Central	-0.17	0.08	0.08	0.11	-0.13	0.11	0.06	0.09
West North Central	-0.11	0.03	0.07	0.10	-0.18	0.10	0.11	0.14
South Atlantic	0.07	0.00	-0.06	-0.05	0.11	-0.03	-0.09	-0.10
East South Central	0.16	0.23	-0.27	-0.19	0.23	0.07	-0.26	-0.18
West South Central	0.27	-0.09	-0.15	-0.23	0.26	-0.16	-0.15	-0.25
Mountain	-0.14	-0.01	0.12	0.15	-0.14	-0.03	0.16	0.17
Pacific	0.06	-0.26	0.12	0.00	-0.02	-0.16	0.12	0.01
State-level								
Adult basic education enrollment rate	0.20	0.31	-0.35	-0.25	0.28	0.14	-0.36	-0.28
Physicians per 100,000 population	-0.18	-0.15	0.23	0.23	-0.20	-0.07	0.23	0.23
Birth rate for teenagers aged 15–19	0.34	0.21	-0.39	-0.36	0.40	0.02	-0.39	-0.38
Average annual pay	-0.11	-0.27	0.25	0.16	-0.15	-0.14	0.23	0.16
Adult secondary education enrollment rate	-0.12	0.13	0.01	0.09	-0.10	0.14	0.01	0.11
Birth rate	0.23	-0.07	-0.13	-0.16	0.18	-0.20	-0.05	-0.13
GED test completion rate	0.13	0.13	-0.18	-0.17	0.18	0.07	-0.22	-0.17
English as a second language enrollment rate	-0.15	-0.33	0.32	0.20	-0.23	-0.18	0.33	0.22
Traffic fatalities per 100 million vehicle miles	0.31	0.25	-0.40	-0.35	0.35	0.09	-0.39	-0.35
Women 15–50 years old who had a birth in the past 12 months	0.10	-0.05	-0.04	-0.06	0.04	-0.09	0.02	-0.03
Average amount of grant and scholarship aid received	-0.26	-0.01	0.20	0.24	-0.26	0.09	0.19	0.24
Graduation rate of postsecondary institutes	-0.01	-0.18	0.12	0.06	-0.08	-0.12	0.15	0.09

See notes at end of table.

Table A-3. PIAAC county- and state-level variable correlations with literacy/numeracy proficiency outcomes: 2012/2014/2017—Continued

Variable	Literacy P1	Literacy P2	Literacy P3	Literacy average	Numeracy P1	Numeracy P2	Numeracy P3	Numeracy average
Homeownership rate	-0.18	0.20	0.02	0.12	-0.14	0.17	0.02	0.13
Infant mortality rate per 1,000 live birth	0.21	0.22	-0.30	-0.24	0.31	0.05	-0.32	-0.29
Average 4th-grade math composite scale scores	-0.23	0.01	0.17	0.22	-0.24	0.06	0.19	0.24
Average 8th-grade math composite scale scores	-0.37	-0.06	0.32	0.35	-0.41	0.07	0.34	0.39
Energy consumption per person	0.23	-0.06	-0.14	-0.18	0.20	-0.13	-0.11	-0.18
State government general revenue per capita	-0.11	-0.25	0.25	0.19	-0.19	-0.08	0.23	0.20
Federal aid to state and local governments per capita	-0.01	-0.07	0.05	0.07	-0.02	-0.06	0.05	0.06
Average 4th-grade reading composite scale scores	-0.22	0.13	0.09	0.18	-0.19	0.12	0.11	0.20
Average 8th-grade reading composite scale scores	-0.41	0.10	0.26	0.35	-0.42	0.19	0.28	0.39
Average weighted monthly salary for full-time instructional staff	-0.34	-0.29	0.33	0.23	-0.25	-0.12	0.32	0.25
Annual college cost (tuition & fees)	-0.25	-0.02	0.21	0.23	-0.24	0.03	0.21	0.24
Violent crime rate per 100,000 population	0.21	-0.15	-0.07	-0.19	0.16	-0.11	-0.09	-0.19

NOTE: P1: proportion at or below Level 1; P2: proportion at Level 2; P3: proportion at or above Level 3; SNAP: Supplemental Nutrition Assistance Program.
SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

Table A-4. PIAAC county- and state-level variable LASSO selection results with literacy/numeracy proficiency outcomes: 2012/2014/2017

	Literacy						Numeracy					
	λ = 0.02		λ = 0.03		λ = 2	λ = 3	λ = 0.02		λ = 0.03		λ = 2	λ = 3
Variable	P1	P3	P1	P3	Avg.	Avg.	P1	P3	P1	P3	Avg.	Avg.
Percentage of population aged 25+: with education less than high school	0.6	-0.5	0.6	-0.5	-109.0	-107.0	0.5	-0.3	0.4	-0.3	-86.7	-88.6
Percentage of population aged 25+: with education more than high school	-0.1	0.4	-0.1	0.4	27.5	22.8	-0.2	0.5	-0.2	0.4	41.2	31.5
Percentage of population below 100 percent poverty line	0.3	-0.3	0.3	-0.3	-31.8	-34.5	0.5	-0.3	0.5	-0.4	-45.4	-59.1
Percentage of Blacks	0.0	0.0	†	†	-1.7	†	0.1	0.0	0.0	0.0	-12.8	-5.3
Percentage of foreign-born people who entered United States after year 2010	0.0	0.0	†	†	1.6	†	†	†	†	†	†	†
Percentage of civilian noninstitutionalized population with no health insurance coverage	0.1	0.0	†	†	-11.7	-6.4	0.2	-0.1	0.1	-0.1	-38.5	-33.7
Birth rate	0.0	0.0	†	†	†	†	†	†	†	†	†	†
Average amount of grant and scholarship aid received	0.0	0.0	†	†	0.0	†	†	†	†	†	0.0	†
Percentage of population born outside of United States	0.0	0.0	†	†	†	†	†	†	†	†	†	†
Unemployment rate	†	†	†	†	0.0	†	†	†	†	†	-0.3	-0.3
Percentage of population aged 16+: service occupation	†	†	†	†	-16.5	-0.7	†	-16.5	-0.7	†	†	†
Percentage of population aged 16+ and didn't work at home: 60+ minutes to work	†	†	†	†	-0.3	†	†	†	-0.3	†	†	†
Percentage of Hispanics	†	†	†	†	†	†	†	†	†	†	-2.0	†

† Not applicable

NOTE: P1: proportion at or below Level 1; P3: proportion at or above Level 3; Avg.: average score.

SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

Appendix B

Simulation Study Results

APPENDIX B

SIMULATION STUDY RESULTS

Simulation studies have frequently been used to evaluate the performance of small area estimators. Although small area estimators are usually supported by well-elaborated statistical models, the models rest on specific assumptions. Simulation studies offer a means to assess the consequences of departures from these assumptions. One strategy in the design of simulation studies is to generate simulated populations based on alternative statistical models. Samples can be drawn from the generated populations and the performance of the small area estimators assessed. An alternative strategy is to treat a very large dataset, such as a census, as the reference population, to draw samples from it, and then to compare the actual performance of the resulting small area estimators to the true area-level values of reference population. The natural appeal of the second strategy, which is the one used here, is that it attempts to represent the possible challenges of applying the small area estimators to data in "the real world" that might not exactly fit any specific set of statistical assumptions. A possible limitation of this approach, however, is in the ability to generalize from the findings of the simulation to the application at hand if the population chosen for the simulation does not approximate key features of the application.

In the summer of 2017, work began on the design of a simulation study based on the 2015 5-year public use file from the American Community Survey (ACS) to evaluate the performance of possible small area estimators under conditions similar to those that will be faced by the Program for the International Assessment of Adult Competencies (PIAAC) application. Most of the initially planned results were obtained by the spring of 2018, before the 2017 PIAAC data were available. A smaller scale supplemental study was conducted in the spring of 2019 to investigate two specific issues that arose during the final implementation of the small area models to the 2012/2014/2017 PIAAC data.

B.1 Overall Design of the Simulation Study

Because of confidentiality restrictions, the geographic detail on ACS public use files is restricted to Public Use Microdata Areas (PUMAs), which are required to have a minimum population size. ACS estimates for each state and most large counties can be assembled by summing the PUMAs within them, but the ACS data for small counties is comingled with other counties, in which case they cannot be individually estimated. In the simulation, the focus on states and counties in PIAAC was modified to evaluate the estimates for states, the large counties that could be assembled from PUMAs, and the remaining ACS PUMAs.

The four stages of sampling in PIAAC were

- sampling primary sampling units (PSUs) within strata, except for a small number of large, certainty PSUs;
- sampling segments within sampled PSUs;
- sampling households within segments; and
- selecting persons within households.

The simulation approximated the PIAAC sampling, combining the second and third stages by artificially forming ACS segments of size equal to the average number of households sampled in the PIAAC third stage, so that the simulation consisted of

- sampling PSUs within strata, except for a small number of large, certainty PSUs;
- sampling ACS segments of households formed for purposes of the simulation; and
- selecting persons within households.

Although the ACS public use file includes weights, the simulation treated this file as the reference population. The simulation weights reflected only the simulated sampling, not the original ACS weights. In both the PIAAC design and the simulation, the last step of sampling is the primary source of variation in the weights. Further details on both the PIAAC and simulation sample designs are presented in the last section.

While the PIAAC universe includes 16- through 74-year-olds in housing units and most noninstitutional group quarters (excluding military barracks and other dwellings on military bases), the simulation was restricted to ACS persons aged 19–74 in housing units-that is, persons in group quarters were excluded-giving a count of 2,089,949 individuals in 1,112,755 households. Households without eligible individuals were excluded from the simulation. Educational attainment was chosen as the dependent variable for the study. To approximate the proportion of adults with literacy Level 1 or below, persons whose highest educational attainment was a GED were combined with those with less than a high school education; together, they are 14.29 percent of the population in the public use file. As an approximation to literacy at Level 2, a second group was formed from the remaining high school graduates plus those who started but did not complete a full year of post-secondary education, representing 29.86 percent of the population in the public use file.

Let s_{ijk} be the set of sampled individuals, l, in sampled households, that are in division i, state j, and PSU k. Direct Hájek (1971) PSU-level estimates, $\hat{p}_{ijk}$, of the proportions in an education group, such as low education, were formed using the weights

$$\hat{p}_{ijk} = \frac{\sum_{l \in s_{ijk}} w_{ijkl} I_{ijkl}}{\sum_{l \in s_{ijk}} w_{ijkl}},$$

where I_{ijkl} is an indicator function for low education for sample individual l in division i, state j, and PSU k, and w_{ijkl} is the survey weight derived as the inverse of the probability of selection considering all stages of selection.

The simulations used the sequence number, 1 to 500, as the seed of the random number generator in R. The random samples based on a given seed are reproducible, so that the simulation can compare the performance of different small area estimators for the same generated samples. For each seed, a sample of PSUs was selected, then a random sample of segments within the PSU was selected with replacement. One or more respondents within each of the households in a sampled segment were selected, and weights were determined for each case. Direct within-PSU variance estimates were computed using segment as the ultimate cluster within the PSU and the weights described above, using the standard linearization formula for a ratio estimator. The variance calculation was performed with the function svymean() from the survey package in R, https://cran.r-project.org/web/packages/survey/survey.pdf.

Small area models. Only area-level small area models were simulated. The PSU-level predictors, treated as fixed effects, are

1. the log of mean household income;
2. the proportion of U.S.-born persons;
3. the proportion of Hispanic persons;
4. the proportion of Black persons; and
5. the proportion working.

All predictors are based on the unweighted full ACS public use file, so they have no sampling error in the simulation setting. Some models also introduce census division, primarily as random effects but also as fixed effects in two models. As defined by the Census Bureau, the census divisions (New England, Middle Atlantic, East North Central, West North Central, South Atlantic, East South Central, West South Central, Mountain, and Pacific divisions) are groups of states. For example, New England comprises Maine, New Hampshire, Vermont, Massachusetts, Rhode Island, and Connecticut. All divisions were represented by one or more sampled PSUs in each of the 500 simulation samples. The notion behind including division in the model was that census division provides some geographic information about PSUs in states with no sampled PSUs in a given sample.

Factors studied. Thus far, the models are all Hierarchical Bayes (HB) implemented through Markov Chain Monte Carlo (MCMC) methods, the same class of models used for the National Assessment of Adult Literacy (NAAL). The models vary

- the form of the small area model, comparing the unmatched model used for NAAL to a matched linear model for the proportion;
- the method of smoothing the estimated sampling variances;
- whether to use the standard weighted estimate of the PSU-level proportion or a survey regression estimate;
- the univariate model for low education similar to one used in NAAL or a bivariate version jointly modeling low education and high school education; and
- whether to include census division as either a random or fixed effect in the model.

A smaller number of simulations were more recently run to investigate

- the effect of the LKJ (Lewandowski, Kurowicka, and Joe) prior distribution in fitting the multivariate model, in place of the historically more traditional inverse Wischart; and
- setting the parameters of the STAN software to more stringent standards to improve the diagnostics for the MCMC chains.

The rationale for examining these two features was to assess the degree to which the simulation results remained applicable to the final estimation approach.

The principal results can be summarized as follows:

- The matched linear model generally performed better than an unmatched model;
- When the small area estimators included modeled variances in order to smooth them, the performance was almost comparable to estimates that would result from knowing the true variances;

- The survey regression estimate provided a more effective starting point for the area-level models than the usual direct survey estimate;
- A bivariate version of a multivariate model jointly modeling low education and high school education generally improved the predictions for low education compared to a univariate model for low education; and
- The best performing model, which included census division as a random effect, was more effective than a similar model treating division as a fixed effect.

One model, M1, was designed to closely follow the unmatched model used for NAAL. First, a fixed-effect logistic regression was used to predict the proportions $\tilde{p}_{ijk}$ of the true proportions p_{ijk} of low education for sampled PSUs k in state j, division i based on the direct estimates $\hat{p}_{ijk}$. Then, the log of the observed relvar (relative variance), $\mathrm{var}(\hat{p}_{ijk})/\hat{p}_{ijk}^2$, was predicted in a weighted regression with terms log($\tilde{p}_{ijk}$), log(1-$\tilde{p}_{ijk}$) and log($n_{seg,ijk} - 1$), using weights $n_{seg,ijk} - 1$, where $n_{seg,ijk}$ is the number of segments in the PSU k. Observations with a direct estimate $\hat{p}_{ijk} = 0$ or a relvar of 0 were excluded from the fitting. But observations with $\tilde{p}_{ijk} > 0$ can be assigned an estimated relvar based on the model. As in the NAAL small area estimation, the estimated relvar is kept fixed during the MCMC cycles, letting the distribution of the variance of $\hat{p}_{ijk}$ depend on the distribution of p_{ijk} through the initially modeled estimate of its relvar.

As in chapter 2, the sampling model is

$$\hat{p}_{ijk} = p_{ijk} + e_{ijk},$$

where $e_{ijk} \sim N(0, p_{ijk}^2\, relvar_{ijk})$. Using a logit link function $z_{ijk} = \ln(p_{ijk}/(1 - p_{ijk}))$, the model is

$$z_{ijk} = \beta_0 + \beta_1 x_{1ijk} + \beta_2 x_{2ijk} + \beta_3 x_{3ijk} + \beta_4 x_{4ijk} + \beta_5 x_{5ijk} + v_{ij} + u_{ijk}, \quad \text{(B.1)}$$

where $v_{ij} \sim N(0, \sigma_s^2)$, $u_{ijk} \sim N(0, \sigma_c^2)$ are random state and PSU level effects, respectively, and $x_{1ijk}, x_{2ijk}, \ldots x_{5ijk}$ are the five predictors mentioned previously.

Many of the models considered are based instead on a matched linear model

$$p_{ijk} = \beta_0 + \beta_1 x_{1ijk} + \beta_2 x_{2ijk} + \beta_3 x_{3ijk} + \beta_4 x_{4ijk} + \beta_5 x_{5ijk} + v_{ij} + u_{ijk}, \quad \text{(B.2)}$$

with the same set of predictors.

Except for three models, the models first applied some form of smoothing to the direct estimates of the sampling variance within PSU. The variance smoothing for M1 was described above. A second model, M2, was similar to M1 but used a different form of smoothing. In place of a model for relvar, a model for the log of the effective sample size, $neff_{ijk} = \hat{p}_{ijk}(1 - \hat{p}_{ijk})/\mathrm{var}(\hat{p}_{ijk})$ based on the direct variance estimate $\mathrm{var}(\hat{p}_{ijk})$, was fitted with a weighted regression using $\ln(\tilde{p}_{ijk})$ and $\ln(n_{ijk})$, where n_{ijk} is the number of sampled individuals in the PSU,

$$\ln(neff_{ijk}) = \beta_0 + \beta_1 \ln(\tilde{p}_{ijk}) + \beta_2 \ln(n_{ijk}) + e'_{ijk},$$

and where e'_{ijk} is an error term with mean 0. The regression used weights proportional to $n_{seg,ijk} - 1$.

The variance smoothing for the matched linear models based on the direct estimates $\hat{p}_{ijk}$ was similar. In place of logistic regression, a fixed effect linear regression is used to predict $\tilde{p}_{ijk}$ for each sampled PSU, converting any predicted value < 0.01 to 0.01. Similar to M2, the log of the observed effective sample size, $neff_{ijk} = \hat{p}_{ijk}(1 - \hat{p}_{ijk})/\mathrm{var}(\hat{p}_{ijk})$ was modeled with a weighted regression using $\ln(\tilde{p}_{ijk})$ and $\ln(n_{ijk})$ as predictors, where n_{ijk} is the number of sampled individuals in the PSU, with weights proportional to $n_{seg,ijk} - 1$. The regression equation and estimated $\tilde{p}_{ijk}$ were then used to predict an effective sample size. The estimated sampling variance, $\tilde{p}_{ijk}(1 - \tilde{p}_{ijk})/\widetilde{neff}_{ijk}$, was then kept fixed during the MCMC simulation.

The true within-PSU sampling variance was estimated by simulating the sampling for 2000 samples in each PSUs and computing the variance of the estimated proportion. Two models (M3 and M5 described later) used these estimated true variances in place of modeled variances.

Some of the linear models, including the best performing one, used a survey regression estimate (or "modified generalized regression [GREG] estimate"), $\check{p}_{ijk}$, of the PSU proportion in place of the simple direct estimate, $\hat{p}_{ijk}$. Rao and Molina (2015, pp. 21-23) described the use of these estimates in small area estimation, their derivation, and the usual Taylor series approach to estimating their variance. In addition to the x predictors available for the sampled cases, the values of the population totals $\mathbf{X}_{ijk}$ in state j and PSU k must be available for this estimator. In the simulation, these were provided by the full ACS. An indicator of low education at the person level was modeled by a linear model using indicators for working, black, Hispanic, U.S. born, and indicators for intervals of household income (loss-$19,999, $20,000–$29,999, $30,000–$39,999, $40,000–$49,999, $50,000–$59,999, $60,000–$69,999, $70,000–$79,999, $80,000–$89,999, $90,000–$99,999, $100,000–$124,999, and $125,000 and above.) The indicator variables for income were selected to replace the log of mean income at the area level because individual income can be negative or zero. Low education did not appear to vary with income below $20,000.

In the modified form proposed by Särndal and Hidiroglou (1989), the survey regression estimate can be written

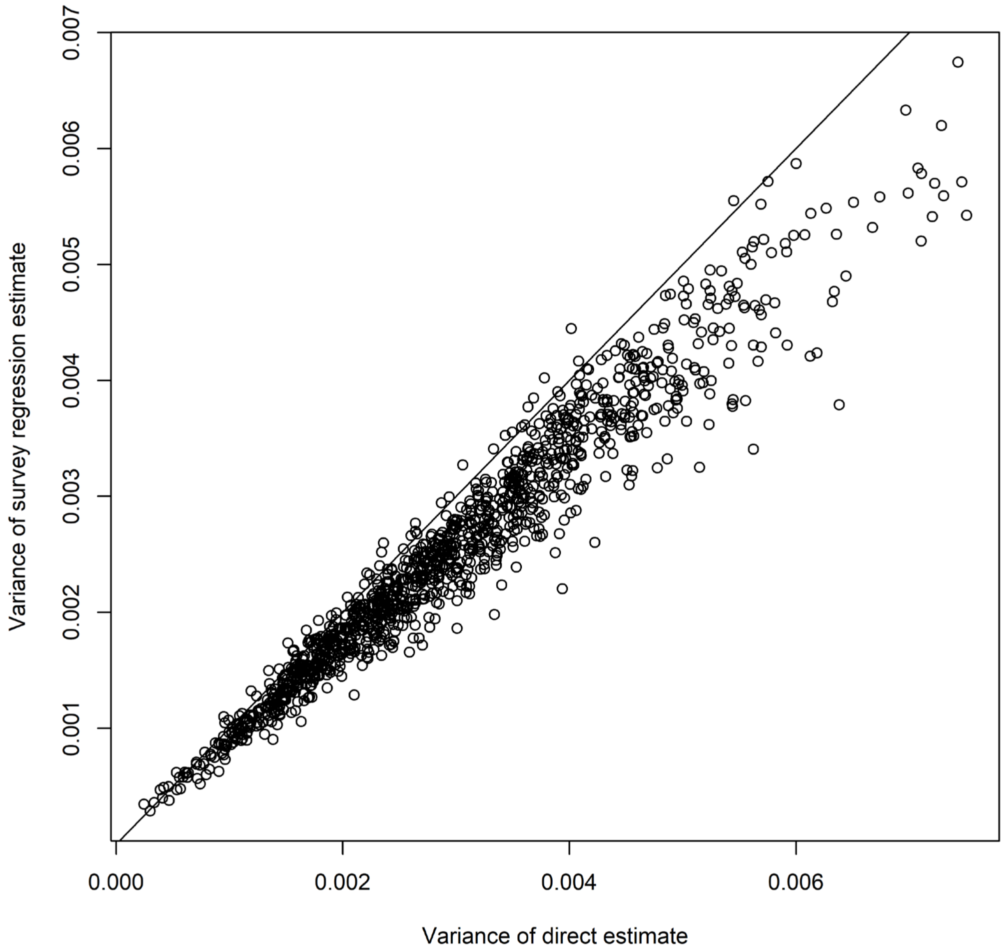
$$\check{p}_{ijk} = \mathbf{X}_{\boldsymbol{ijk}}^{\boldsymbol{T}}\widehat{\mathbf{B}}/N_{\boldsymbol{ijk}} + \textstyle\sum_{\boldsymbol{l}\in s_{\boldsymbol{ijk}}} w_{\boldsymbol{ijkl}} e_{ijkl} / \sum_{l\in s_{ijk}} w_{ijkl},$$

where $\widehat{\mathbf{B}}$ is the vector of survey-weighted regression coefficients from the unit-level regression based on the whole sample, N_{ijk} is the known size of the eligible population in the PSU, the e_{ijkl} are the unit-level residuals from the regression fit in state j and PSU k, and the w_{ijkl} are the corresponding survey weights. The Taylor series variance estimate is based on applying the standard variance expression (in this case appropriate for the simple random sampling with replacement [SRSWR] sampling of segments within PSU) to the e_{ijkl}. As is well known, GREG estimators can occasionally yield negative weights in some cases, leading to negative estimates of nonnegative characteristics (Rao and Molina 2015, p.14). The survey regression estimator did so during the simulations for a small number of PSUs, and in the simulations negative estimates were replaced by small positive quantities or excluded from the modeling.

The variance modeling was modified slightly for the survey regression estimate. The observed effective sample size $neff_{ijk} = \check{p}_{ijk}(1 - \check{p}_{ijk})/\text{var}(\check{p}_{ijk})$ was again modeled with a weighted regression using $\ln(\tilde{p}_{ijk})$ and $\ln(n_{ijk})$, with weights proportional to $n_{seg,ijk} - 1$.

The true sampling variance of the PSU-level survey regression estimates was estimated by simulating the sampling 40,000 times, including the estimation of $\widehat{\mathbf{B}}$, and finding the variance of the PSU-level estimates for those samples including the PSU. Figure B-1 compares the variances of the estimates of $\hat{p}_{ijk}$ and $\check{p}_{ijk}$. On average, the variance of the survey regression estimate is about 14 percent lower.

Figure B-1. Comparison of estimated true variances of the direct and survey regression estimates for the 1,234 PSUs in the simulation



NOTE: The line $y = x$ is included for comparison. PSU: primary sampling unit.
SOURCE: U.S. Department of Commerce, U.S. Census Bureau, American Community Survey, 2011–2015.

The ratio of average variance for the two estimators is shown in table B-1, disaggregated by groupings of PSUs. The first distinction is between the 803 PSUs of individual PUMAs that were not uniquely assigned to counties, many of which contained multiple small counties, and the 431 PSUs corresponding to large individual counties, divided by size into two noncertainty groups and the 10 certainty PSUs in the simulation. The table indicates that the survey regression estimate yields the greatest relative improvement in the largest counties, but average improvements are present in the other groups of PSUs. Comparisons are also made according to the level of low education in two ways. In the first, the PSUs are grouped according to the actual low education values. In the second, the PSUs are grouped according to the predicted proportion from the linear regression of low education on the same fixed effects as in equation (B.2). The first comparison stems from a finite population perspective, whereas the second brings more of a modeling perspective to the comparisons. Note that the distribution of PSUs by ACS low-education level is somewhat more dispersed than the distribution of the modeled estimates, as might be expected. These two perspectives produce relatively similar outcomes for the ratios of average variance in table B-1, showing variance reductions for all of the groups.

Table B-1. Ratio of the estimated true sampling variance of the PSU-level survey regression estimate to the average direct estimate for the simulation PSUs

PSU grouping	Number of PSUs	Population aged 19–74 (in millions)	Ratio of average variance
PSUs by type and size			
Split counties	803	73.3	0.866
Counties w/pop 19–74 < 500,000	353	61.9	0.851
Counties w/pop 19–74 > 500,000	68	53.9	0.788
Certainty counties	10	28.2	0.755
PSUs by ACS percent low education			
0–4.9 percent	41	5.1	0.902
5.0–9.9 percent	251	43.6	0.886
10.0–4.9 percent	438	87.9	0.856
15.0–19.9 percent	290	44.7	0.860
20.0–24.9 percent	159	28.6	0.842
25.0–49 percent	55	7.4	0.841
PSUs by low predicted education			
0–4.9 percent	35	4.1	0.885
5.0–9.9 percent	187	38.3	0.858
10.0–14.9 percent	461	80.0	0.856
15.0–19.9 percent	423	66.3	0.862
20.0–24.9 percent	106	22.5	0.851
25.0–49 percent	22	6.1	0.816

NOTE: ACS: American Community Survey; PSU: primary sampling unit.
SOURCE: U.S. Department of Commerce, U.S. Census Bureau, American Community Survey, 2011–2015.

The small area model for NAAL produced estimates for a single characteristic, at or below Level 1 literacy, but a multivariate approach permits expansion of the analysis to simultaneously produce estimates for the proportions at or below Level 1 literacy, for Level 2 literacy, and for Level 3 literacy and above. As previously noted, high school education is used as a proxy for Level 2 literacy in the simulations. The multivariate models used the same predictors for both education levels, but the beta coefficients for the two levels were not forced to be equal and indeed were quite different.

All of the models were implemented through HB using MCMC methods in the same basic way. Models were fitted in R by calling either the JAGS or STAN software. For JAGS, the protocol for NAAL was followed: three chains with burn-in of 10,000 and total of 100,000 for each chain, which was sampled at 1 in 10 for analysis. All models used diffuse uniform (-1000, 1000) priors on the betas and gamma priors on the precision (reciprocal variance) of the state and PSU variances with shape and inverse scale parameters 0.001. For STAN, the three chains were shortened to a burn-in of 2,000 and total of 20,000, with again 1 in 10 sampling. Univariate models and multivariate used diffuse N(0, 100) priors on the betas. Univariate models used inverse gamma priors for the state and count variances with shape and scale parameters 0.001 for the variance components. Multivariate models used inverse Wishart distributions each with 2 degrees of freedom; the scale matrix for the PSU covariance was taken to have elements .001156, .000098, .000098, .001156. For states and divisions, the elements were .000576, .000289, .000289, .000961.

The hierarchical model was fitted to the 147 sampled PSUs, with random effects fitted for those states with PSUs in sample. The number of observed states varied from simulation to simulation. The resulting chains for the 147 PSUs and observed states were expanded into chains for all 1,234 PSUs and 50 states plus DC. In nonsample PSUs in sampled states, equation (B.1) or equation (B.2) was computed for each cycle of the chain from the current values of the βs, the current value of ν_i for the state, and a random draw $u_{ijk} \sim N(0, \sigma_c^2)$, based on the current value of σ_c^2. For states without any PSUs in sample, a value of ν_{ij} was drawn from $\nu_{ij} \sim N(0, \sigma_s^2)$ based on the current value of σ_s^2 and used in equation (B.1) or equation (B.2) for each PSU in the state. For the unmatched model, equation (B.1), the calculations were performed on the logistic scale, and then the result was transformed back to predicted proportions. State proportions were created by weighting their PSU proportions by population aged 19–74.

Models considered. Models M1, the closest analogue to the NAAL model, and M2, with the alternative variance smoothing, have already been introduced. Except where noted, models M1-M7 were fitted with JAGS and the remaining models with STAN. The other models are as follows:

- M3: the same estimator as M1 or M2 if the within-sampling variances are known, using the variance estimates from the 2,000 simulated samples in each PSU;
- M4: the linear model based on $\hat{p}_{ij}$;
- M4S: M4 implemented in STAN;
- M5: M4 using the variance estimates from the 2,000 simulated samples in each PSU;
- M6: the linear model implemented on the survey regression estimate, $\check{p}_{ij}$;

- M6S: M6 implemented in STAN;
- M7: M6 using the within-PSU variance estimates from the 40,000 simulated samples;
- M8: the multivariate extension of M4, with modifications to the variance smoothing noted below;
- M9: the multivariate extension of M6;
- M10: an extension of M4 with random effects for census division;
- M11: a multivariate version of M10;
- M12: an extension of M6 with random effects for census division;
- M13: a multivariate version of M12;
- M14: an extension of M6 with fixed effects for census division; and
- M15: a multivariate version of M14.

Except for the procedure to smooth the variances, table B-2 summarizes these choices.

Table B-2. Features of the models studied, comparing unmatched, "x" (vs. matched, "."), use of the estimated true variance, "x" (vs. smoothed "."), STAN, "x" (vs. JAGS, "."), use of the survey regression estimate, "x" (vs. Hájek "."), multivariate, "x" (vs. univariate, ".") and inclusion, "fixed" or "random" of census division effect in the model (vs. not ".")

Model	Unmatched	True variance	STAN	Based on $\check{p}_{ij}$	Multivariate	Census division
M1	x	.	.	.	.	.
M2	x	.	.	.	.	.
M3	x	x	.	.	.	.
M4	.	.	.	.	.	.
M4S	.	.	x	.	.	.
M5	.	x	.	.	.	.
M6	.	.	.	x	.	.
M6S	.	.	x	x	.	.
M7	.	x	.	x	.	.
M8	.	.	x	.	x	.
M9	.	.	x	x	x	.
M10	.	.	x	.	.	random
M11	.	.	x	.	x	random
M12	.	.	x	x	.	random
M13	.	.	x	x	x	random
M14	.	.	x	x	.	fixed
M15	.	.	x	x	x	fixed

SOURCE: Author's definition.

In the multivariate models an additional predictor was added to the individual-level regression used in the survey regression estimator: an indicator variable for household income $125,000–$174,999. The motivation for the addition was to further reflect the income distribution when modeling higher education levels.

The multivariate models required estimates of the sampling covariance matrix within PSU for the two components of education. The variance for each of the three education proportions was modeled separately, giving $var_{ijk}^{(1)}$ for the modeled variance of low education, $var_{ijk}^{(2)}$ for the modeled variance of high school education, and $var_{ijk}^{(3)}$ for the modeled variance of higher education, that is, 1 or more years of college. In general, if three proportions with respective variances $var^{(1)}, var^{(2)}$ and $var^{(3)}$ sum to 1, then The variance of the third proportion must equal the variance of the sum of the first two, or

$$var^{(3)} = var^{(1)} + var^{(2)} + 2cov^{(12)}.$$

A modeled covariance was derived by

$$cov_{ijk}^{(12)} = .5 \times (var_{ijk}^{(3)} - var_{ijk}^{(1)} - var_{ijk}^{(2)}).$$

After some initial experimentation, each of the variance models were simplified to modeling the log of the effective sample size as

$$\ln(neff_{ijk}) = \beta_0 + \beta_1 \ln(n_{ijk}) + e'_{ijk}.$$

Two previous variance models including a term or terms involving p_{ijk} occasionally produced covariance matrices that were not positive definite for one or more of the simulated samples. The simplified approach yielded satisfactory results for all 500 simulation samples.

B.2 Results

Mean Square Error (MSE) Results at the PSU Level. Table B-3 presents average MSE for PSUs overall and for the groupings of PSUs by population, table B-4 gives results for groupings by levels of actual low education, and table B-5 for groupings by levels of modeled low education.

Table B-3. Average mean square errors ($\times 10^4$) based on 500 simulated samples under different HB area-level models for low education, averaged over all PSUs and for groups of PSUs, classified by the size of the population aged 19–74

Model	Overall	PSU type, population aged 19–74			
		Split counties	Population < 500,000	Population 500,000+, noncertainty	Certainty
M1	13.19	14.16	12.33	6.34	12.79
M2	13.09	14.09	12.11	6.46	12.16
M3	13.36	14.57	12.09	5.67	13.24
M4	12.37	13.38	11.36	6.14	9.96
M4S	12.36	13.36	11.36	6.13	9.89
M5	12.55	13.82	11.16	5.26	9.74
M6	12.06	13.09	11.07	5.60	8.82
M6S	12.07	13.08	11.09	5.65	8.77
M7	12.26	13.53	10.89	5.01	8.74
M8	12.16	13.29	10.94	5.36	10.52
M9	11.88	13.02	10.70	4.97	9.45
M10	12.40	13.17	11.74	7.02	10.05
M11	11.92	12.85	11.01	5.88	10.47
M12	12.02	12.81	11.37	6.42	8.88
M13	11.59	12.53	10.71	5.42	9.40
M14	12.62	13.37	12.05	7.31	9.12
M15	12.24	13.02	11.61	6.74	9.63

NOTE: HB: Hierarchical Bayes; PSU: primary sampling unit.
SOURCE: U.S. Department of Commerce, U.S. Census Bureau, American Community Survey, 2011–2015.

Table B-4. Average mean square errors ($\times 10^4$) based on 500 simulated samples under different HB area-level models for low education, for groups of PSUs, classified by the percentage of low education

	Actual percentage of low education					
Model	0–4 percent	5–9 percent	10–14 percent	15–19 percent	20–24 percent	25 percent+
M1	10.26	8.76	8.14	12.04	23.95	50.85
M2	10.62	8.65	7.88	11.85	24.36	50.62
M3	8.07	7.05	7.36	13.10	27.42	54.53
M4	7.37	9.61	8.27	9.56	21.84	48.82
M4S	7.37	9.63	8.29	9.55	21.74	48.72
M5	6.47	8.37	7.51	10.20	24.58	54.00
M6	7.09	9.35	8.01	9.30	21.44	47.88
M6S	7.09	9.38	8.05	9.30	21.35	47.74
M7	6.34	8.23	7.32	9.91	24.09	52.62
M8	6.98	9.18	7.63	9.18	22.61	51.18
M9	6.68	8.98	7.45	8.95	22.15	50.05
M10	7.90	9.88	8.69	9.98	20.79	45.20
M11	7.05	9.16	7.72	9.20	21.52	48.14
M12	7.56	9.55	8.36	9.65	20.31	44.13
M13	6.73	8.90	7.49	8.94	21.00	46.88
M14	8.63	10.21	9.15	10.55	20.41	42.58
M15	8.27	9.73	8.52	10.09	20.55	43.62

NOTE: HB: Hierarchical Bayes; PSU: primary sampling unit.
SOURCE: U.S. Department of Commerce, U.S. Census Bureau, American Community Survey, 2011–2015.

Table B-5. Average mean square errors ($\times 10^4$) based on 500 simulated samples under different HB area-level models for low education, for groups of PSUs, classified by the modeled estimate of the percentage of low education

	Modeled percentage of low education					
Model	0–4 percent	5–9 percent	10–14 percent	15–19 percent	20–24 percent	25 percent+
M1	5.64	5.59	9.58	15.47	26.35	58.33
M2	5.95	5.66	9.65	15.64	25.64	50.15
M3	4.33	5.22	10.16	16.36	25.73	46.56
M4	4.81	6.19	9.64	15.19	23.40	26.97
M4S	4.81	6.17	9.64	15.17	23.37	27.04
M5	4.65	5.81	9.64	15.70	24.20	27.02
M6	4.50	5.91	9.42	14.90	22.80	25.49
M6S	4.52	5.93	9.43	14.89	22.77	25.59
M7	4.32	5.60	9.44	15.40	23.65	25.44
M8	4.20	5.67	9.25	15.20	23.64	27.13
M9	3.93	5.47	9.07	14.92	23.06	25.78
M10	5.82	6.81	10.03	14.92	21.68	26.65
M11	4.70	5.92	9.31	14.70	22.06	26.57
M12	5.44	6.48	9.75	14.55	20.99	25.05
M13	4.41	5.67	9.09	14.37	21.39	25.12
M14	6.71	7.43	10.51	15.05	20.53	25.56
M15	6.24	6.98	10.04	14.73	20.30	26.02

NOTE: HB: Hierarchical Bayes; PSU: primary sampling unit.
SOURCE: U.S. Department of Commerce, U.S. Census Bureau, American Community Survey, 2011–2015.

Because the results are based on only 500 simulated samples, many but not all comparisons between models are statistically significant. Furthermore, in some instances where one model gives better overall performance, the results are not uniform across subsets of the PSUs. Two results stand out immediately, however,

- Even though some of the differences between M4 and M4S and between M6 and M6S are statistically significant, the MSE results are for all practical purposes identical, so these two available comparisons suggest that JAGS and STAN were producing essentially equivalent results for these models.

- Comparisons between M4 and M6, between M8 and M9, between M10 and M12, and between M11 and M13 each reflect the effect of the direct versus the survey regression estimator. The comparisons favor the survey regression estimator for each of the 4 pairs and each of the 17 comparisons within each pair. The values of the 68 t-statistics range from a low of 4.04 to a high of 35.1; in other words, all are highly significant. The simulation results uniformly favor use of the survey regression estimate over the direct estimate.

Three other results show overall differences but somewhat inconsistent results:

- Comparison of M4 and M8, M6 and M9, M10 and M11, and M12 and M13 each assess the impact of using a multivariate model on the quality of the estimates for low education. There is a significant overall gain from the multivariate model, particularly when census division is a random effect, and gains appear in most but not all of the groups, with exceptions for the certainty PSUs (not significant) and for the highest proportions of low education (significantly worse).

- Including random effects for census divisions appears to improve the estimation in some circumstances but not others. Here, the pairings are M4 with M10, M6 with M12, M8 with M11, and M9 with M13. The first two pairings of univariate models yield nonsignificant overall comparisons and significant differences for some subgroups with both positive and negative signs. The overall improvements are significant for the pairings of multivariate models, M8 with M11 and M9 with M13, and most but not all of the significant findings for subgroups of PSUs are in the direction of showing a benefit for the use of random effects at the division level.

- Expressing the division effect as a random effect in the model rather than a fixed effect is substantially better overall and for most subsets of the PSUs, except for differences that are not significant for certainty PSUs and for differences that are not significant or reversed for PSUs with the highest levels of low education.

Two other results seem entirely mixed:

- The comparison of M1 and M2 is not statistically significant overall and produces a mixture of positive, negative, and not significant differences for the subgroups considered.

- Somewhat surprisingly, the substitution of estimated true variances for modeled variances does not consistently improve the results over the modeled variance. In fact, overall M2 is statistically significantly better than M3, but the comparison reverses in some of the subgroups. Similarly, M4 outperforms M5 for overall MSE, as does M6 over M7, but again the comparison reverses for some of the subgroups. The simulation results do not indicate an appreciable loss from using smoothed variances in place of the unknown true variances.

The unmatched model has an advantage over the matched linear model by guaranteeing positive estimates when applied to the direct survey estimates. The chance for negative estimates may be increased further by use of the survey regression estimate, which itself may give negative estimates. For model M13, which includes use of the survey regression estimates, the average number of negative HB estimates among the 147 sampled PSUs was 0.5, and the average was 1.1 among all of the 1,234 PSU HB estimates. The simulation results suggest that the magnitude of the problem is small, but that it should be addressed in advance. One approach would be to assign any negative estimate a value known to be small in the population, such as 0.02, or alternatively the smallest positive HB estimate among the entire set of 1,234. This one disadvantage of the matched model should be assessed in the context of its other advantages, such as its possible use of the survey regression estimator and extensions to multivariate estimation. (It may be possible to develop a multivariate extension of the unmatched model, but this was not attempted.)

For the 147 PSUs in sample, one might ask how often the HB estimate is better than the sample estimate. In fact, the simulation shows that the conditional variance of a sampled PSU is smaller than the estimated average MSE of the HB estimate when the PSU is in sample in 33 of the 1,234 PSUs, including 1 of the 10 certainty PSUs. (Random error from the simulation may affect these results, because the conditional MSE of some individual PSUs may be based on fewer than 50 simulation samples.) Because it is unlikely that a rule could be developed to successfully identify the small number of PSUs better off with only their direct estimates, the HB estimates may be viewed as offering improvement for all but a small proportion of the PSUs.

MSE Results at the State Level. The twelve largest states in 2015 included almost 60 percent of the population aged 19–74. These states are California, Florida, Georgia, Illinois, Michigan, New Jersey, New York, North Carolina, Ohio, Pennsylvania, Texas, and Virginia. Table B-6 compares state-level MSE results. The MSEs are systematically lower at the state level than at the PSU level. Interestingly, the unmatched models appear to produce better state-level estimates than many of the matched models, but the best matched models are competitive, with M13 again appearing best overall.

Table B-6. Average mean square errors ($\times 10^4$) based on 500 simulated samples under different HB area-level models for low education, state-level estimates

Model	Overall	Smaller states	Largest 12 states
M1	3.04	3.34	2.08
M2	3.02	3.34	2.00
M3	3.03	3.31	2.11
M4	3.44	3.78	2.31
M4S	3.43	3.78	2.31
M5	3.33	3.68	2.20
M6	3.23	3.58	2.09
M6S	3.25	3.60	2.10
M7	3.17	3.53	2.00
M8	3.11	3.42	2.12
M9	2.95	3.26	1.93
M10	3.71	4.02	2.69
M11	3.03	3.24	2.33
M12	3.45	3.77	2.41
M13	2.83	3.05	2.11
M14	4.39	4.91	2.71
M15	3.92	4.36	2.51

NOTE: HB: Hierarchical Bayes.
SOURCE: U.S. Department of Commerce, U.S. Census Bureau, American Community Survey, 2011–2015.

Estimates of the National Total. As noted previously, the national low education rate for the simulated population based on weighting the unweighted PSU-level rates by the estimated population aged 19–74 was 14.22 percent. Except for M3 (13.70 percent), M5 (13.84 percent), and M7 (13.90 percent)—the three models using estimated true variances—all other models produced national rates with means within 0.10 percentage points of 14.22 percent. (The values were 14.28 percent, 14.20 percent, 14.17 percent, 14.17 percent, 14.17 percent, 14.17 percent, 14.15 percent, 14.15 percent, 14.17 percent, 14.15 percent, 14.17 percent, 14.15 percent, 14.17 percent, and 14.15 percent, for M1, M2, M4, M4S, M6, M6S, M8, M9, M10, M11, M12, M13, M14, and M15, respectively.)

Interval results. Because the simulation is based on a single population, the proportion of times the HB credible interval covers an individual true value will be different from the nominal level. The true value for some PSUs may lie close to the model prediction, and for these PSUs the credible intervals may cover at an average rate above the nominal value. In the case of PSUs with outlier values compared to the assumed model, the credible intervals may cover the true values at a rate lower than the nominal level. But from a frequentist perspective, a set of intervals may be judged as successful for a fixed population if their actual coverage is close to the nominal coverage when averaged over groups of areas. Table B-7 presents the average coverage of the 95 percent credible intervals from the MCMC overall and for PSUs classified by PSU type and population. In this table, the performance appears relatively satisfactory for all of the models. The large, noncertainty PSUs are covered at a rate above the nominal level, whereas coverage is a bit below for the certainty. M2 appears to improve on M1 in terms of coverage.

Table B-7. Average percentage of coverage of 95 percent credible intervals produced by MCMC based on 500 simulated samples under different HB area-level models for low education, averaged over all PSUs, and classified by the population aged 19–74 of the PSU

		PSU type, population aged 19–74			
Model	Overall	Split PSUs	Population < 500,000	Population 500,000+, noncertainty	Certainty
M1	91.9	91.8	91.3	96.8	91.9
M2	95.1	95.0	94.9	98.0	93.5
M3	90.8	90.5	90.5	96.9	89.5
M4	95.6	95.4	95.7	98.3	93.2
M4S	95.7	95.5	95.7	98.3	93.1
M5	93.9	93.3	94.4	98.8	93.0
M6	95.1	94.8	95.2	98.5	93.4
M6S	95.3	95.0	95.4	98.4	93.5
M7	93.6	92.9	94.1	98.7	92.7
M8	93.5	92.9	94.0	98.1	90.0
M9	93.2	92.5	93.7	98.2	90.2
M10	96.1	96.1	95.9	98.0	93.5
M11	94.1	93.8	94.2	97.8	90.5
M12	95.7	95.6	95.4	98.1	93.6
M13	93.7	93.3	93.9	97.8	90.8
M14	95.8	95.8	95.5	97.7	93.3
M15	94.1	94.0	93.9	97.1	91.1

NOTE: HB: Hierarchical Bayes; MCMC: Markov Chain Monte Carlo; PSU: primary sampling unit.
SOURCE: U.S. Department of Commerce, U.S. Census Bureau, American Community Survey, 2011–2015.

Tables B-8 and B-9 provide the same groupings of PSUs as tables B-4 and B-5. Table B-8 shows that the unmatched HB models provide low true coverage for areas with extremely low levels of low education. For the same PSUs, the matched models cover at rates far above the nominal level, suggesting that their intervals are too wide there. At the other end of the scale, when the proportion of low education is 25 percent or more, the coverage of intervals for matched models is also poor, where the unmatched results are somewhat better. The range 20–24 percent remains challenging for all of the methods, and the best performance is seen in the middle of the distribution. The results in table B-9 appear somewhat better, with improvements for the unmatched models for PSUs with modeled estimates less than 5 percent and for matched models for PSUs with modeled estimates of 25 percent or more. The difference between table B-8 and table B-9 can be understood as a consequence of regression to the mean. In other words, PSUs with extreme values will be more difficult to model. When PSUs are averaged over groups based on their predicted values under a model, the regression effect is mitigated.

Table B-8. Average percentage of coverage of 95 percent credible intervals produced by MCMC based on 500 simulated samples under different HB area-level models for low education, classified by the percentage of low education in the PSU

	Actual percentage of low education					
Model	0–4 percent	5–9 percent	10–14 percent	15–19 percent	20–24 percent	25 percent+
M1	57.6	90.2	96.2	94.6	89.5	83.8
M2	66.7	94.7	98.1	97.0	93.4	88.5
M3	63.8	92.1	95.6	91.7	85.3	78.9
M4	99.2	97.4	98.1	97.8	89.7	71.1
M4S	99.3	97.4	98.1	97.8	89.8	71.1
M5	99.2	97.1	97.9	96.4	83.5	61.0
M6	99.0	97.0	97.9	97.5	88.4	68.6
M6S	99.2	97.1	97.9	97.6	88.9	69.1
M7	99.1	96.8	97.7	96.2	82.7	59.8
M8	98.7	96.0	97.3	96.5	83.5	60.6
M9	98.7	95.8	97.1	96.3	82.8	59.6
M10	99.2	97.5	98.2	97.9	91.0	75.6
M11	98.8	96.3	97.5	96.8	85.5	64.8
M12	99.2	97.2	98.0	97.6	90.0	73.7
M13	98.7	95.9	97.2	96.5	84.6	63.4
M14	98.9	97.2	97.9	97.5	90.5	76.4
M15	98.2	95.9	97.1	96.3	96.4	69.8

NOTE: HB: Hierarchical Bayes; MCMC: Markov Chain Monte Carlo; PSU: primary sampling unit.
SOURCE: U.S. Department of Commerce, U.S. Census Bureau, American Community Survey, 2011–2015.

Table B-9. Average percentage of coverage of 95 percent credible intervals produced by MCMC based on 500 simulated samples under different HB area-level models for low education, classified by the modeled estimate of the percentage of low education in the PSU

	Modeled percentage of low education					
Model	0–4 percent	5–9 percent	10–14 percent	15–19 percent	20–24 percent	25 percent+
M1	72.4	90.4	92.5	93.2	93.7	89.9
M2	78.2	93.9	95.7	96.1	96.4	95.0
M3	76.7	90.4	90.9	91.7	92.8	90.9
M4	99.3	99.0	96.9	94.2	90.0	89.7
M4S	99.6	99.1	96.9	94.2	90.0	89.3
M5	99.6	98.9	95.8	91.5	86.2	87.7
M6	99.4	98.9	96.6	93.4	88.7	89.1
M6S	99.6	99.0	96.7	93.6	88.9	89.0
M7	99.6	98.8	95.6	91.0	85.4	87.0
M8	99.5	98.6	95.6	90.8	85.3	86.0
M9	99.4	98.6	95.4	90.3	84.7	85.7
M10	99.4	98.9	97.0	94.9	91.9	90.2
M11	99.3	98.6	95.8	91.9	87.4	86.8
M12	99.4	98.9	96.7	94.3	90.9	89.7
M13	99.3	98.5	95.5	91.3	86.5	86.5
M14	99.2	98.7	96.7	94.5	91.9	90.1
M15	98.8	98.1	95.5	92.1	88.8	86.7

NOTE: HB: Hierarchical Bayes; MCMC: Markov Chain Monte Carlo; PSU: primary sampling unit.
SOURCE: U.S. Department of Commerce, U.S. Bureau of the Census, American Community Survey, 2011–2015.

Table B-10 shows the average coverage rates of the state estimates for all models to be satisfactory.

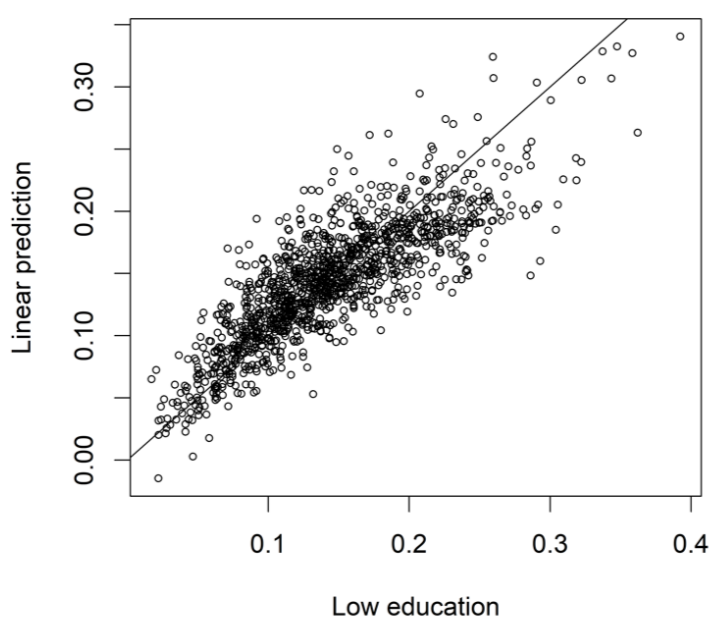
Remarks on Two Aspects of the Simulation Results. Two simple analyses of the ACS data offer some insight into results from the simulation. One concerns the finding that the matched model outperformed the unmatched model in the simulation. A re-examination of the characteristics of the simulation population suggests how population characteristics might influence the relative performance of the small area models. Figures B-2 and B-3 contrast the predictions of a logistic and linear model for the PSU low-education proportions, using only fixed effects, based on the entire ACS sample used in the simulation. The graphs indicate that both model expressions capture a large proportion of the underlying variation, but the logistic fit appears to slant away from the 45-degree line. Very few predicted values from the logistic model are below 0.05, while there is a larger number of actual values below 0.05. This last observation may be a consequence of using population proportions as predictors of the logit of a probability, as is the case in this application with all of the variables representing proportions except the log of income.

Table B-10. Average percentage of coverage of 95 percent credible intervals produced by MCMC based on 500 simulated samples under different HB area-level models for low education, state-level estimates

Model	Overall	Smaller states	Largest 12 states
M1	95.4	95.8	94.2
M2	96.3	96.7	95.0
M3	95.4	95.7	94.3
M4	97.9	98.3	96.6
M4S	97.9	98.4	96.4
M5	98.1	98.4	97.2
M6	97.9	98.2	96.7
M6S	97.8	98.3	96.5
M7	97.9	98.3	96.8
M8	96.1	96.5	94.6
M9	96.0	96.3	94.7
M10	98.2	98.8	96.2
M11	97.4	98.1	94.9
M12	98.1	98.7	96.2
M13	97.2	98.0	94.7
M14	97.9	98.5	95.9
M15	97.0	97.7	94.5

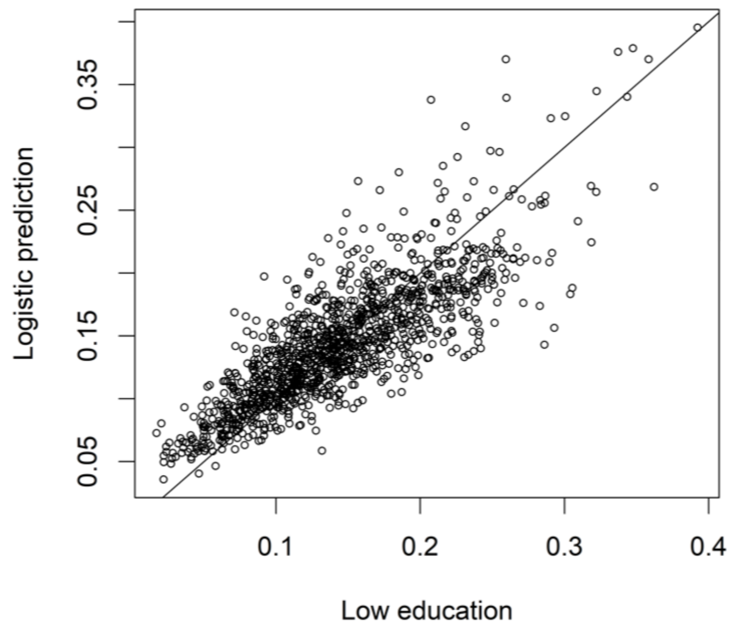
NOTE: HB: Hierarchical Bayes; MCMC: Markov Chain Monte Carlo.
SOURCE: U.S. Department of Commerce, U.S. Census Bureau, American Community Survey, 2011–2015.

Figure B-2. Comparison of low education vs. the predictions under the matched model using only the fixed effects for the 1,234 PSUs in the simulation, based on the entire ACS population used in the study



NOTE: ACS: American Community Survey; PSU: primary sampling unit.
SOURCE: U.S. Department of Commerce, U.S. Census Bureau, American Community Survey, 2011–2015.

Figure B-3. Comparison of low education vs. the predictions under the unmatched model using only the fixed effects for the 1,234 PSUs in the simulation, based on the entire ACS population used in the study

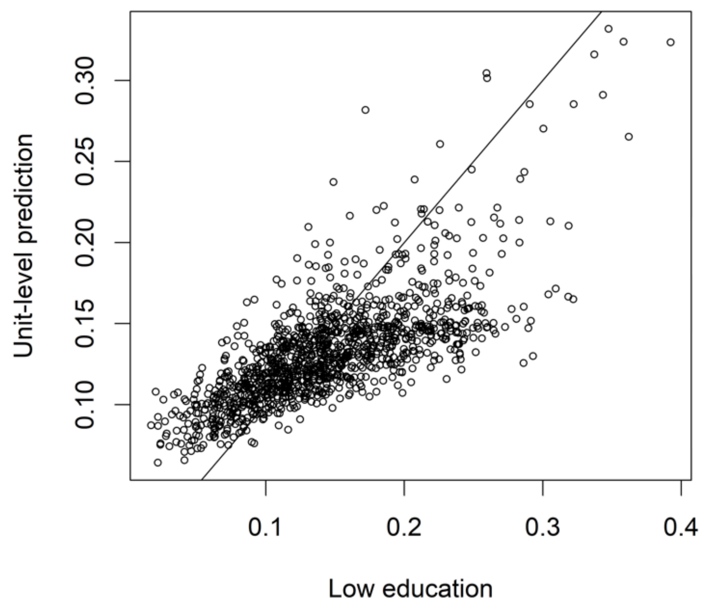


NOTE: ACS: American Community Survey; PSU: primary sampling unit.
SOURCE: U.S. Department of Commerce, U.S. Census Bureau, American Community Survey, 2011–2015.

The average MSE of the predictions in figure B-2 is 10.99 ($\times 10^{-4}$), to place the measure on the same scale as table B-3. The average MSE for the predictions in figure B-3 is 11.53 on the same scale. (Note that these values are not far below the average MSEs realized by the small area models based on a much smaller sample.) Thus, the matched model has an initial advantage over the unmatched model in the simulation, and there is a likely connection between the somewhat better overall fit of the matched model to the full population and its relative performance in the simulation. One might conjecture that for a different population where the unmatched model fits the full population better than the matched model, the simulation results would show the small area estimators based on the unmatched model to do better than the matched model. Repeating the simulation on another population where the relative success of the matched and unmatched models resembles figures B-2 and B-3 is likely to produce a similar outcome to the one here, where the SAE results are better for the matched models. It is an open question of how often the unmatched models would provide better fit to an ACS data or other natural populations. Repeating the simulation on another population would be of future interest.

A second remark concerns the omission from the simulation of any attempt to implement a unit-level small area model. Unit-level small area models are a well-recognized strategy in small area estimation problems, and some researchers express a preference for them when sufficient data are available to support them. But unit-level models may not perform well in the presence of ecological effects unless those effects are carefully included. Figure B-4 shows the consequences of producing estimates as the sum of unit-level predictions, which are the basic building blocks of a unit-level model. The level of predictions given by summing unit-level predictions is distinctly poorer, as reflected in the average scaled MSE of 18.81. Consequently, this simple comparison suggests that a unit-level SAE model would not be competitive in this situation.

Figure B-4. Comparison of low education vs. the predictions of a linear unit-level model using only the fixed effects for the 1,234 PSUs in the simulation, based on the entire ACS population used in the study



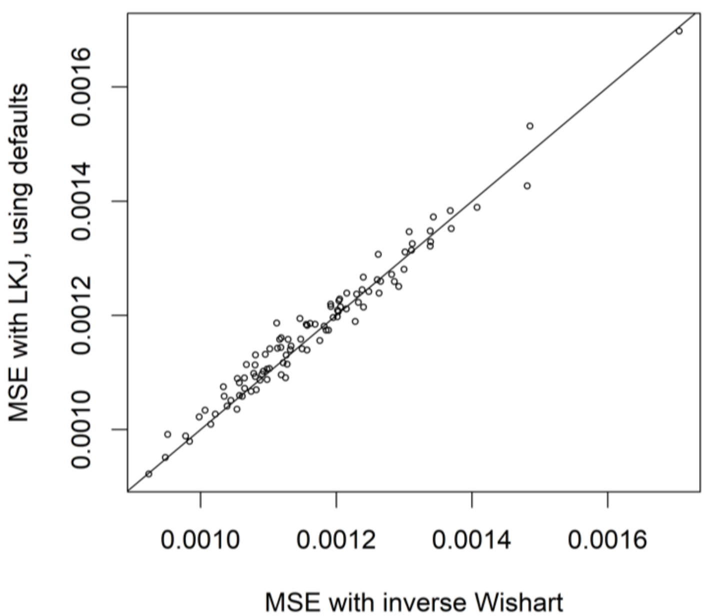
NOTE: ACS: American Community Survey; PSU: primary sampling unit.
SOURCE: U.S. Department of Commerce, U.S. Census Bureau, American Community Survey, 2011–2015.

B.3 Supplemental Simulation

The simulation results reported in the previous section were obtained in the winter and spring of 2018, before the results from the 2017 PIAAC sample were available. The results set a general direction for the modeling of the PIAAC results without dictating every detail. One major change in the final model was to replace the simulation study's use of inverse Wishart distributions as priors for the variance-covariance matrices of the random effects at the county, state, and division levels. Instead, an $LKJ_{corr}(1)$ prior, introduced in section 5.2, was used. This section reports on a small simulation study to investigate whether the revised prior would substantially alter the conclusions of the previous simulation. If the performance changed substantially with the use of the LKJ prior, then the preceding simulation results from 2018 would be of diminished value.

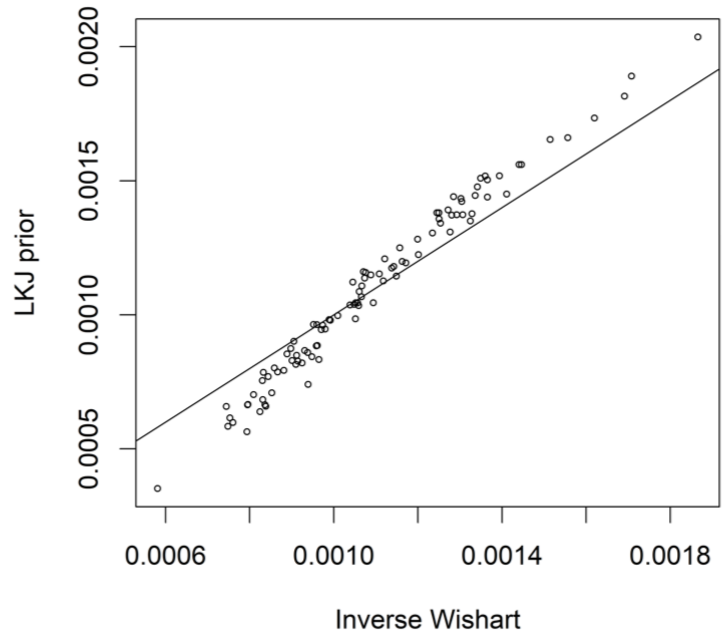
A simulation using the first 100 samples was conducted with the LKJ prior using the default parameters used by STAN in R. The default parameters permitted considerably faster runs than the settings used in the final model. Figure B-5 compares the average MSE using two different priors for M13, the inverse Wishart used in the original simulation and the LKJ prior with the default parameters used by STAN. Figure B-5 indicates that revising the prior has a relatively small effect, generally no more than a few percentage points, on the average MSE, with no clear trend. In contrast, figure B-6 compares the average posterior variance from the simulation, showing generally greater observed differences between the results from the two priors and a tendency for the LKJ prior to produce somewhat more dispersed posterior variances, either systematically higher or systematically lower than the inverse Wishart, depending on the sample.

Figure B-5. Comparison of average MSE in predicting low education, 100 simulations of M13 with inverse Wishart vs. LKJ priors



NOTE: LKJ: Lewandowski, Kurowicka, and Joe. MSE: mean square error.
SOURCE: U.S. Department of Commerce, U.S. Census Bureau, American Community Survey, 2011–2015.

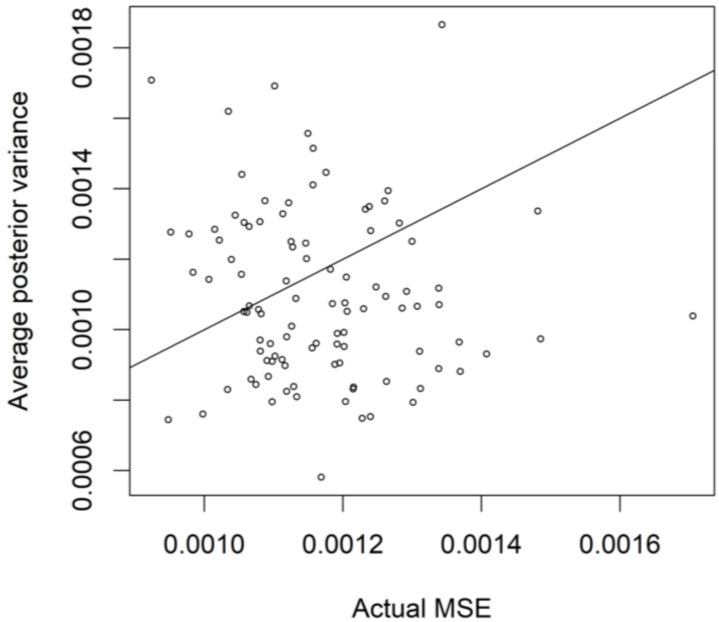
Figure B-6. Comparison of average posterior variance in predicting low education, 100 simulations of M13 with inverse Wishart vs. LKJ priors



NOTE: LKJ: Lewandowski, Kurowicka, and Joe.
SOURCE: U.S. Department of Commerce, U.S. Census Bureau, American Community Survey, 2011–2015.

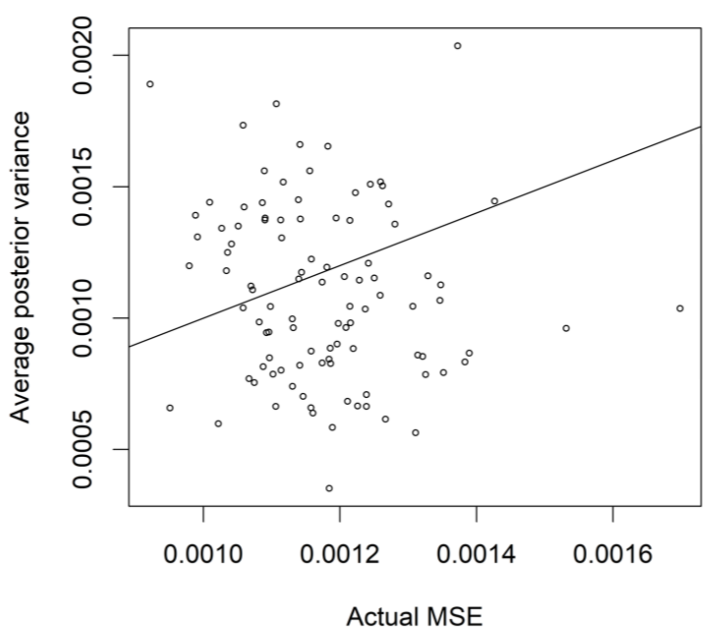
Using the same results as figures B-5 and B-6, figure B-7 compares the average MSE to the average posterior variance using the LKJ prior, and figure B-8 does so for the inverse Wishart. From a finite population perspective, these last two results suggest that the average posterior variance is only a rough approximation to the true MSE.

Figure B-7. Comparison of actual average MSE to the average posterior variance in predicting low education, 100 simulations of M13 with inverse Wishart prior



NOTE: MSE: mean square error.
SOURCE: U.S. Department of Commerce, U.S. Census Bureau, American Community Survey, 2011–2015.

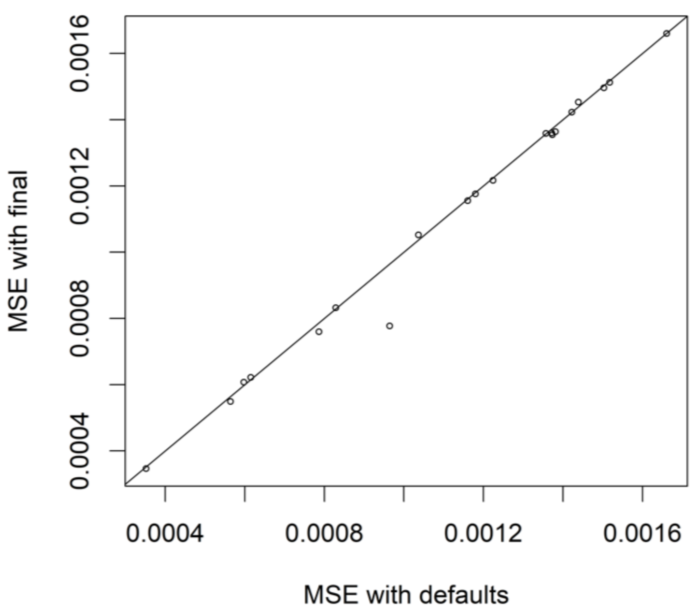
Figure B-8. Comparison of actual average MSE to the average posterior variance in predicting low education, 100 simulations of M13 with LKJ prior



NOTE: LKJ: Lewandowski, Kurowicka, and Joe; MSE: mean square error.
SOURCE: U.S. Department of Commerce, U.S. Census Bureau, American Community Survey, 2011–2015.

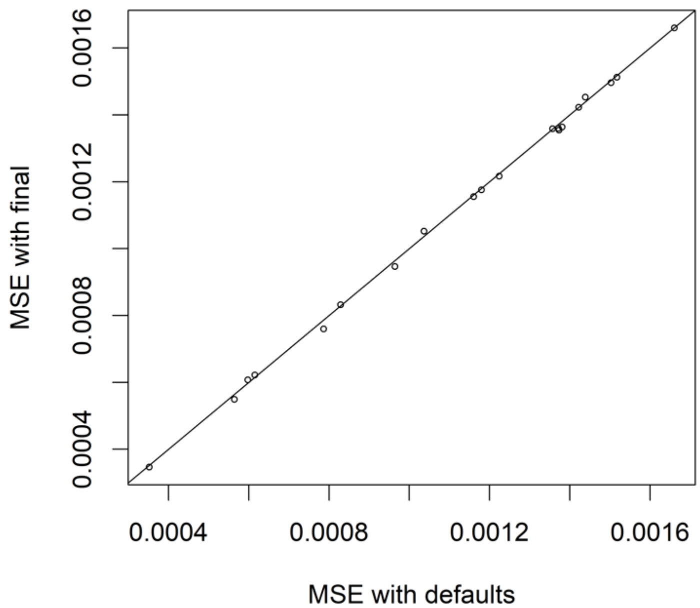
Figures B-5 through B-8 show that the inverse Wishart and LKJ priors yield relative similar results, suggesting that the simulation findings can still be a useful guide for the PIAAC application. A remaining point, however, is that the 100 simulations were based on the default settings rather than the final settings used for PIAAC, adapt_delta=.99 and max_treedepth=13. There was sufficient time to run 20 simulations using the final settings. In one case, replicate sample 17, the running time was exceptionally long. Figure B-9 compares the average MSE of the LKJ prior with the default against the results with the final settings. The result for replicate sample 17 appears in this graph as a distinct outlier. As a check on the computation, all of the random seeds were fixed at the same values, and the same result was exactly reproduced when it was rerun. In an additional run, by changing the value of the random seed just before the call to stan(), the revised finding fell in line with other results as shown in figure B-10. A possible interpretation is that the initial MCMC results for replicate sample 17 were unsatisfactory, illustrating the usefulness of diagnostics to review MCMC results. With the revised results, figure B-10 indicates that findings from the LKJ and inverse Wishart priors are quite similar.

Figure B-9. Comparison of average posterior variance in predicting low education, 20 simulations of M13, LKJ prior with default settings vs. LKJ with the final settings



NOTE: LKJ: Lewandowski, Kurowicka, and Joe; MSE: mean square error.
SOURCE: U.S. Department of Commerce, U.S. Census Bureau, American Community Survey, 2011–2015.

Figure B-10. Comparison of average posterior variance in predicting low education, 20 simulations of M13, LKJ prior with default settings vs. LKJ with the final settings, with a revised random seed for replicate 17



NOTE: LKJ: Lewandowski, Kurowicka, and Joe; MSE: mean square error.
SOURCE: U.S. Department of Commerce, U.S. Census Bureau, American Community Survey, 2011–2015.

B.4 Summary and Discussion

The HB model implemented for NAAL performs well when applied to a similar application based on treating an entire ACS sample as a population. Nonetheless, several variants of the approach are possible, and the simulations show that it is possible to improve on the original approach.

- It was possible to create a survey regression estimator that uniformly improved the performance of the small area estimates over various subsets of the PSUs.
- In the simulation, most models were based on smoothed estimates of the variance. The models generally performed well relative to use of the estimated true variances obtained from simulation. This conclusion must be tempered by the observation that, because the within-sampling variances were affected by the method to form segments rather than arising naturally from information in the ACS data, the apparent conclusion that variance smoothing had a negligible effect might not hold in some other situations.
- The comparisons of models here showed some advantage to the multivariate models in predicting low education, with M9 and M13 producing the generally best results. The multivariate models have other important advantages. First, it is likely that they provide more closely coordinated estimates of high school education concurrently than if high school education were modeled in a univariate model. Second, the multivariate models will support

interval estimates for the sum of low and high school education, equivalent in the PIAAC application to obtaining interval estimates for literacy Level 2 and below as well as for Level 1 and below and for Level 2.

- Incorporating division as a random effect in the model improved the overall results, but the simulation findings are not entirely consistent on this point. The simulation suggested including division as a random effect rather than a fixed effect was more effective.
- The data used for modeling the NAAL, PIAAC, or each simulated sample in the simulation covered a relatively small fraction of the total number of counties. This differs from many of the applications in the small area literature in which most or all of the small areas are represented by at least some sample. The current HB model does not explicitly account for PSU-level variation in estimating the coefficients of the area-level regression model. The generally reasonable average coverage rates of the credible intervals suggest that this limitation did not substantially affect the simulation findings.

B.5 Additional Details on the Design of the Simulation

Major Features of the PIAAC 2012/2014/2017 Sample Design. Although the PIAAC design is described elsewhere in the report and by Hogan et al. (2016), a summary of key features here will facilitate comparison between the PIAAC design and the simulation. The PIAAC sample was drawn through a multistage design, beginning with a first-stage selection of PSUs. Each PSU was a county or a group of counties. PIAAC divided the country into 1,949 PSUs, of which 1,213 were single counties. Most of the largest counties were individual PSUs, including four counties included with certainty.

Sampling for the 2012/2014/2017 PIAAC involved two stratified samples of PSUs. For the 2012/2014, a stratified sample of 80 PSUs was selected, with the 2014 sample using the same PSUs that were selected in 2012. Of the 80 PSUs in the 2012/2014 sample, four were self-representing, that is, selected with certainty. The remaining 1,945 PSUs were grouped into 18 major strata. Each major stratum was divided into an even number of minor strata, and a single PSU was sampled with probability proportional to size from each minor stratum. For 2017, the PIAAC PSUs were restratified into 11 major strata and then into new minor strata, and a first-stage sample of 4 certainty and 76 noncertainty PSUs was selected. Except for the four certainty PSUs, the sampling procedure was designed to minimize the overlap of the 2017 sample with the 2012/2014 sample. As a result, the large PSUs with a probability of selection of 0.5 or more in the original 2012/2014 sample that were not selected then had a conditional probability of selection close to 1.0 into the 2017 sample. Across the combined 2012/2014/2017 samples, 147 unique PSUs were selected for PIAAC.

At the second stage of selection, dwelling units in the sampled PSUs were divided into segments by grouping census blocks. An average of approximately 11 segments were sampled in each PSU. In PSUs composed of more than one county, counties received only a share of the segments allotted to the PSU, occasionally only one or none.

Dwelling units were sampled within the segment as the third stage of selection. Finally, the fourth stage of selection sampled eligible respondents within sampled dwellings. A single individual was sampled if three or fewer persons were eligible, and two individuals were sampled otherwise.

The Simulated Population. The public use file from the 2015 ACS provided the data for the simulation study. The simulation used only variables available from the ACS, so the unweighted ACS data can be regarded as an unknown but true population. Ideally, the simulation would have used ACS data at the county level, but many counties cannot be identified on the public use file. For reasons of protecting the confidentiality of ACS respondents, the Census Bureau restricts the geographic information released on the public use file to a set of PUMAs. There are 2,531 such areas on the 2015 file. PUMAs are allowed to cross county but not state boundaries, and in defining PUMAs, small counties were combined with others in order to satisfy a minimum population requirement of 100,000 people. Generally, however, the largest counties are exactly divided into one or more PUMAs, so that estimates for these counties can be based on reassembling their constituent PUMAs.

As already noted, the PIAAC sample design treated most of the largest U.S. counties as PSUs, including four certainty counties. To mimic this pattern, large counties in the ACS file were reassembled from their constituent PUMAs, including the four PIAAC certainty counties. A total of 431 individual counties were reassembled from 1,728 PUMAs, including one instance that combined Monroe County with Miami-Dade in Florida. (One ACS PUMA includes Monroe with part of Miami-Dade, but Miami-Dade is otherwise the sum of a set of PUMAs. The exception was made in order to treat Miami-Dade/Monroe as a PSU.) Counting Miami-Dade/Monroe as a single county, the 431 counties were treated as PSUs in the simulation. The counties included approximately 144 million persons aged 19–74 in households in 2015.

The remaining 803 PUMAs (=2,531-1,728) represented approximately 73 million persons aged 19–74 in households in 2015. They were treated as individual PSUs in the simulation, bringing the total PSUs in the simulation to 1,234. The PSUs were matched to their major strata in the 2017 PIACC design: The 431 PSUs representing individual counties could be directly assigned their 2017 major strata, and the remaining 803 PSU/PUMAs were assigned the major stratum of the most populous county in the PUMA. Minor strata were not assigned in the simulation.

Simulation of Sampling. The sample of PSUs was meant to approximate but not fully mimic the more complex sampling procedure used for PIAAC. In place of the combination of 4 certainty PSUs and the sampling of large PSUs by a procedure to reduce overlap, the simulation simplified this procedure by treating 10 counties, all large, as self-representing. The number of PSUs selected from each of the 11 major strata was adjusted accordingly and varied between 8 and 15. The PSUs were selected through random systematic sampling using the function UPrandomsystematic() in the R sampling package. The ACS population aged 19–74 was used as a measure of size, but the measure of size was reduced in a few noncertainty PSUs to ensure a probability of selection less than 1. A total of 147 PSUs was sampled in each simulation, the number of unique PSUs in the PIAAC 2012/2014/2017 design.

To simulate the second and third stages of selection for PIAAC, the ACS households in each PSU were grouped into segments of size 7. An interclass correlation of 0.25 corresponding to a design effect of about 2.5 (=1+0.25×(7-1)) was induced by sorting the households by the number of low education adults and assigning a linearly increasing set of initial probabilities, with the first probability approximately one-sixth the last. (The choice of slope for the increase was determined by trial and error.) The design effect was approximately achieved by drawing a household sample of the same size without replacement using these initial probabilities, thereby permuting the sorted households in the PSU. The resulting permutation was partitioned into segments. This assignment of ACS households to segments was performed once and kept fixed throughout the simulation. The unconditional sampling rate for households was set at 0.0105, approximating the overall sampling rate for the combined 2012/2014/2017 PIAAC sample.

A simplified procedure was implemented to approximate the sampling of eligible respondents within households, which is the fourth stage of selection in PIAAC. A single respondent was drawn for households of 3 or fewer, two respondents for 4-6, three respondents for 7-9, etc. Although PIAAC capped the number of respondents at two, the weights for large households were trimmed in PIAAC, a process not replicated in the simulation. The sampling of eligible respondents within sampled households is a primary source of weight variation in the PIAAC design.

The simulated samples were weighted by assigning each sampled individual the reciprocal of the product of the following factors:

- $\pi_{jk}^{(1)}$ = the probability of selection of PSU k in state j. This probability was 1 for the 10 certainty PSUs; otherwise when the PSU is in stratum S_h

$$\pi_{jk}^{(1)} = \frac{n_h m_{jk}}{\sum_{j'k' \in S_h} m_{j'k'}},$$

where $m_{j'k'}$ is the measure of size for each PSU $j'k'$ in the same major stratum-typically equal to the ACS estimated population aged 19–74 but in a few cases a value to prevent $\pi_{jk}^{(1)}$ from exceeding 1-and n_h is the number of PSUs sampled from the major stratum h;

- $\pi_{jk}^{(2)}$ = the expected number of hits for an individual segment resulting from SRSWR within the PSU, computed as the ratio of the number of segments assigned to be sampled with replacement from the PSU to the number of segments in the PSU; and

- $\pi_{jk\ell}^{(3)} = n_{jk\ell s}/n_{jk\ell c}$, where $n_{jk\ell c}$ is the number of eligible persons in the household ℓ and $n_{jk\ell s}$ is the number sampled, that is, $n_{jk\ell s} = 1$ if $n_{jk\ell c} \leq 3$; $n_{jk\ell s} = 2$ if $4 \leq n_{jk\ell c} \leq 6$; etc.

The number of segments to sample in each PSU was determined to keep variation in $\pi_{jk}^{(1)}\pi_{jk}^{(2)}$ to a minimum, that is, to produce an approximately self-weighting sample of households. When applied to the simulation samples, the weights for sampled person l in household ℓ in state j and PSU k,

$$w_{jkl} = 1/(\pi_{jk}^{(1)}\pi_{jk}^{(2)}\pi_{jk\ell}^{(3)}),$$

can be used to form unconditionally design unbiased estimates of unweighted ACS PSU-level totals from the ACS public use file.

Analysis of state and national rates. Although unweighted proportions from the full ACS were the reference population in the simulation, a small complication was introduced in analyzing rates for states and the U.S, although not for the rates for PSUs. ACS-weighted estimates of the eligible population by PSU were used to aggregate the PSU-level rates to state and national rates. This approach was used consistently throughout the analysis. (In hindsight, a simpler alternative would have been to form aggregates using the unweighted PSU counts.) With this approach, 14.22 percent were estimated to fall in the low education group instead of the 14.29 percent in the public use file. Similarly, 29.13 percent were in the high school education group instead of 29.86 percent on the public use file. The same weighting was used to aggregate HB PSU-level estimates to the state and national levels. Thus, the effect on the analysis can be assumed to be small.

Appendix C

Select Study Results

APPENDIX C

SELECT STUDY RESULTS

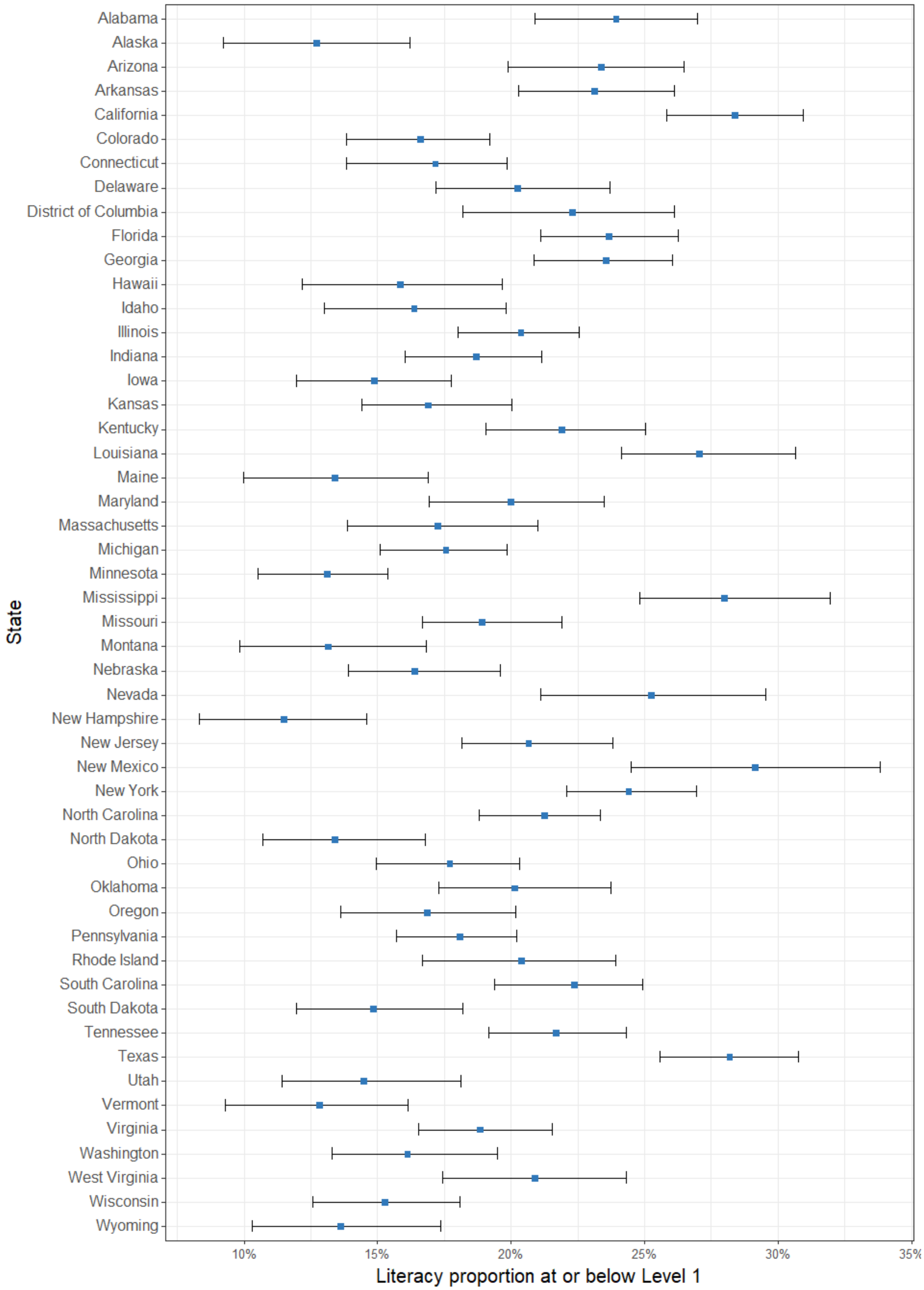
This appendix provides the following:

- state-level results for all outcomes;
- additional tables and graphs related to chapter 5; and
- additional tables and graphs related to chapter 6.

State-level results for all outcomes

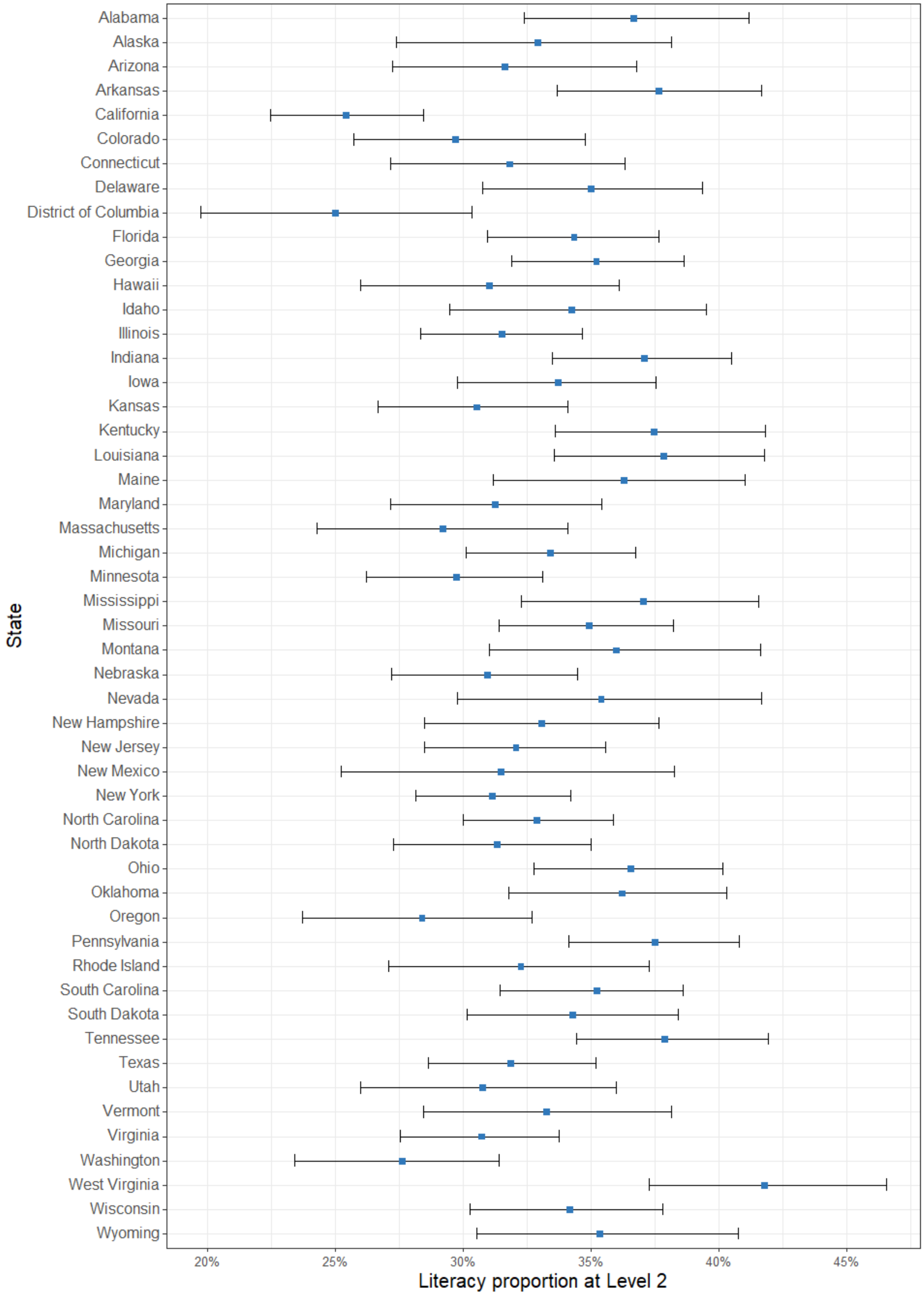
For each outcome, figure C-1 provides a graph with estimates and corresponding 95 percent credible intervals on the *x* axis and state indicators on the *y* axis. As part of our dissemination strategy, a web page is available to allow easy access to the final predictions, while allowing the user to compare the predicted quantities for a given pair of domains of interest.

Figure C-1. Small area estimates and credible intervals for states, by outcome: 2012/2014/2017



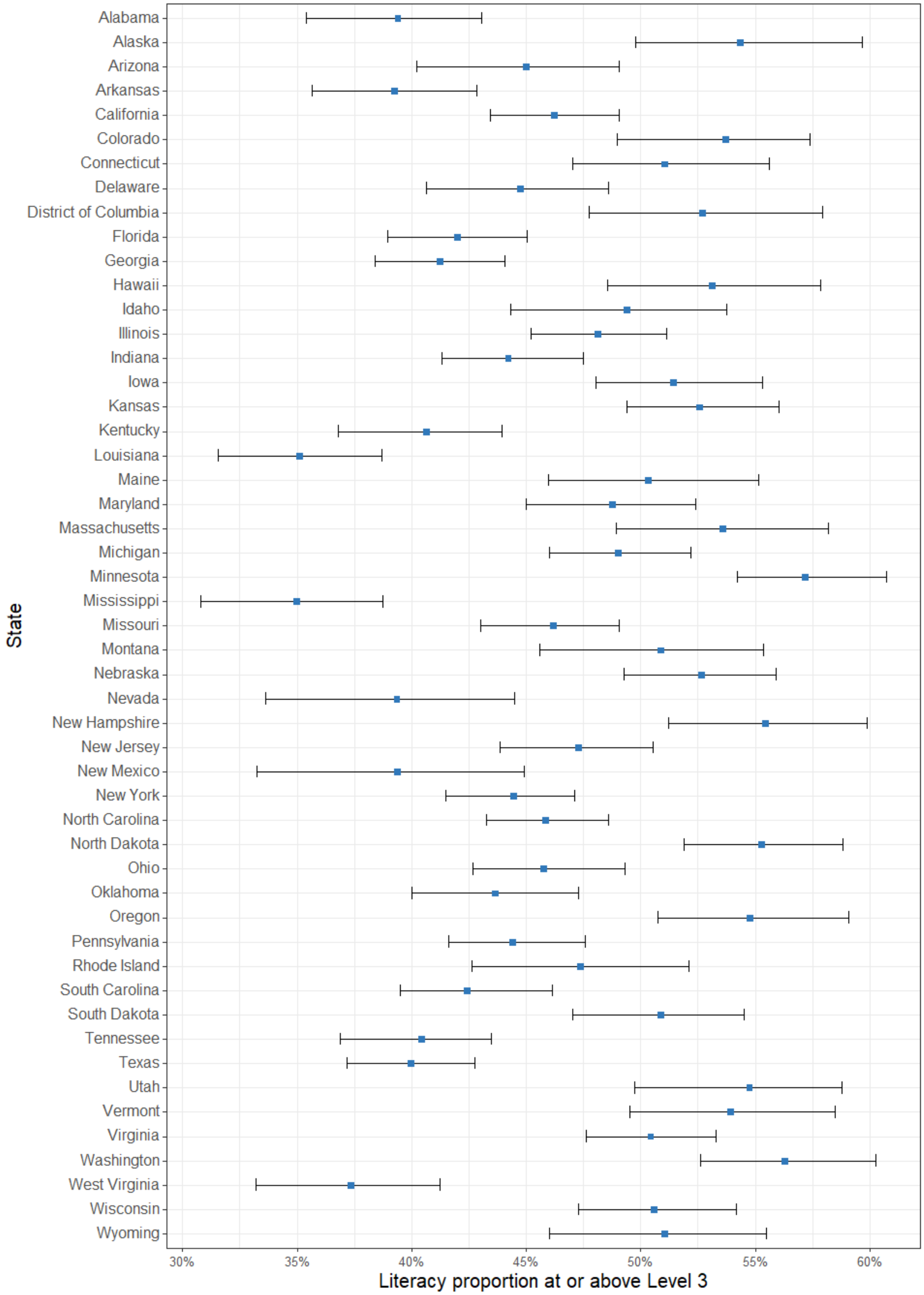
See notes at end of figure.

Figure C-1. Small area estimates and credible intervals for states, by outcome: 2012/2014/2017—Continued



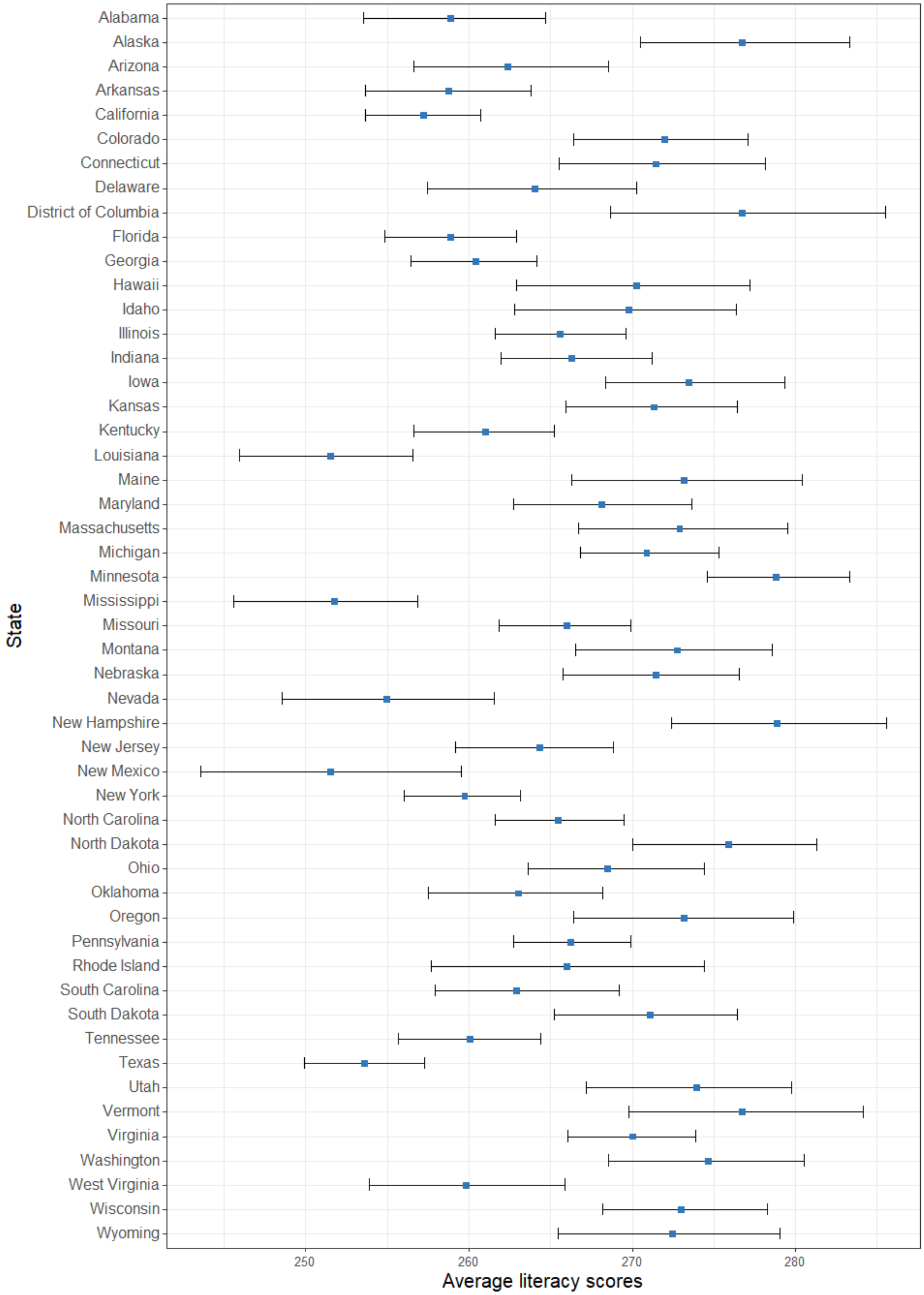
See notes at end of figure.

Figure C-1. Small area estimates and credible intervals for states, by outcome: 2012/2014/2017—Continued



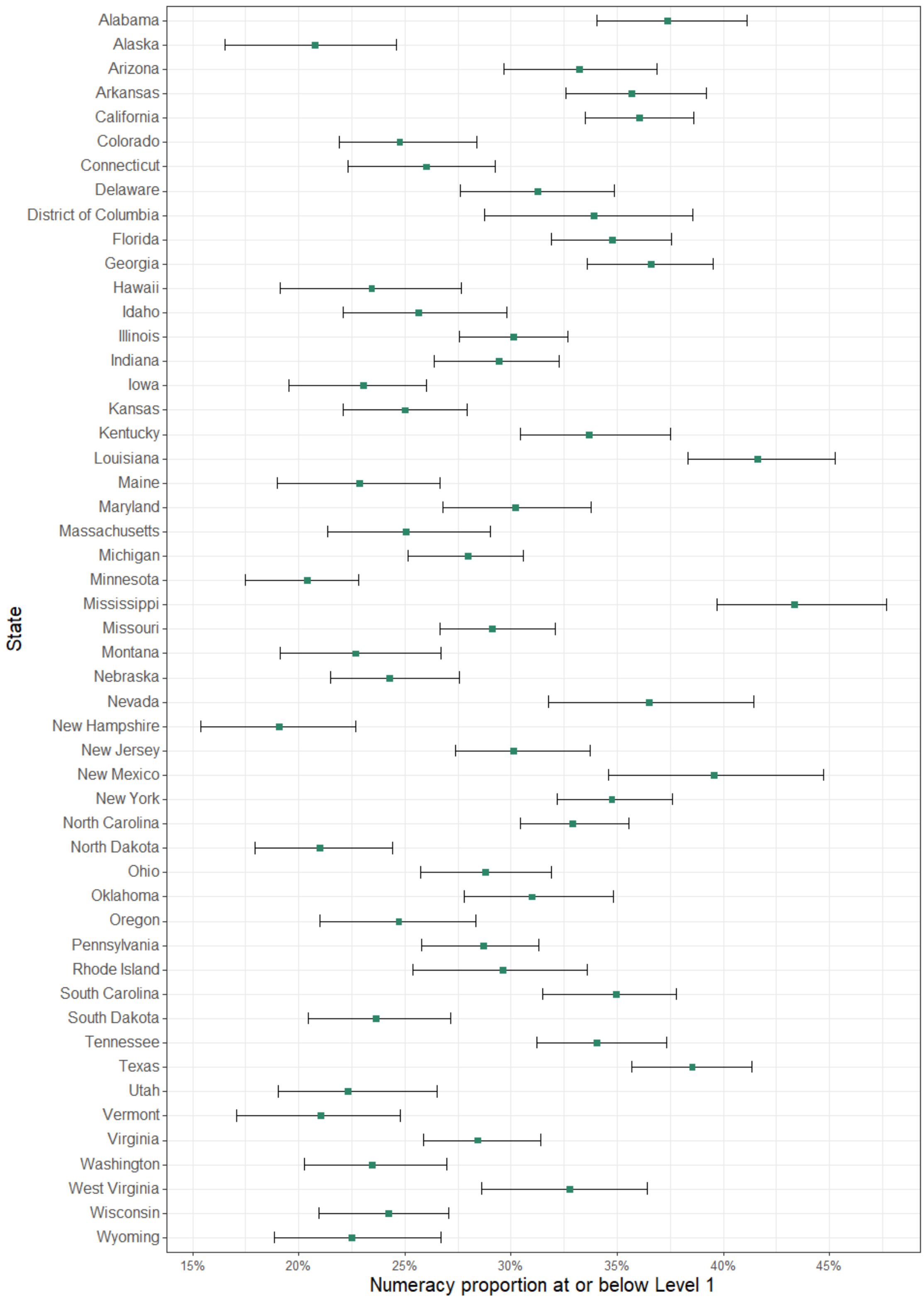
See notes at end of figure.

Figure C-1. Small area estimates and credible intervals for states, by outcome: 2012/2014/2017—Continued



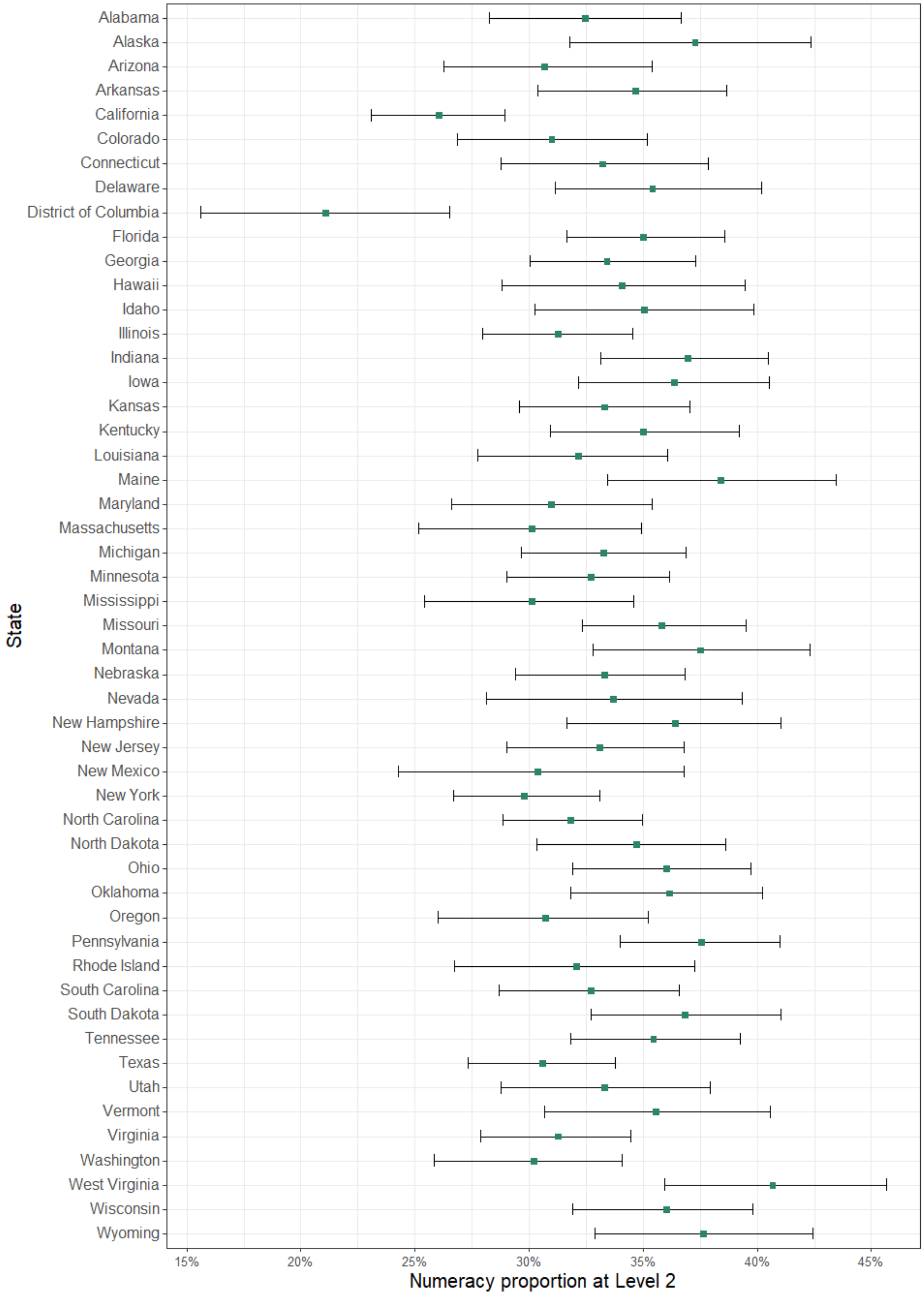
See notes at end of figure.

Figure C-1. Small area estimates and credible intervals for states, by outcome: 2012/2014/2017—Continued



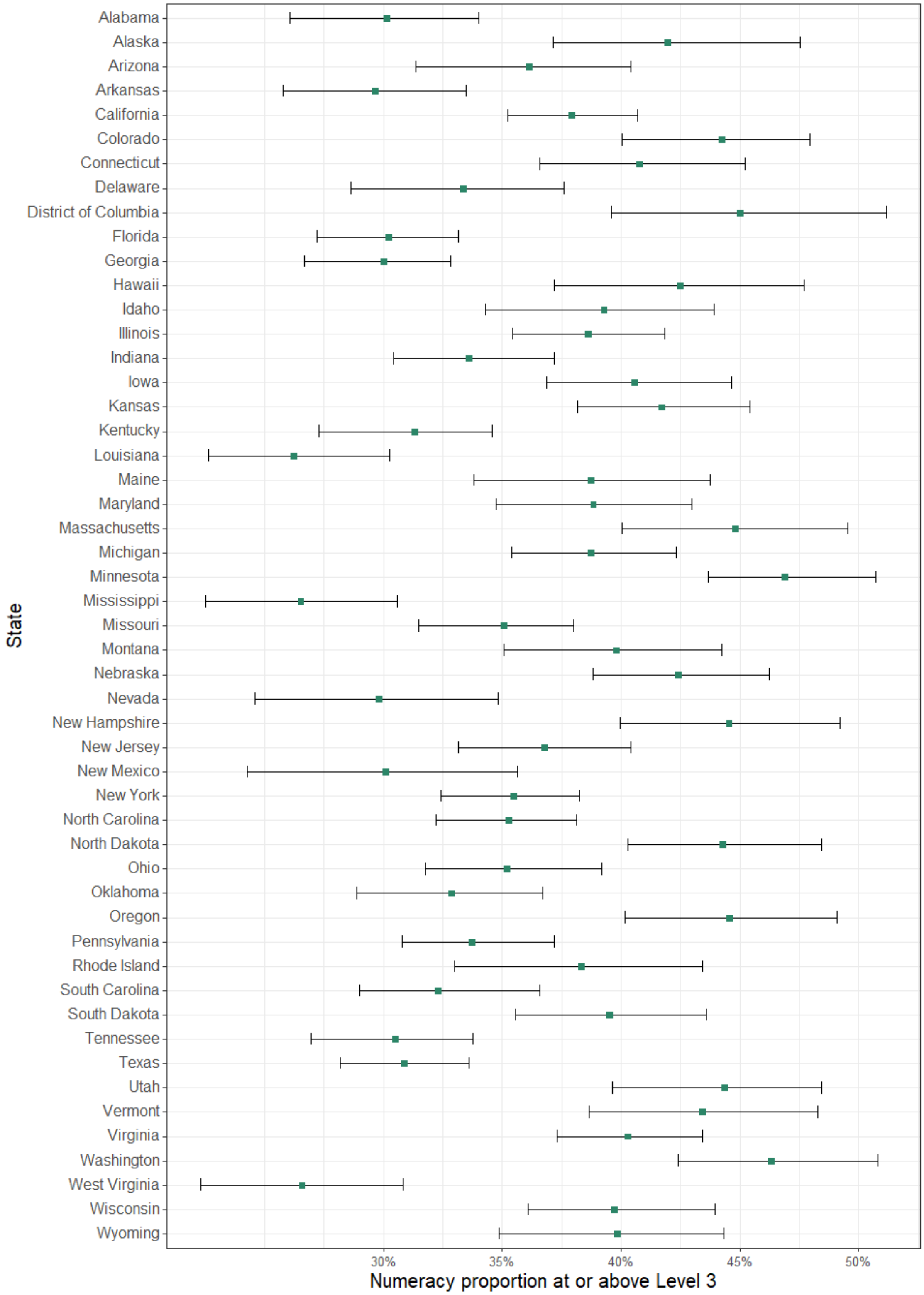
See notes at end of figure.

Figure C-1. Small area estimates and credible intervals for states, by outcome: 2012/2014/2017—Continued



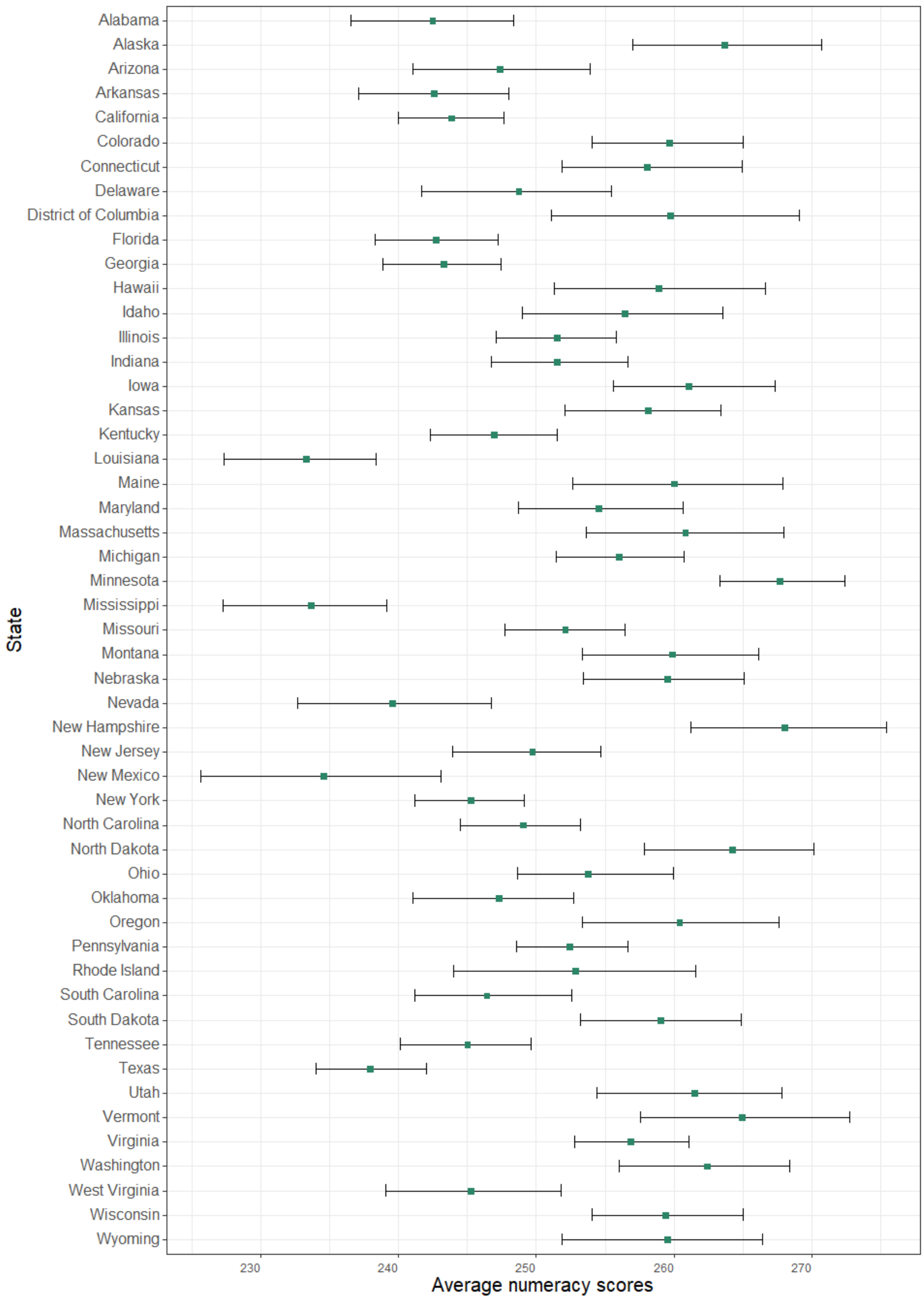
See notes at end of figure.

Figure C-1. Small area estimates and credible intervals for states, by outcome: 2012/2014/2017—Continued



See notes at end of figure.

Figure C-1. Small area estimates and credible intervals for states, by outcome: 2012/2014/2017—Continued



SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

Additional Chapter 5 Tables and Graphs

The appendix provides the results for all other outcomes not covered in chapter 5. Tables C-1 through C-3 provide the different sets of covariates for literacy average, numeracy proportions, and numeracy average, respectively. Table C-4 provides distribution of credible interval widths and coefficients of variation for small area estimates, for each outcome.

Table C-1. Different sets of covariates used in cross validation for literacy average: 2012/2014/2017

Variable	Scenarios 1	2	3	4
Education—LH	X	X	X	X
Education—MH	X	X	X	X
Poverty	X	X	X	X
Black	X		X	X
Enter U.S. 2010	X			
Health insurance	X	X	X	X
Unemployment rate	X			
Grant/Scholarship received	X			
Journey to work	X			
Hispanic			X	X
Service occupations	X	X		X
Sum of squared differences between predicted averages and direct estimates over 44 counties with sample size at least 100	2,573.61	2,931.72	2,778.59	2,527.72

NOTE: LH: Percentage of population aged 25 and over with less than high school education; MH: Percentage of population aged 25 and over with more than high school education.
SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

Table C-2. Different sets of covariates used in cross validation for numeracy proportions: 2012/2014/2017

Variable	Scenarios 1	2	3
Education—LH	X	X	X
Education—MH	X	X	X
Poverty	X	X	X
Black	X	X	X
Health insurance	X	X	X
Diabetes rate	X		
Hispanic		X	X
Service occupations			X
Sum of squared differences between predicted proportions and direct estimates over 44 counties with sample size at least 100			
P1	0.140	0.116	0.113
P2	0.132	0.132	0.133
P3	0.217	0.196	0.190

NOTE: LH: Percentage of population aged 25 and over with less than high school education; MH: Percentage of population aged 25 and over with more than high school education; P1: Proportion at or below Level 1; P2: Proportion at Level 2; P3: Proportion at or above Level 3.
SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

Table C-3. Different sets of covariates used in cross validation for numeracy average: 2012/2014/2017

Variable	Scenarios 1	2	3	4	5
Education—LH	X	X	X	X	X
Education—MH	X	X	X	X	X
Poverty	X	X	X	X	X
Black	X	X	X	X	X
Unemployment rate	X	X			X
Health insurance	X	X	X	X	X
Grant/Scholarship received	X				
Hispanic	X		X	X	X
Service occupations	X	X		X	
Sum of squared differences between predicted averages and direct estimates over 44 counties with sample size at least 100	2,753.02	3,266.64	3,426.92	3,193.44	3,313.61

NOTE: LH: Percentage of population aged 25 and over with less than high school education; MH: Percentage of population aged 25 and over with more than high school education.
SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

Table C-4. Distribution of credible interval widths and coefficients of variation for small area estimates: 2012/2014/2017

Proficiency domain	Statistic	Percentile 20	40	60	80	Median
Literacy proportion—at Level 2	County estimates					
	95 percent credible interval width (percent)	9.7	10.4	11.3	12.7	10.8
	Coefficient of variation (percent)	6.4	6.9	7.5	8.7	7.2
	Sampled county estimates					
	95 percent credible interval width (percent)	8.8	9.4	10.0	11.0	9.8
	Coefficient of variation (percent)	6.5	7.0	7.8	8.7	7.3
	Nonsampled county estimates					
	95 percent credible interval width (percent)	9.7	10.5	11.4	12.7	10.9
	Coefficient of variation (percent)	6.4	6.9	7.5	8.7	7.2
	State estimates					
	95 percent credible interval width (percent)	6.8	7.7	8.8	9.9	8.2
	Coefficient of variation (percent)	5.1	5.8	6.3	7.8	6.0
Literacy proportion—at or above Level 3	County estimates					
	95 percent credible interval width (percent)	9.4	9.9	10.7	11.7	10.3
	Coefficient of variation (percent)	4.9	5.8	7.0	9.3	6.3
	Sampled county estimates					
	95 percent credible interval width (percent)	8.6	9.1	9.5	10.3	9.3
	Coefficient of variation (percent)	4.1	4.6	5.4	7.4	5.0
	Nonsampled county estimates					
	95 percent credible interval width (percent)	9.4	10.0	10.7	11.8	10.3
	Coefficient of variation (percent)	5.0	5.9	7.1	9.4	6.4
	State estimates					
	95 percent credible interval width (percent)	6.2	6.9	7.9	9.2	7.3
	Coefficient of variation (percent)	3.3	3.6	4.2	4.8	3.9
Literacy average	County estimates					
	95 percent credible interval width	18.1	18.8	19.4	20.3	19.1
	Coefficient of variation (percent)	1.7	1.8	1.9	2.0	1.8
	Sampled county estimates					
	95 percent credible interval width	14.1	15.0	15.8	16.7	15.3
	Coefficient of variation (percent)	1.3	1.4	1.5	1.6	1.5
	Nonsampled county estimates					
	95 percent credible interval width	18.2	18.9	19.5	20.4	19.2
	Coefficient of variation (percent)	1.7	1.8	1.9	2.0	1.8
	State estimates					
	95 percent credible interval width	8.5	10.6	11.3	13.0	11.0
	Coefficient of variation (percent)	0.8	1.0	1.1	1.2	1.0

See notes at end of table.

Table C-4. Distribution of credible interval widths and coefficients of variation for small area estimates: 2012/2014/2017—Continued

Proficiency domain	Statistic	Percentile 20	Percentile 40	Percentile 60	Percentile 80	Median
Numeracy proportion—at or below Level 1	County estimates					
	95 percent credible interval width (percent)	8.6	9.2	9.9	11.0	9.5
	Coefficient of variation (percent)	6.2	7.2	8.3	10.0	7.7
	Sampled county estimates					
	95 percent credible interval width (percent)	7.9	8.3	8.8	9.5	8.5
	Coefficient of variation (percent)	5.8	6.8	7.8	9.4	7.3
	Nonsampled county estimates					
	95 percent credible interval width (percent)	8.7	9.2	9.9	11.1	9.6
	Coefficient of variation (percent)	6.2	7.2	8.3	10.0	7.7
	State estimates					
	95 percent credible interval width (percent)	5.7	6.5	7.1	7.7	6.7
	Coefficient of variation (percent)	4.6	5.4	6.5	7.5	5.9
Numeracy proportion—at Level 2	County estimates					
	95 percent credible interval width (percent)	10.7	11.3	12.1	13.3	11.7
	Coefficient of variation (percent)	7.0	7.5	8.3	9.7	7.8
	Sampled county estimates					
	95 percent credible interval width (percent)	9.7	10.3	10.8	11.5	10.5
	Coefficient of variation (percent)	6.9	7.6	8.3	9.7	7.9
	Nonsampled county estimates					
	95 percent credible interval width (percent)	10.8	11.4	12.2	13.3	11.8
	Coefficient of variation (percent)	7.0	7.5	8.3	9.7	7.8
	State estimates					
	95 percent credible interval width (percent)	7.2	8.2	8.8	9.6	8.3
	Coefficient of variation (percent)	5.4	5.9	6.5	7.2	6.0
Numeracy proportion—at or above Level 3	County estimates					
	95 percent credible interval width (percent)	11.4	12.0	12.5	13.3	12.2
	Coefficient of variation (percent)	7.6	9.2	11.3	15.3	10.1
	Sampled county estimates					
	95 percent credible interval width (percent)	10.1	10.6	11.1	11.6	10.9
	Coefficient of variation (percent)	6.0	7.0	8.4	11.4	7.5
	Nonsampled county estimates					
	95 percent credible interval width (percent)	11.5	12.0	12.5	13.3	12.3
	Coefficient of variation (percent)	7.8	9.3	11.4	15.5	10.3
	State estimates					
	95 percent credible interval width (percent)	6.8	7.6	8.2	9.4	7.9
	Coefficient of variation (percent)	4.5	5.0	5.6	6.3	5.2

See notes at end of table.

Table C-4. Distribution of credible interval widths and coefficients of variation for small area estimates: 2012/2014/2017—Continued

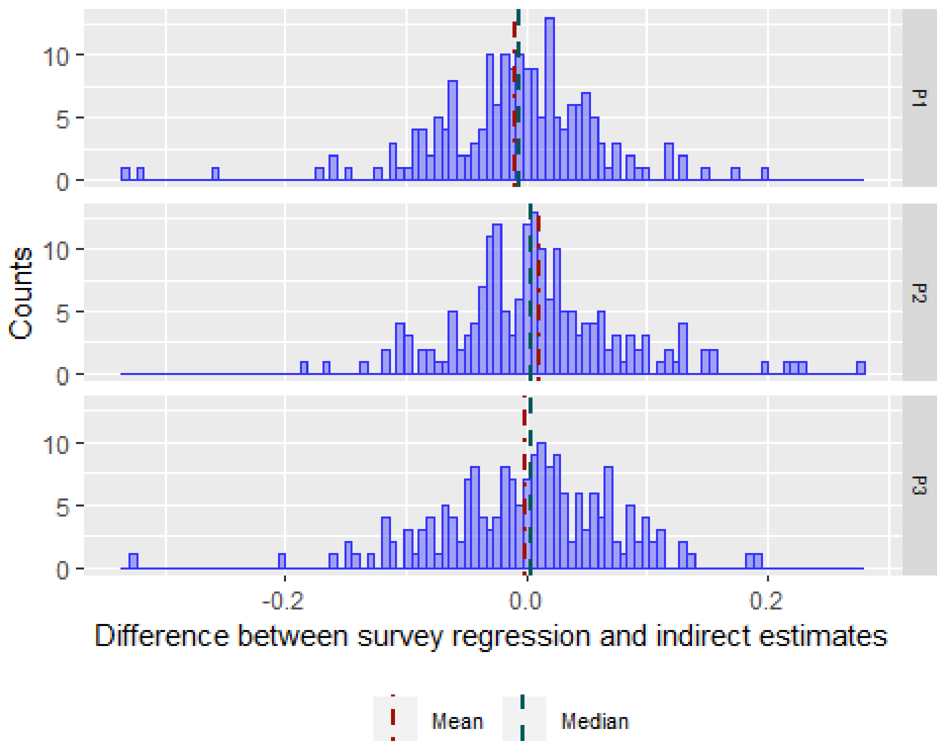
Proficiency domain	Statistic	Percentile 20	40	60	80	Median
Numeracy average	County estimates					
	95 percent credible interval width	19.5	20.2	20.8	21.8	20.5
	Coefficient of variation (percent)	1.9	2.0	2.1	2.3	2.1
	Sampled county estimates					
	95 percent credible interval width	15.3	16.4	17.1	18.2	16.7
	Coefficient of variation (percent)	1.5	1.6	1.7	1.9	1.7
	Nonsampled county estimates					
	95 percent credible interval width	19.6	20.3	20.9	21.9	20.6
	Coefficient of variation (percent)	2.0	2.1	2.2	2.3	2.1
	State estimates					
	95 percent credible interval width	9.1	11.3	12.3	14.2	11.7
	Coefficient of variation (percent)	0.9	1.1	1.2	1.4	1.2

SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

Additional Chapter 6 Tables and Graphs

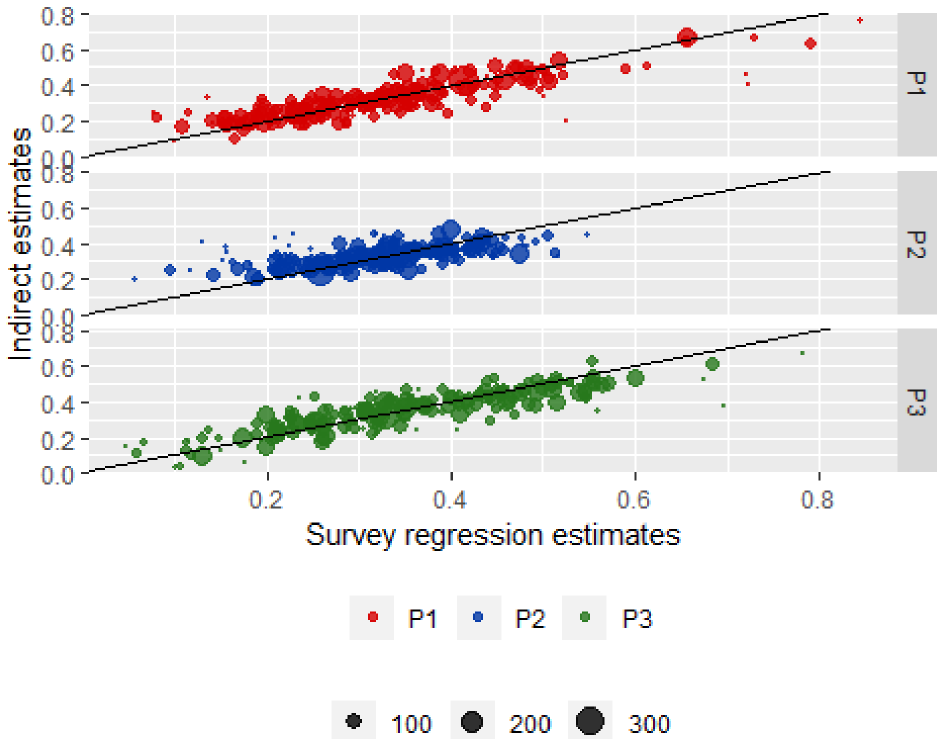
The results for all other outcomes not covered in chapter 6 are provided below for the external validation approaches. Figures C-2 through C-16 provide the evaluation graphs for literacy average, numeracy proportions, and numeracy average, respectively. Tables C-5 through C-11 provide the evaluation results for aggregates to various domains.

Figure C-2. Numeracy proportion—Histogram of differences between survey regression estimates and indirect estimates: 2012/2014/2017



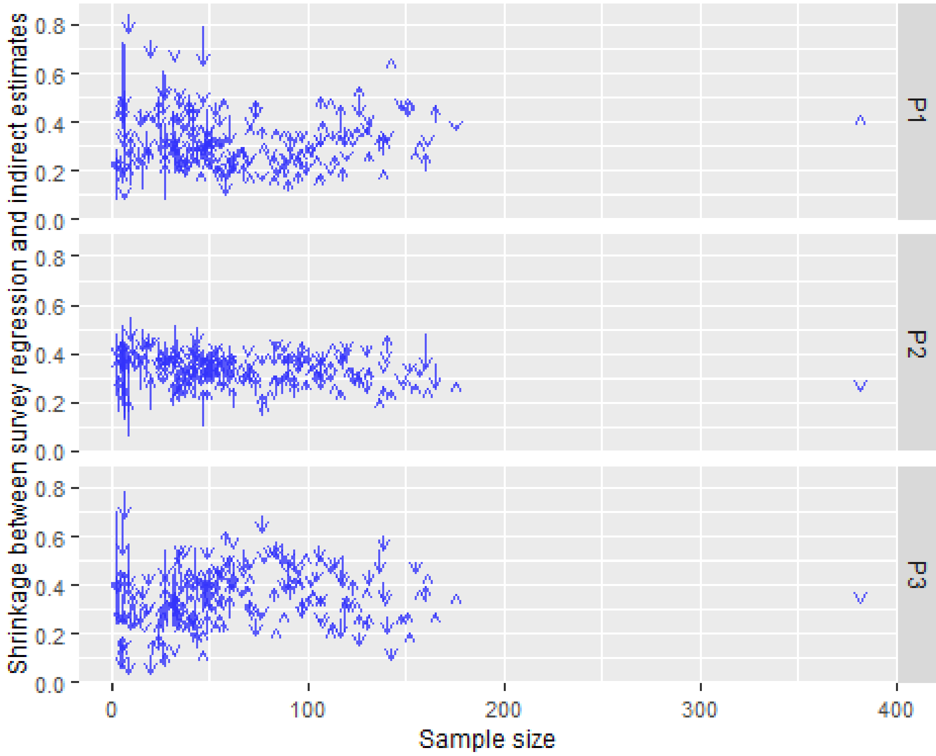
NOTE: P1: proportion at or below Level 1; P2: proportion at Level 2; P3: proportion at or above Level 3.
SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

Figure C-3. Numeracy proportion—Comparison between survey regression estimates and indirect estimates: 2012/2014/2017



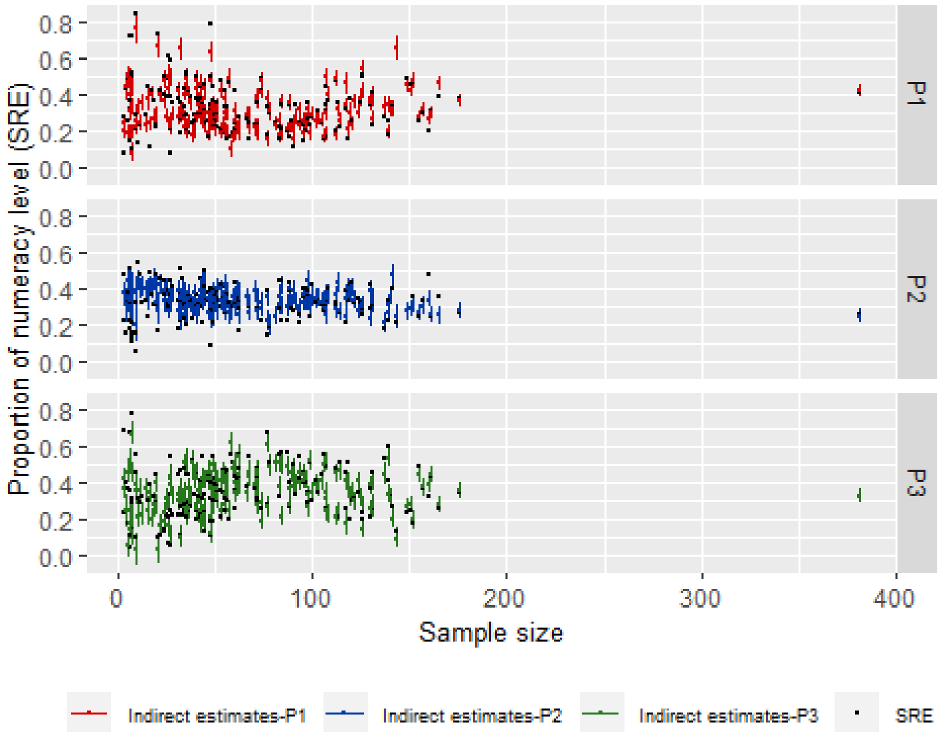
NOTE: P1: proportion at or below Level 1; P2: proportion at Level 2; P3: proportion at or above Level 3.
SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

Figure C-4. Numeracy proportion—Shrinkage plots of point estimates, by sample size: 2012/2014/2017



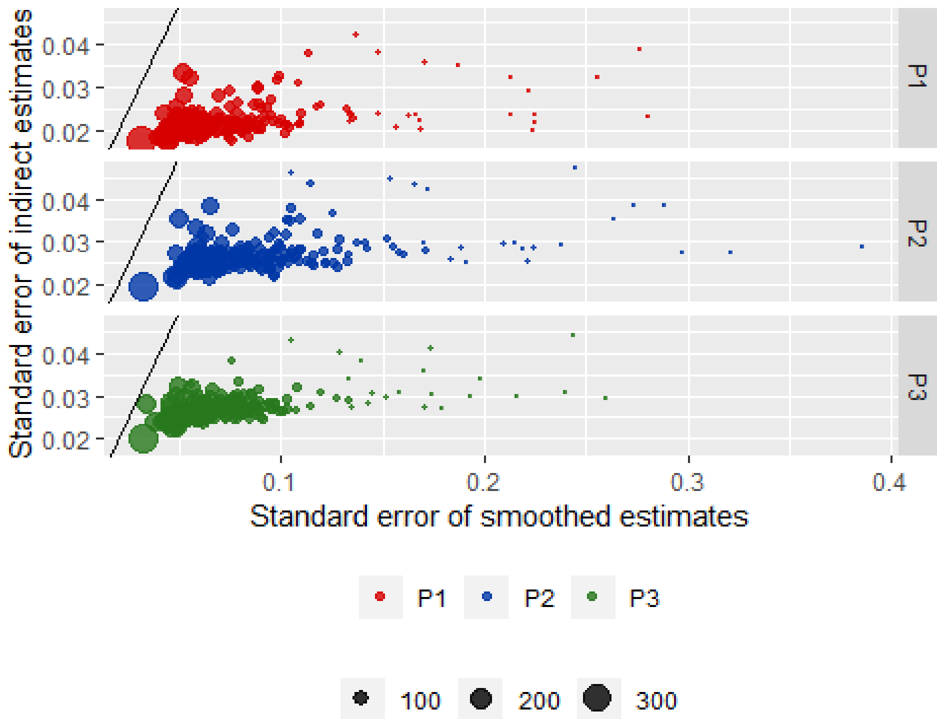
NOTE: P1: proportion at or below Level 1; P2: proportion at Level 2; P3: proportion at or above Level 3.
SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

Figure C-5. Numeracy proportion—Indication of coverage by credible interval: 2012/2014/2017



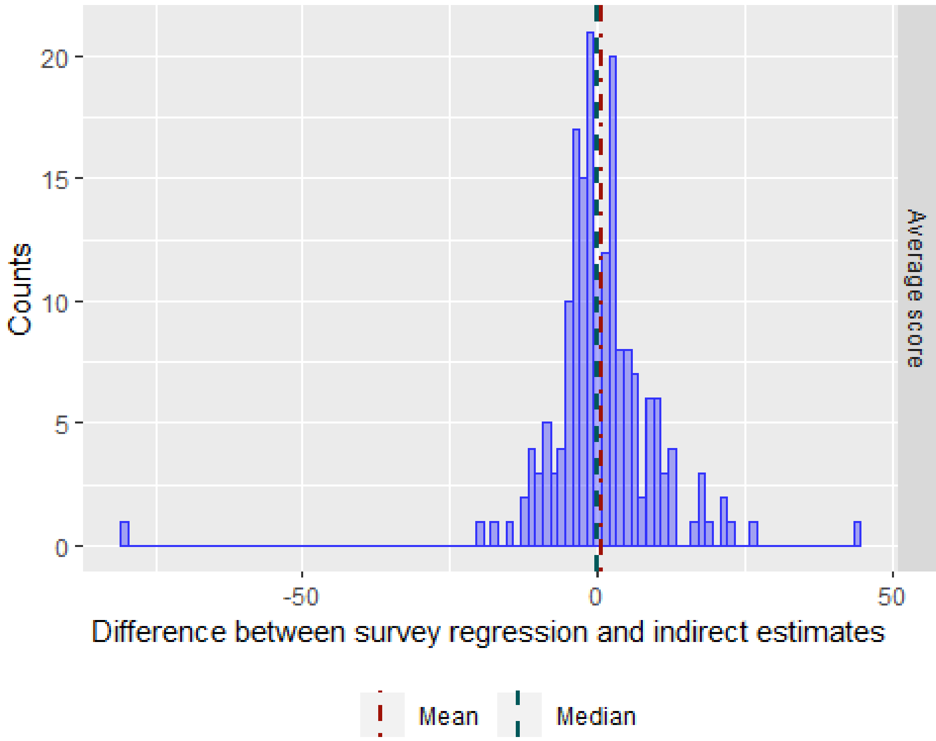
NOTE: P1: proportion at or below Level 1; P2: proportion at Level 2; P3: proportion at or above Level 3; SRE: Survey regression estimates.
SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

Figure C-6. Numeracy proportion—Comparison of standard errors between model and smoothed approaches: 2012/2014/2017



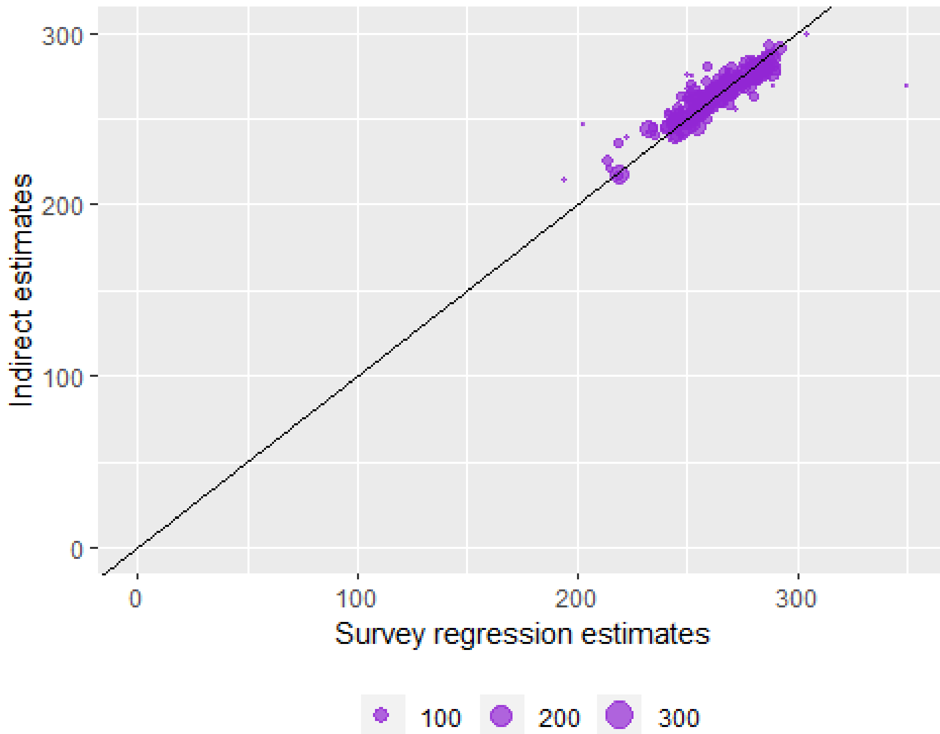
NOTE: P1: proportion at or below Level 1; P2: proportion at Level 2; P3: proportion at or above Level 3.
SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

Figure C-7. Literacy average—Histogram of differences between SRE and indirect estimates: 2012/2014/2017



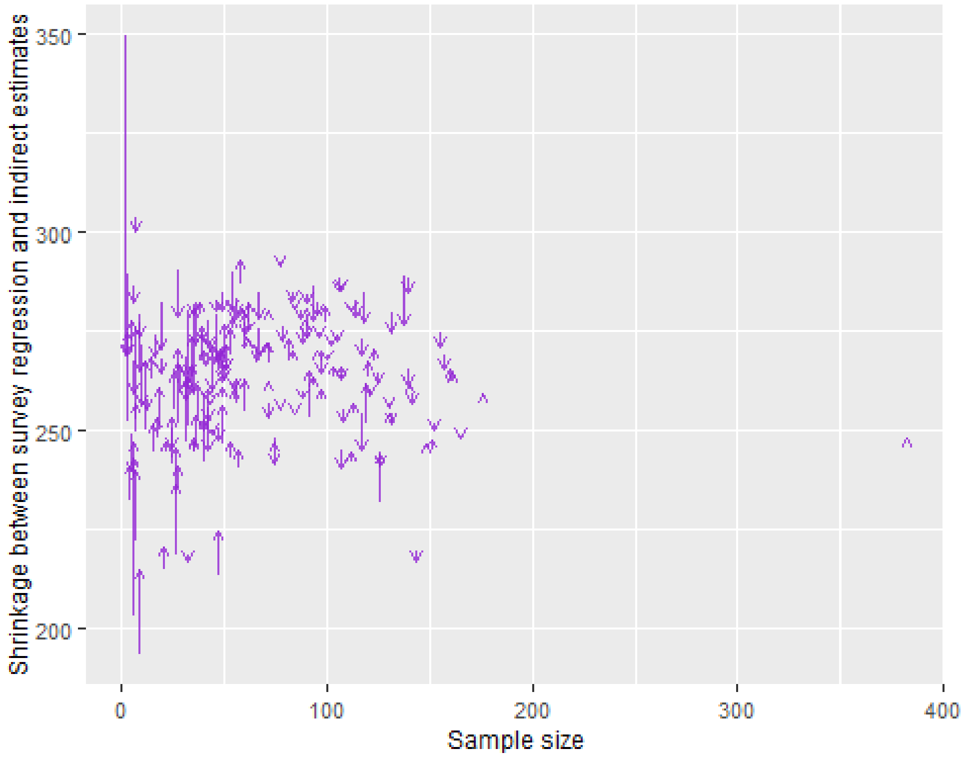
NOTE: SRE: Survey regression estimates.
SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

Figure C-8. Literacy average—Scatterplot of SRE and indirect estimates, with sample size as bubbles: 2012/2014/2017



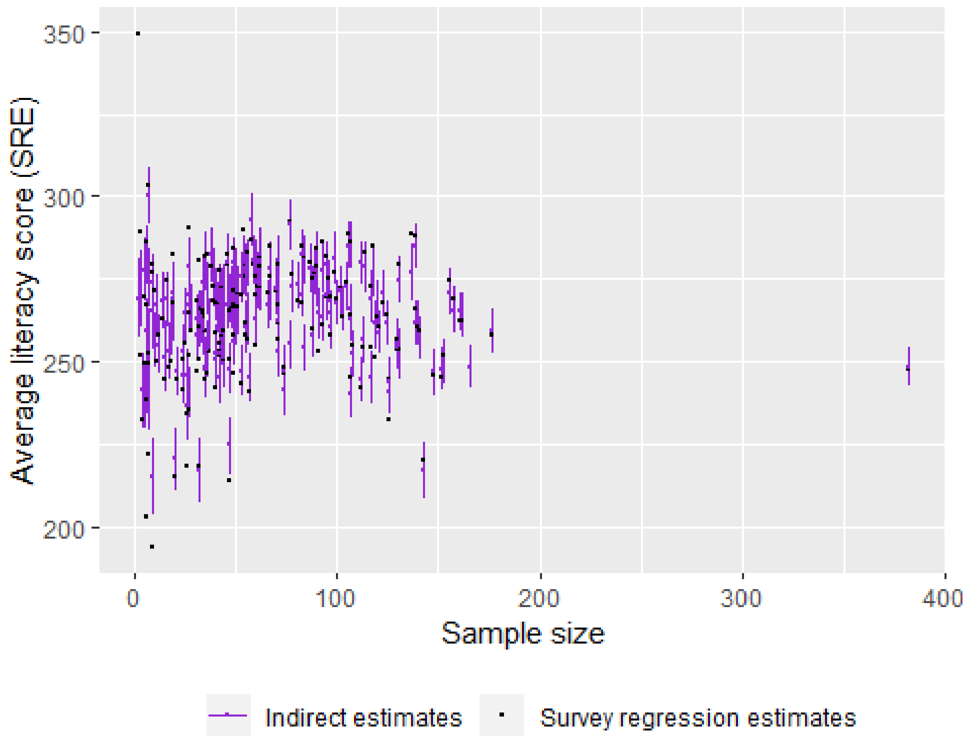
NOTE: SRE: Survey regression estimates.
SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

Figure C-9. Literacy average—Shrinkage plots of point estimates, by sample size: 2012/2014/2017



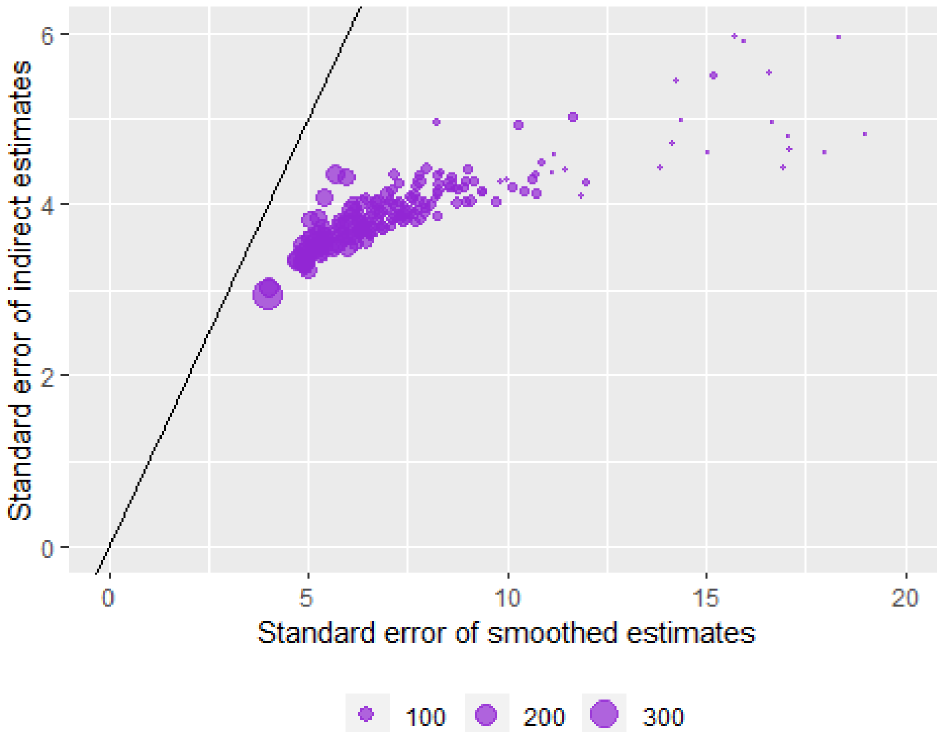
SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

Figure C-10. Literacy average—Indication of coverage by credible interval: 2012/2014/2017



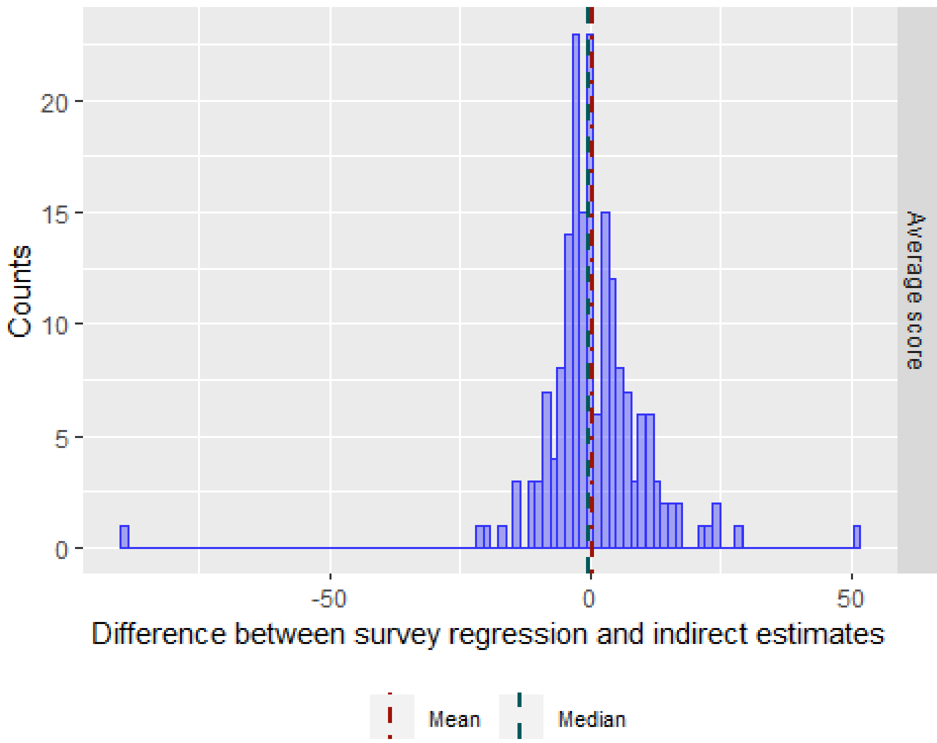
NOTE: SRE: Survey regression estimates.
SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

Figure C-11. Literacy average—Comparison of standard errors between model and smoothed approaches: 2012/2014/2017



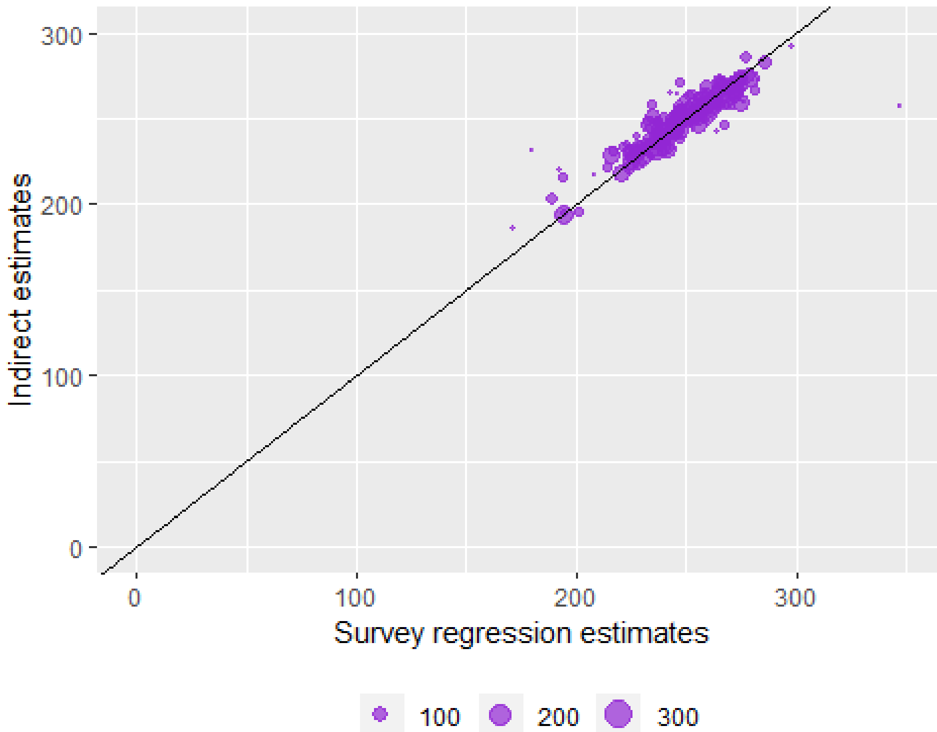
SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

Figure C-12. Numeracy average—Histogram of differences between SRE and indirect estimates: 2012/2014/2017



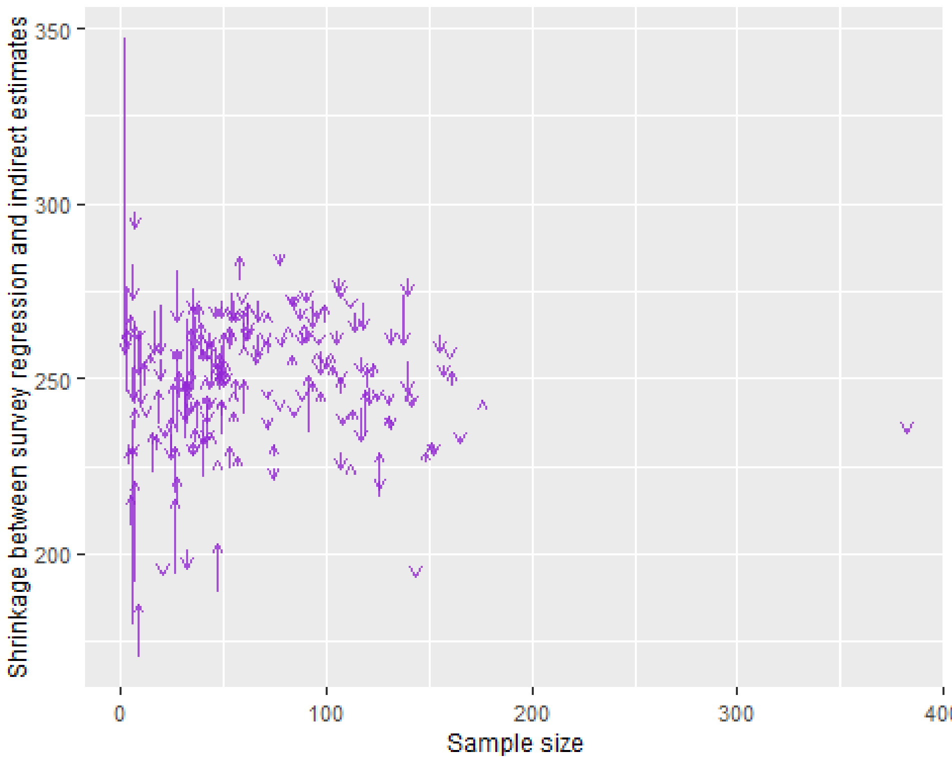
NOTE: SRE: Survey regression estimates.
SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

Figure C-13. Numeracy average—Scatterplot of SRE and indirect estimates, with sample size as bubbles: 2012/2014/2017



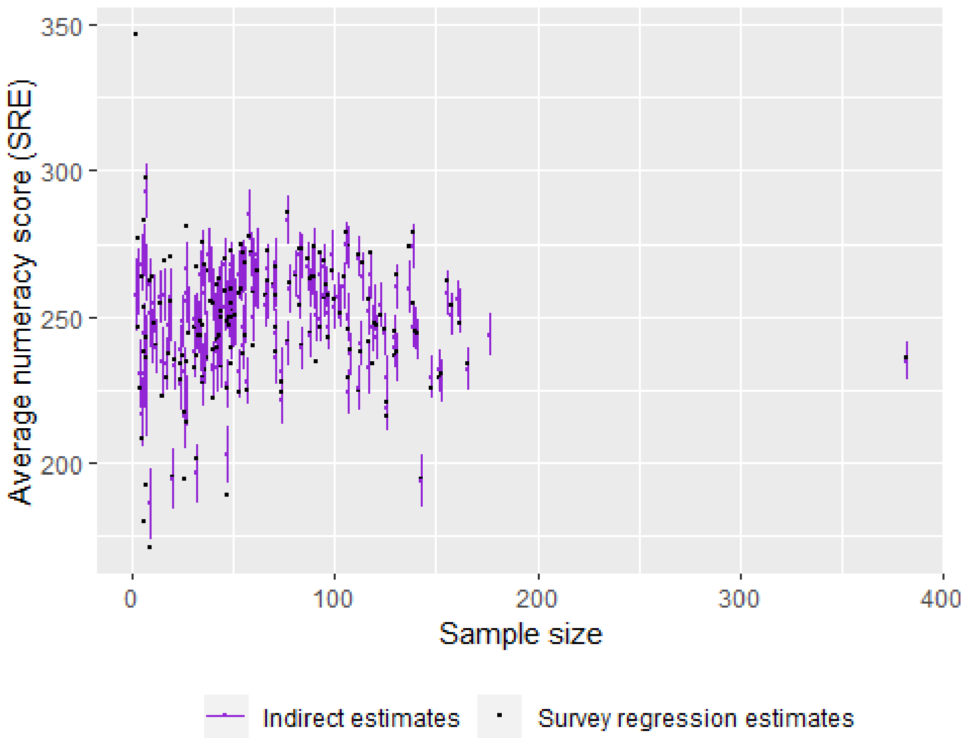
NOTE: SRE: Survey regression estimates.
SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

Figure C-14. Numeracy average—Shrinkage plots of point estimates, by sample size: 2012/2014/2017



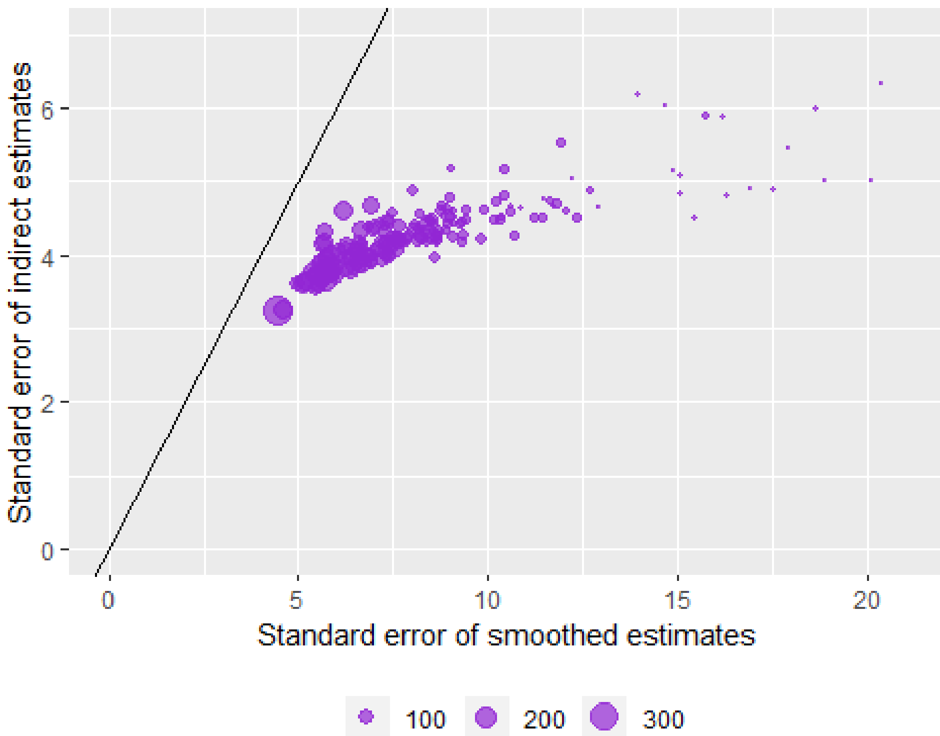
SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

Figure C-15. Numeracy average—Indication of coverage by credible interval: 2012/2014/2017



NOTE: SRE: Survey regression estimates.
SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

Figure C-16. Numeracy average—Comparison of standard errors between model and smoothed approaches: 2012/2014/2017



SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

Table C-5. Evaluation of aggregate estimates for literacy proportion at Level 2: 2012/2014/2017

Subgroup (source: year)	Indirect estimate: Number of counties	Indirect estimate: Weighted estimate	Direct estimate: Sample size	Direct estimate: Estimate	Direct estimate: Standard error	Percentage point difference	Relative difference (percent)
Variables used in the model:							
Percentage of population aged 25+: with education less than high school (ACS: 2013–2017)							
<9.28	862	30.3	4,099	29.8	1.21	0.6	1.9
9.28–13.82	938	32.7	4,168	33.9	1.09	-1.2	-3.4
≥13.82	1,342	33.7	4,063	33.3	1.25	0.4	1.3
Percentage of population aged 25+: with education more than high school (ACS: 2013–2017)							
<56.57	2,107	37.3	4,079	37.3	1.09	0.0	-0.1
56.57–64.27	652	32.4	4,097	33.1	1.12	-0.8	-2.3
≥64.27	383	27.5	4,154	27.2	1.02	0.3	1.0
Percentage of population below 100 percent poverty line (ACS: 2013–2017)							
<11.99	919	30.3	4,102	30.7	1.15	-0.4	-1.2
11.99–16.78	984	33.1	4,097	32.4	1.16	0.7	2.1
≥16.78	1,239	33.5	4,131	33.8	1.27	-0.3	-1.0
Percentage of Blacks (ACS: 2013–2017)							
<4.35	1,937	33.6	4,042	33.6	1.31	0.0	-0.1
4.35–12.91	523	30.7	4,154	30.2	1.16	0.4	1.4
≥12.91	682	33.0	4,134	33.0	1.05	0.0	-0.1
Percentage of Hispanics (ACS: 2013–2017)							
<5.16	1,860	36.6	4,093	35.6	1.24	1.0	2.8
5.16–16.27	810	32.1	4,117	31.3	1.12	0.8	2.7
≥16.27	472	29.8	4,120	30.3	1.10	-0.5	-1.6
Percentage of civilian noninstitutionalized population with no health insurance coverage (ACS: 2013–2017)							
<8.34	1,057	31.5	4,091	29.6	1.06	1.9	6.3
8.34–12.32	1,001	32.2	4,131	33.5	1.43	-1.3	-3.8
≥12.32	1,084	33.6	4,108	33.7	1.30	-0.1	-0.3
Percentage of population aged 16+: service occupation (ACS: 2013–2017)							
<16.37	1,054	30.8	4,025	30.5	1.13	0.2	0.8
16.37–18.60	901	32.7	4,150	32.4	1.18	0.3	0.8
≥18.60	1,187	33.3	4,155	33.6	1.11	-0.3	-1.0
Variables not used in the model:							
Total population (ACS: 2013–2017)							
<164,110	2,754	37.5	4,088	37.1	1.23	0.4	1.0
164,110–837,288	319	31.5	4,076	31.5	1.21	0.0	-0.1
≥837,288	69	29.0	4,166	28.6	1.07	0.3	1.1

See notes at end of table.

Table C-5. Evaluation of aggregate estimates for literacy proportion at Level 2: 2012/2014/2017—Continued

Subgroup (source: year)	Indirect estimate: Number of counties	Indirect estimate: Weighted estimate	Direct estimate: Sample size	Direct estimate: Estimate	Direct estimate: Standard error	Percentage point difference	Relative difference (percent)
Census region (ACS: 2013–2017)							
Northwest	217	32.7	2,140	32.0	1.39	0.7	2.3
Midwest	1,055	33.7	2,995	33.0	1.63	0.7	2.0
South	1,422	34.2	5,058	34.5	1.16	-0.3	-0.9
West	448	28.0	2,137	28.2	1.40	-0.2	-0.8
Beale codes (USDA: 2013)							
Counties in metro area of 1 million population or more	432	30.1	6,636	30.2	0.91	-0.2	-0.6
Counties in metro areas of less than 1 million population	733	33.4	3,616	32.6	1.19	0.8	2.5
Nonmetro counties	1,976	38.9	2,078	38.9	1.64	0.1	0.1
Percentage of population 1+ moved from abroad in the past year (ACS: 2013–2017)							
<0.358	2,207	37.6	4,010	37.2	1.05	0.4	1.2
0.358–0.8207	665	31.4	4,276	32.2	1.06	-0.8	-2.4
≥0.8207	270	28.4	4,044	27.8	1.18	0.6	2.1
Percentage of population aged 5+: speak other language and speak English not at all or not well (ACS: 2013–17)							
<14.54	1,581	34.6	4,104	32.9	1.08	1.6	4.9
14.54–21.24	787	31.6	4,129	33.0	1.31	-1.3	-4.1
≥21.24	774	31.5	4,097	30.8	1.25	0.7	2.2
Percentage of foreign-born people who entered United States after year 2010 (ACS: 2013–2017)							
<12.08	1,668	33.2	4,170	33.1	1.13	0.1	0.3
12.08–17.86	666	31.9	3,977	32.4	1.03	-0.5	-1.7
≥17.86	808	31.7	4,183	31.1	1.27	0.6	1.8

See notes at end of table.

Table C-5. Evaluation of aggregate estimates for literacy proportion at Level 2: 2012/2014/2017—Continued

Subgroup (source: year)	Indirect estimate: Number of counties	Indirect estimate: Weighted estimate	Direct estimate: Sample size	Direct estimate: Estimate	Direct estimate: Standard error	Percentage point difference	Relative difference (percent)
Percentage of population aged 16+ and didn't work at home: less than 30 minutes to work (ACS: 2013–2017)							
<56.15	526	30.4	4,103	31.1	1.19	-0.8	-2.5
56.15–68.69	989	32.7	4,021	32.9	1.33	-0.2	-0.6
≥68.69	1,627	33.8	4,206	32.7	1.27	1.1	3.3
Percentage of population receiving SNAP/Food stamps (ACS: 2013–2017)							
<9.5	883	28.5	4,040	28.7	1.18	-0.2	-0.8
9.5–13.658	839	33.5	4,139	33.4	1.19	0.1	0.3
≥13.658	1,420	34.7	4,151	35.4	1.19	-0.7	-2.0
Median household income—ACS (ACS: 2013–2017)							
<52,017	2,021	36.5	4,070	36.8	1.11	-0.3	-0.7
52,017–62,293	707	32.1	4,120	32.4	1.24	-0.3	-1.0
≥62,293	414	28.4	4,140	28.5	1.14	0.0	-0.1
Percentage of diagnosed diabetes (DDT: 2013)							
<8.7	451	27.8	3,935	27.1	1.08	0.7	2.5
8.7–10.5	785	31.6	4,137	32.8	1.23	-1.3	-3.9
≥10.5	1,906	37.2	4,258	37.3	1.21	-0.1	-0.3
Average amount of grant and scholarship aid received (IPEDS: 2014–2015)							
<7,030	1,226	33.6	4,041	34.0	1.35	-0.4	-1.2
7,030–7,996	949	30.8	4,134	31.5	1.14	-0.7	-2.3
≥7,996	967	32.7	4,155	31.2	1.15	1.5	4.7
Graduation rate of postsecondary institutes (IPEDS: 2014–2015)							
<50	1,122	33.7	3,386	33.9	1.41	-0.1	-0.4
50–57	1,551	33.1	4,738	32.9	1.21	0.2	0.6
≥57	469	30.6	4,206	30.5	0.99	0.1	0.5
Infant mortality rate per 1,000 live births (NCHS: 2013)							
<5.68	767	29.2	4,136	28.9	1.10	0.3	1.1
5.68–6.69	1,139	32.8	4,026	33.9	1.25	-1.1	-3.3
≥6.69	1,236	36.0	4,168	34.9	1.10	1.0	2.9

NOTE: ACS: American Community Survey; USDA: U.S. Department of Agriculture; DDT: Centers for Disease Control and Prevention's Division of Diabetes Translation; IPEDS: Integrated Postsecondary Education Data System; NCHS: National Center for Health Statistics; SNAP: Supplemental Nutrition Assistance Program.
SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

Table C-6. Evaluation of aggregate estimates for literacy proportion at or above Level 3: 2012/2014/2017

	Indirect estimate		Direct estimate			Percentage	Relative
Subgroup (source: year)	Number of counties	Weighted estimate	Sample size	Estimate	Standard error	point difference	difference (percent)
Variables used in the model:							
Percentage of population aged 25+: with education less than high school (ACS: 2013–2017)							
<9.28	862	55.9	4,099	55.7	1.24	0.2	0.4
9.28–13.82	938	46.9	4,168	44.0	1.30	2.8	6.4
≥13.82	1,342	35.8	4,063	35.2	1.53	0.6	1.8
Percentage of population aged 25+: with education more than high school (ACS: 2013–17)							
<56.57	2,107	35.9	4,079	32.9	1.25	3.0	9.1
56.57–64.27	652	45.4	4,097	45.5	1.54	-0.1	-0.1
≥64.27	383	56.0	4,154	55.0	1.27	0.9	1.7
Percentage of population below 100 percent poverty line (ACS: 2013–2017)							
<11.99	919	54.2	4,102	53.2	1.18	1.0	2.0
11.99–16.78	984	45.6	4,097	44.7	1.36	0.9	2.0
≥16.78	1,239	37.7	4,131	36.6	1.47	1.2	3.2
Percentage of Blacks (ACS: 2013–2017)							
<4.35	1,937	47.2	4,042	45.4	1.56	1.8	3.9
4.35–12.91	523	48.1	4,154	48.2	1.33	-0.1	-0.2
≥12.91	682	42.5	4,134	41.3	1.56	1.1	2.7
Percentage of Hispanics (ACS: 2013–2017)							
<5.16	1,860	44.9	4,093	45.3	1.80	-0.4	-0.9
5.16–16.27	810	49.6	4,117	49.4	1.46	0.2	0.4
≥16.27	472	43.3	4,120	41.3	1.34	2.0	4.8
Percentage of civilian noninstitutionalized population with no health insurance coverage (ACS: 2013–2017)							
<8.34	1,057	52.3	4,091	53.8	1.22	-1.5	-2.7
8.34–12.32	1,001	45.1	4,131	42.8	1.47	2.4	5.6
≥12.32	1,084	38.6	4,108	38.7	1.61	-0.1	-0.3
Percentage of population aged 16+: service occupation (ACS: 2013–2017)							
<16.37	1,054	52.5	4,025	51.3	1.11	1.2	2.3
16.37–18.60	901	45.8	4,150	46.5	1.28	-0.7	-1.5
≥18.60	1,187	40.5	4,155	38.4	1.37	2.1	5.5
Variables not used in the model:							
Total population (ACS: 2013–2017)							
<164,110	2,754	41.6	4,088	40.7	1.82	0.9	2.2
164,110–837,288	319	49.2	4,076	49.5	1.74	-0.3	-0.7
≥837,288	69	46.0	4,166	45.2	1.24	0.7	1.7

See notes at end of table.

Table C-6. Evaluation of aggregate estimates for literacy proportion at or above Level 3: 2012/2014/2017—Continued

Subgroup (source: year)	Indirect estimate: Number of counties	Indirect estimate: Weighted estimate	Direct estimate: Sample size	Direct estimate: Estimate	Direct estimate: Standard error	Percentage point difference	Relative difference (percent)
Census region (ACS: 2013–2017)							
Northwest	217	47.0	2,140	48.1	1.54	-1.1	-2.3
Midwest	1,055	48.8	2,995	48.9	2.08	-0.1	-0.2
South	1,422	42.1	5,058	40.3	1.50	1.8	4.4
West	448	48.3	2,137	47.4	1.39	0.9	1.9
Beale codes (USDA: 2013)							
Counties in metro area of 1 million population or more	432	47.8	6,636	47.5	0.93	0.3	0.6
Counties in metro areas of less than 1 million population	733	45.6	3,616	45.3	1.62	0.3	0.6
Nonmetro counties	1,976	38.7	2,078	36.4	2.34	2.3	6.3
Percentage of population 1+ moved from abroad in the past year (ACS: 2013–2017)							
<0.358	2,207	41.8	4,010	41.3	1.62	0.5	1.3
0.358–0.8207	665	46.6	4,276	45.7	1.27	1.0	2.1
≥0.8207	270	48.6	4,044	48.2	1.58	0.5	1.0
Percentage of population aged 5+: speak other language and speak English not at all or not well (ACS: 2013–17)							
<14.54	1,581	48.0	4,104	50.2	1.61	-2.1	-4.2
14.54–21.24	787	48.3	4,129	45.3	1.37	3.0	6.6
≥21.24	774	40.5	4,097	40.1	1.44	0.5	1.2
Percentage of foreign-born people who entered United States after year 2010 (ACS: 2013–2017)							
<12.08	1,668	43.4	4,170	42.7	1.49	0.7	1.7
12.08–17.86	666	46.0	3,977	42.8	1.33	3.1	7.3
≥17.86	808	49.8	4,183	50.3	1.51	-0.5	-1.0

See notes at end of table.

Table C-6. Evaluation of aggregate estimates for literacy proportion at or above Level 3: 2012/2014/2017—Continued

Subgroup (source: year)	Indirect estimate: Number of counties	Indirect estimate: Weighted estimate	Direct estimate: Sample size	Direct estimate: Estimate	Direct estimate: Standard error	Percentage point difference	Relative difference (percent)
Percentage of population aged 16+ and didn't work at home: less than 30 minutes to work (ACS: 2013–2017)							
<56.15	526	46.1	4,103	45.8	1.56	0.3	0.8
56.15–68.69	989	46.2	4,021	44.6	1.42	1.6	3.5
≥68.69	1,627	45.2	4,206	45.2	1.61	0.1	0.1
Percentage of population receiving SNAP/Food stamps (ACS: 2013–2017)							
<9.5	883	53.6	4,040	52.3	1.33	1.3	2.5
9.5–13.658	839	46.1	4,139	45.5	1.48	0.6	1.3
≥13.658	1,420	39.0	4,151	35.9	1.30	3.2	8.8
Median household income—ACS (ACS: 2013–2017)							
<52,017	2,021	38.3	4,070	35.8	1.37	2.5	7.0
52,017–62,293	707	44.5	4,120	43.9	1.55	0.6	1.4
≥62,293	414	54.6	4,140	53.7	1.25	0.9	1.7
Percentage of diagnosed diabetes (DDT: 2013)							
<8.7	451	52.8	3,935	52.7	1.53	0.1	0.2
8.7–10.5	785	45.5	4,137	43.8	1.31	1.8	4.0
≥10.5	1,906	40.0	4,258	38.4	1.79	1.5	4.0
Average amount of grant and scholarship aid received (IPEDS: 2014–2015)							
<7,030	1,226	43.2	4,041	41.5	1.59	1.7	4.1
7,030–7,996	949	46.3	4,134	44.0	1.38	2.3	5.2
≥7,996	967	48.1	4,155	50.2	1.39	-2.0	-4.0
Graduation rate of postsecondary institutes (IPEDS: 2014–2015)							
<50	1,122	45.7	3,386	47.0	1.76	-1.3	-2.7
50–57	1,551	45.9	4,738	44.0	1.65	1.9	4.4
≥57	469	45.8	4,206	45.1	1.46	0.7	1.6
Infant mortality rate per 1,000 live births (NCHS: 2013)							
<5.68	767	48.8	4,136	48.5	1.19	0.3	0.5
5.68–6.69	1,139	44.6	4,026	41.7	1.85	2.9	6.9
≥6.69	1,236	43.4	4,168	44.4	1.39	-1.0	-2.2

NOTE: ACS: American Community Survey; USDA: U.S. Department of Agriculture; DDT: Centers for Disease Control and Prevention's Division of Diabetes Translation; IPEDS: Integrated Postsecondary Education Data System; NCHS: National Center for Health Statistics; SNAP: Supplemental Nutrition Assistance Program.

SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

Table C-7. Evaluation of aggregate estimates for literacy average: 2012/2014/2017

Subgroup (source: year)	Indirect estimate: Number of counties	Indirect estimate: Weighted estimate	Direct estimate: Sample size	Direct estimate: Estimate	Direct estimate: Standard error	Percentage point difference	Relative difference (percent)
Variables used in the model:							
Percentage of population aged 25+: with education less than high school (ACS: 2013–2017)							
<9.28	862	277.0	4,099	277.0	1.39	0.0	0.0
9.28–13.82	938	265.5	4,168	263.7	1.51	1.8	0.7
≥13.82	1,342	249.4	4,063	248.5	1.82	0.9	0.4
Percentage of population aged 25+: with education more than high school (ACS: 2013–2017)							
<56.57	2,107	252.7	4,079	249.2	1.62	3.5	1.4
56.57–64.27	652	262.7	4,097	264.0	1.79	-1.4	-0.5
≥64.27	383	274.9	4,154	274.4	1.65	0.6	0.2
Percentage of population below 100 percent poverty line (ACS: 2013–2017)							
<11.99	919	274.3	4,102	273.8	1.45	0.5	0.2
11.99–16.78	984	263.7	4,097	262.8	1.52	0.9	0.3
≥16.78	1,239	252.6	4,131	252.0	1.83	0.6	0.2
Percentage of Blacks (ACS: 2013–2017)							
<4.35	1,937	266.2	4,042	264.0	1.72	2.2	0.8
4.35–12.91	523	265.3	4,154	266.3	1.60	-1.0	-0.4
≥12.91	682	259.6	4,134	259.1	2.12	0.5	0.2
Percentage of Hispanics (ACS: 2013–2017)							
<5.16	1,860	266.3	4,093	266.8	1.94	-0.4	-0.2
5.16–16.27	810	269.4	4,117	269.5	1.74	-0.2	-0.1
≥16.27	472	256.8	4,120	254.9	1.67	1.8	0.7
Percentage of civilian noninstitutionalized population with no health insurance coverage (ACS: 2013–2017)							
<8.34	1,057	272.8	4,091	274.4	1.48	-1.6	-0.6
8.34–12.32	1,001	262.4	4,131	260.3	1.41	2.1	0.8
≥12.32	1,084	253.3	4,108	254.9	2.03	-1.6	-0.6
Percentage of population aged 16+: service occupation (ACS: 2013–2017)							
<16.37	1,054	272.8	4,025	271.7	1.31	1.1	0.4
16.37–18.60	901	263.9	4,150	265.4	1.52	-1.6	-0.6
≥18.60	1,187	255.7	4,155	253.7	1.69	1.9	0.8
Variables not used in the model:							
Total population (ACS: 2013–2017)							
<164,110	2,754	261.7	4,088	261.2	2.07	0.5	0.2
164,110–837,288	319	268.0	4,076	269.0	2.03	-1.0	-0.4
≥837,288	69	260.6	4,166	260.2	1.61	0.3	0.1
Census region (ACS: 2013–2017)							
Northwest	217	265.4	2,140	267.7	1.76	-2.3	-0.9
Midwest	1,055	269.6	2,995	269.2	2.24	0.4	0.2
South	1,422	259.8	5,058	258.6	1.76	1.2	0.5
West	448	262.7	2,137	262.3	1.46	0.3	0.1

See notes at end of table.

Table C-7. Evaluation of aggregate estimates for literacy average: 2012/2014/2017—Continued

Subgroup (source: year)	Indirect estimate: Number of counties	Indirect estimate: Weighted estimate	Direct estimate: Sample size	Direct estimate: Estimate	Direct estimate: Standard error	Percentage point difference	Relative difference (percent)
Beale codes (USDA: 2013)							
Counties in metro area of 1 million population or more	432	264.5	6,636	265.1	1.13	-0.6	-0.2
Counties in metro areas of less than 1 million population	733	264.1	3,616	263.6	1.87	0.5	0.2
Nonmetro counties	1,976	258.4	2,078	256.4	2.70	2.0	0.8
Percentage of population 1+ moved from abroad in the past year (ACS: 2013–2017)							
<0.358	2,207	262.2	4,010	262.0	1.82	0.2	0.1
0.358–0.8207	665	263.9	4,276	263.7	1.59	0.2	0.1
≥0.8207	270	264.3	4,044	264.2	2.01	0.1	0.0
Percentage of population aged 5+: speak other language and speak English not at all or not well (ACS: 2013–2017)							
<14.54	1,581	269.0	4,104	272.0	1.79	-3.0	-1.1
14.54–21.24	787	266.6	4,129	264.3	1.35	2.4	0.9
≥21.24	774	254.6	4,097	253.7	1.54	0.9	0.4
Percentage of foreign-born people who entered United States after year 2010 (ACS: 2013–2017)							
<12.08	1,668	260.2	4,170	259.9	1.64	0.2	0.1
12.08–17.86	666	263.2	3,977	259.3	1.61	3.9	1.5
≥17.86	808	269.8	4,183	271.2	1.64	-1.4	-0.5
Percentage of population aged 16+ and didn't work at home: less than 30 minutes to work (ACS: 2013–17)							
<56.15	526	261.8	4,103	263.4	1.98	-1.6	-0.6
56.15–68.69	989	264.6	4,021	263.1	1.50	1.5	0.6
≥68.69	1,627	264.0	4,206	263.6	1.81	0.4	0.2
Percentage of population receiving SNAP/Food stamps (ACS: 2013–2017)							
<9.5	883	271.7	4,040	270.4	1.48	1.3	0.5
9.5–13.658	839	264.7	4,139	264.9	1.64	-0.2	-0.1
≥13.658	1,420	255.6	4,151	252.8	1.73	2.9	1.1

See notes at end of table.

Table C-7. Evaluation of aggregate estimates for literacy average: 2012/2014/2017—Continued

Subgroup (source: year)	Indirect estimate: Number of counties	Indirect estimate: Weighted estimate	Direct estimate: Sample size	Direct estimate: Estimate	Direct estimate: Standard error	Percentage point difference	Relative difference (percent)
Median household income—ACS (ACS: 2013–2017)							
<52,017	2,021	256.1	4,070	253.4	1.64	2.7	1.0
52,017–62,293	707	261.1	4,120	261.6	1.79	-0.6	-0.2
≥62,293	414	273.4	4,140	272.7	1.69	0.7	0.3
Percentage of diagnosed diabetes (DDT: 2013)							
<8.7	451	269.8	3,935	270.0	2.04	-0.2	-0.1
8.7–10.5	785	262.3	4,137	261.8	1.56	0.5	0.2
≥10.5	1,906	259.2	4,258	257.7	2.02	1.5	0.6
Average amount of grant and scholarship aid received (IPEDS: 2014–2015)							
<7,030	1,226	260.4	4,041	259.3	1.77	1.1	0.4
7,030–7,996	949	262.7	4,134	260.6	1.49	2.2	0.8
≥7,996	967	267.8	4,155	270.5	1.58	-2.7	-1.0
Graduation rate of postsecondary institutes (IPEDS: 2014–2015)							
<50	1,122	265.4	3,386	267.9	1.89	-2.4	-0.9
50–57	1,551	264.2	4,738	262.4	1.88	1.8	0.7
≥57	469	261.5	4,206	261.1	1.68	0.4	0.2
Infant mortality rate per 1,000 live births (NCHS: 2013)							
<5.68	767	264.6	4,136	264.8	1.39	-0.2	-0.1
5.68–6.69	1,139	261.8	4,026	259.2	2.14	2.5	1.0
≥6.69	1,236	263.9	4,168	265.6	1.70	-1.6	-0.6

NOTE: ACS: American Community Survey; USDA: U.S. Department of Agriculture; DDT: Centers for Disease Control and Prevention's Division of Diabetes Translation; IPEDS: Integrated Postsecondary Education Data System; NCHS: National Center for Health Statistics; SNAP: Supplemental Nutrition Assistance Program.
SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

Table C-8. Evaluation of aggregate estimates for numeracy proportion at or below Level 1: 2012/2014/2017

Subgroup (source: year)	Indirect estimate: Number of counties	Indirect estimate: Weighted estimate	Direct estimate: Sample size	Direct estimate: Estimate	Direct estimate: Standard error	Percentage point difference	Relative difference (percent)
Variables used in the model:							
Percentage of population aged 25+: with education less than high school (ACS: 2013–2017)							
<9.28	862	21.7	4,099	22.0	0.98	-0.3	-1.4
9.28–13.82	938	30.6	4,168	32.0	1.10	-1.4	-4.4
≥13.82	1,342	42.2	4,063	42.8	1.69	-0.6	-1.4
Percentage of population aged 25+: with education more than high school (ACS: 2013–2017)							
<56.57	2,107	39.0	4,079	42.3	1.55	-3.3	-7.8
56.57–64.27	652	32.5	4,097	30.4	1.49	2.1	7.0
≥64.27	383	24.3	4,154	25.3	1.24	-1.0	-3.9
Percentage of population below 100 percent poverty line (ACS: 2013–2017)							
<11.99	919	23.2	4,102	23.6	1.14	-0.4	-1.7
11.99–16.78	984	31.4	4,097	32.2	1.16	-0.8	-2.4
≥16.78	1,239	41.0	4,131	41.7	1.55	-0.7	-1.6
Percentage of Blacks (ACS: 2013–2017)							
<4.35	1,937	28.2	4,042	29.7	1.49	-1.6	-5.2
4.35–12.91	523	30.1	4,154	29.9	1.19	0.2	0.7
≥12.91	682	36.7	4,134	37.3	1.45	-0.7	-1.8
Percentage of Hispanics (ACS: 2013–17)							
<5.16	1,860	29.4	4,093	29.2	1.53	0.2	0.8
5.16–16.27	810	28.1	4,117	28.7	1.32	-0.6	-2.2
≥16.27	472	36.6	4,120	37.5	1.54	-0.9	-2.4
Percentage of civilian noninstitutionalized population with no health insurance coverage (ACS: 2013–2017)							
<8.34	1,057	24.7	4,091	24.3	1.18	0.5	1.9
8.34–12.32	1,001	33.0	4,131	33.7	1.26	-0.7	-2.1
≥12.32	1,084	39.3	4,108	38.5	1.75	0.8	2.2
Percentage of population aged 16+: service occupation (ACS: 2013–2017)							
<16.37	1,054	25.1	4,025	26.2	1.03	-1.1	-4.1
16.37–18.60	901	31.7	4,150	30.8	1.29	0.9	2.9
≥18.60	1,187	37.4	4,155	38.6	1.52	-1.2	-3.1
Variables not used in the model:							
Total population (ACS: 2013–2017)							
<164,110	2,754	32.1	4,088	33.0	1.65	-1.0	-3.0
164,110–837,288	319	28.9	4,076	28.1	1.69	0.8	2.8
≥837,288	69	34.7	4,166	34.7	1.13	-0.1	-0.2

See notes at end of table.

Table C-8. Evaluation of aggregate estimates for numeracy proportion at or below Level 1: 2012/2014/2017—Continued

Subgroup (source: year)	Indirect estimate		Direct estimate			Percentage point difference	Relative difference (percent)
	Number of counties	Weighted estimate	Sample size	Estimate	Standard error		
Census region (ACS: 2013–2017)							
Northwest	217	30.0	2,140	28.7	1.49	1.3	4.5
Midwest	1,055	27.3	2,995	26.6	1.49	0.7	2.6
South	1,422	35.3	5,058	36.7	1.35	-1.4	-3.9
West	448	31.9	2,137	32.2	1.51	-0.3	-0.9
Beale codes (USDA: 2013)							
Counties in metro area of 1 million population or more	432	31.6	6,636	31.0	1.02	0.5	1.7
Counties in metro areas of less than 1 million population	733	31.2	3,616	31.5	1.47	-0.3	-0.9
Nonmetro counties	1,976	34.2	2,078	36.8	2.25	-2.6	-7.0
Percentage of population 1+ moved from abroad in the past year (ACS: 2013–2017)							
<0.358	2,207	31.7	4,010	32.2	1.47	-0.5	-1.5
0.358–0.8207	665	31.7	4,276	31.5	1.37	0.1	0.4
≥0.8207	270	32.3	4,044	32.5	1.49	-0.2	-0.6
Percentage of population aged 5+: speak other language and speak English not at all or not well (ACS: 2013–2017)							
<14.54	1,581	27.3	4,104	26.0	1.42	1.3	5.1
14.54–21.24	787	29.6	4,129	31.2	1.14	-1.6	-5.1
≥21.24	774	38.8	4,097	39.1	1.30	-0.3	-0.7
Percentage of foreign-born people who entered United States after year 2010 (ACS: 2013–2017)							
<12.08	1,668	33.4	4,170	33.5	1.47	-0.2	-0.5
12.08–17.86	666	32.2	3,977	34.1	1.55	-1.9	-5.5
≥17.86	808	28.7	4,183	28.4	1.19	0.2	0.9
Percentage of population aged 16+ and didn't work at home: less than 30 minutes to work (ACS: 2013–17)							
<56.15	526	33.1	4,103	31.6	1.56	1.5	4.7
56.15–68.69	989	31.0	4,021	32.3	1.20	-1.3	-4.0
≥68.69	1,627	31.6	4,206	32.3	1.53	-0.7	-2.2
Percentage of population receiving SNAP/Food stamps (ACS: 2013–2017)							
<9.5	883	25.4	4,040	26.1	1.23	-0.7	-2.8
9.5–13.658	839	30.5	4,139	30.6	1.32	-0.1	-0.4
≥13.658	1,420	38.4	4,151	41.1	1.47	-2.7	-6.5

See notes at end of table.

Table C-8. Evaluation of aggregate estimates for numeracy proportion at or below Level 1: 2012/2014/2017—Continued

Subgroup (source: year)	Indirect estimate: Number of counties	Indirect estimate: Weighted estimate	Direct estimate: Sample size	Direct estimate: Estimate	Direct estimate: Standard error	Percentage point difference	Relative difference (percent)
Median household income—ACS (ACS: 2013–2017)							
<52,017	2,021	37.6	4,070	39.7	1.47	-2.1	-5.2
52,017–62,293	707	33.5	4,120	33.2	1.36	0.3	0.9
≥62,293	414	24.5	4,140	25.1	1.36	-0.6	-2.4
Percentage of diagnosed diabetes (DDT: 2013)							
<8.7	451	27.5	3,935	27.7	1.63	-0.2	-0.9
8.7–10.5	785	32.5	4,137	32.5	1.35	0.0	0.0
≥10.5	1,906	35.1	4,258	36.4	1.73	-1.3	-3.6
Average amount of grant and scholarship aid received (IPEDS: 2014–2015)							
<7,030	1,226	33.9	4,041	35.0	1.61	-1.1	-3.1
7,030–7,996	949	32.4	4,134	33.7	1.43	-1.3	-3.9
≥7,996	967	29.0	4,155	27.3	1.20	1.7	6.2
Graduation rate of postsecondary institutes (IPEDS: 2014–2015)							
<50	1,122	31.5	3,386	29.2	1.54	2.3	8.0
50–57	1,551	31.2	4,738	33.3	1.58	-2.2	-6.5
≥57	469	32.9	4,206	32.9	1.36	0.0	0.1
Infant mortality rate per 1,000 live births (NCHS: 2013)							
<5.68	767	30.6	4,136	30.4	1.26	0.2	0.6
5.68–6.69	1,139	32.9	4,026	34.3	1.72	-1.4	-4.2
≥6.69	1,236	32.4	4,168	32.0	1.44	0.4	1.4

NOTE: ACS: American Community Survey; USDA: U.S. Department of Agriculture; DDT: Centers for Disease Control and Prevention's Division of Diabetes Translation; IPEDS: Integrated Postsecondary Education Data System; NCHS: National Center for Health Statistics; SNAP: Supplemental Nutrition Assistance Program.

SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

Table C-9. Evaluation of aggregate estimates for numeracy proportion at Level 2: 2012/2014/2017

Subgroup (source: year)	Indirect estimate: Number of counties	Indirect estimate: Weighted estimate	Direct estimate: Sample size	Direct estimate: Estimate	Direct estimate: Standard error	Percentage point difference	Relative difference (percent)
Variables used in the model:							
Percentage of population aged 25+: with education less than high school (ACS: 2013–2017)							
<9.28	862	33.1	4,099	32.0	1.04	1.1	3.3
9.28–13.82	938	32.8	4,168	33.4	1.26	-0.6	-1.9
≥13.82	1,342	30.7	4,063	30.8	1.18	-0.1	-0.3
Percentage of population aged 25+: with education more than high school (ACS: 2013–2017)							
<56.57	2,107	34.8	4,079	34.0	1.04	0.8	2.2
56.57–64.27	652	32.2	4,097	34.0	1.25	-1.9	-5.6
≥64.27	383	29.7	4,154	28.6	1.28	1.1	3.8
Percentage of population below 100 percent poverty line (ACS: 2013–2017)							
<11.99	919	32.9	4,102	32.8	1.10	0.1	0.3
11.99–16.78	984	33.2	4,097	32.5	1.20	0.7	2.1
≥16.78	1,239	30.3	4,131	30.8	1.26	-0.5	-1.6
Percentage of Blacks (ACS: 2013–2017)							
<4.35	1,937	34.8	4,042	34.3	1.13	0.5	1.4
4.35–12.91	523	31.5	4,154	31.1	1.22	0.5	1.5
≥12.91	682	30.7	4,134	30.7	1.21	-0.1	-0.2
Percentage of Hispanics (ACS: 2013–2017)							
<5.16	1,860	36.1	4,093	35.9	1.16	0.2	0.6
5.16–16.27	810	32.6	4,117	31.5	1.25	1.1	3.6
≥16.27	472	29.3	4,120	29.4	1.11	-0.1	-0.5
Percentage of civilian noninstitutionalized population with no health insurance coverage (ACS: 2013–17)							
<8.34	1,057	33.1	4,091	31.2	1.22	1.9	6.2
8.34–12.32	1,001	31.5	4,131	33.3	1.09	-1.8	-5.5
≥12.32	1,084	31.8	4,108	31.8	1.30	0.1	0.2
Percentage of population aged 16+: service occupation (ACS: 2013–2017)							
<16.37	1,054	32.5	4,025	31.6	1.39	0.9	2.9
16.37–18.60	901	32.4	4,150	32.5	1.10	-0.2	-0.5
≥18.60	1,187	31.7	4,155	32.1	1.29	-0.4	-1.1
Variables not used in the model:							
Total population (ACS: 2013–2017)							
<164,110	2,754	36.7	4,088	36.4	1.09	0.3	0.7
164,110–837,288	319	32.1	4,076	31.4	1.21	0.7	2.2
≥837,288	69	28.6	4,166	29.0	1.07	-0.4	-1.2

See notes at end of table.

Table C-9. Evaluation of aggregate estimates for numeracy proportion at Level 2: 2012/2014/2017—Continued

Subgroup (source: year)	Indirect estimate: Number of counties	Indirect estimate: Weighted estimate	Direct estimate: Sample size	Direct estimate: Estimate	Direct estimate: Standard error	Percentage point difference	Relative difference (percent)
Census region (ACS: 2013–2017)							
Northwest	217	32.8	2,140	32.5	1.42	0.3	1.0
Midwest	1,055	34.3	2,995	34.7	1.44	-0.3	-1.0
South	1,422	32.8	5,058	32.3	1.29	0.5	1.4
West	448	28.8	2,137	29.1	1.48	-0.2	-0.7
Beale codes (USDA: 2013)							
Counties in metro area of 1 million population or more	432	30.3	6,636	30.5	0.87	-0.2	-0.6
Counties in metro areas of less than 1 million population	733	33.2	3,616	32.9	1.33	0.3	1.0
Nonmetro counties	1,976	37.3	2,078	36.5	1.48	0.9	2.4
Percentage of population 1+ moved from abroad in the past year (ACS: 2013–2017)							
<0.358	2,207	36.9	4,010	36.8	1.14	0.0	0.1
0.358–0.8207	665	31.5	4,276	32.0	0.99	-0.5	-1.7
≥0.8207	270	28.6	4,044	27.9	1.37	0.7	2.4
Percentage of population aged 5+: speak other language and speak English not at all or not well (ACS: 2013–17)							
<14.54	1,581	35.3	4,104	33.6	1.24	1.6	4.9
14.54–21.24	787	32.1	4,129	32.8	1.26	-0.6	-2.0
≥21.24	774	29.7	4,097	29.8	1.05	-0.1	-0.5
Percentage of foreign-born people who entered United States after year 2010 (ACS: 2013–2017)							
<12.08	1,668	33.0	4,170	33.3	1.22	-0.2	-0.7
12.08–17.86	666	31.7	3,977	32.1	1.27	-0.3	-1.1
≥17.86	808	31.5	4,183	30.8	1.21	0.7	2.2

See notes at end of table.

Table C-9. Evaluation of aggregate estimates for numeracy proportion at Level 2: 2012/2014/2017—Continued

Subgroup (source: year)	Indirect estimate: Number of counties	Indirect estimate: Weighted estimate	Direct estimate: Sample size	Direct estimate: Estimate	Direct estimate: Standard error	Percentage point difference	Relative difference (percent)
Percentage of population aged 16+ and didn't work at home: less than 30 minutes to work (ACS: 2013–2017)							
<56.15	526	30.3	4,103	30.8	1.30	-0.5	-1.6
56.15–68.69	989	32.8	4,021	33.1	1.20	-0.2	-0.7
≥68.69	1,627	33.3	4,206	32.5	1.13	0.8	2.5
Percentage of population receiving SNAP/Food stamps (ACS: 2013–2017)							
<9.5	883	30.9	4,040	30.6	1.10	0.3	1.1
9.5–13.658	839	33.7	4,139	33.8	1.10	-0.1	-0.2
≥13.658	1,420	32.1	4,151	32.2	1.30	-0.1	-0.3
Median household income—ACS (ACS: 2013–2017)							
<52,017	2,021	34.0	4,070	34.3	1.14	-0.2	-0.7
52,017–62,293	707	31.7	4,120	32.1	1.05	-0.4	-1.3
≥62,293	414	30.8	4,140	30.3	1.30	0.5	1.7
Percentage of diagnosed diabetes (DDT: 2013)							
<8.7	451	29.4	3,935	28.2	1.22	1.2	4.4
8.7–10.5	785	31.7	4,137	33.0	1.13	-1.3	-4.0
≥10.5	1,906	35.1	4,258	35.4	1.09	-0.3	-0.8
Average amount of grant and scholarship aid received (IPEDS: 2014–2015)							
<7,030	1,226	32.9	4,041	32.8	1.39	0.0	0.1
7,030–7,996	949	30.9	4,134	31.2	1.18	-0.3	-1.0
≥7,996	967	32.9	4,155	32.4	1.21	0.6	1.8
Graduation rate of postsecondary institutes (IPEDS: 2014–2015)							
<50	1,122	33.0	3,386	34.5	1.28	-1.5	-4.4
50–57	1,551	32.8	4,738	31.1	1.30	1.8	5.7
≥57	469	30.9	4,206	31.4	1.01	-0.4	-1.4
Infant mortality rate per 1,000 live births (NCHS: 2013)							
<5.68	767	29.9	4,136	29.8	1.08	0.2	0.6
5.68–6.69	1,139	32.7	4,026	33.1	1.25	-0.4	-1.1
≥6.69	1,236	34.5	4,168	34.1	1.22	0.5	1.4

NOTE: ACS: American Community Survey; USDA: U.S. Department of Agriculture; DDT: Centers for Disease Control and Prevention's Division of Diabetes Translation; IPEDS: Integrated Postsecondary Education Data System; NCHS: National Center for Health Statistics; SNAP: Supplemental Nutrition Assistance Program.
SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

Table C-10. Evaluation of aggregate estimates for numeracy proportion at or above Level 3: 2012/2014/2017

	Indirect estimate		Direct estimate				
Subgroup (source: year)	Number of counties	Weighted estimate	Sample size	Estimate	Standard error	Percentage point difference	Relative difference (percent)
Variables used in the model:							
Percentage of population aged 25+: with education less than high school (ACS: 2013–2017)							
<9.28	862	45.2	4,099	46.0	1.41	-0.7	-1.6
9.28–13.82	938	36.6	4,168	34.6	1.47	2.1	6.0
≥13.82	1,342	27.1	4,063	26.4	1.29	0.7	2.7
Percentage of population aged 25+: with education more than high school (ACS: 2013–2017)							
<56.57	2,107	26.2	4,079	23.7	1.05	2.5	10.7
56.57–64.27	652	35.4	4,097	35.6	1.39	-0.2	-0.6
≥64.27	383	46.1	4,154	46.1	1.43	-0.1	-0.2
Percentage of population below 100 percent poverty line (ACS: 2013–2017)							
<11.99	919	43.9	4,102	43.6	1.39	0.3	0.6
11.99–16.78	984	35.4	4,097	35.4	1.40	0.1	0.2
≥16.78	1,239	28.7	4,131	27.5	1.41	1.2	4.2
Percentage of Blacks (ACS: 2013–2017)							
<4.35	1,937	37.0	4,042	35.9	1.38	1.1	3.0
4.35–12.91	523	38.4	4,154	39.1	1.48	-0.7	-1.7
≥12.91	682	32.7	4,134	31.9	1.58	0.8	2.4
Percentage of Hispanics (ACS: 2013–2017)							
<5.16	1,860	34.5	4,093	35.0	1.81	-0.5	-1.3
5.16–16.27	810	39.3	4,117	39.8	1.57	-0.5	-1.2
≥16.27	472	34.2	4,120	33.1	1.29	1.0	3.2
Percentage of civilian noninstitutionalized population with no health insurance coverage (ACS: 2013–2017)							
<8.34	1,057	42.1	4,091	44.5	1.38	-2.4	-5.4
8.34–12.32	1,001	35.5	4,131	33.0	1.32	2.5	7.6
≥12.32	1,084	28.8	4,108	29.7	1.59	-0.9	-3.0
Percentage of population aged 16+: service occupation (ACS: 2013–2017)							
<16.37	1,054	42.4	4,025	42.2	1.41	0.1	0.4
16.37–18.60	901	35.9	4,150	36.6	1.26	-0.7	-2.0
≥18.60	1,187	30.8	4,155	29.3	1.22	1.6	5.3
Variables not used in the model:							
Total population (ACS: 2013–2017)							
<164,110	2,754	31.3	4,088	30.5	1.80	0.7	2.4
164,110–837,288	319	39.1	4,076	40.5	1.74	-1.5	-3.6
≥837,288	69	36.7	4,166	36.3	1.32	0.4	1.2

See notes at end of table.

Table C-10. Evaluation of aggregate estimates for numeracy proportion at or above Level 3: 2012/2014/2017—Continued

Subgroup (source: year)	Indirect estimate: Number of counties	Indirect estimate: Weighted estimate	Direct estimate: Sample size	Direct estimate: Estimate	Direct estimate: Standard error	Percentage point difference	Relative difference (percent)
Census region (ACS: 2013–2017)							
Northwest	217	37.2	2,140	38.8	1.41	-1.6	-4.2
Midwest	1,055	38.4	2,995	38.8	2.20	-0.4	-0.9
South	1,422	31.9	5,058	30.9	1.66	1.0	3.2
West	448	39.3	2,137	38.8	1.45	0.5	1.3
Beale codes (USDA: 2013)							
Counties in metro area of 1 million population or more	432	38.1	6,636	38.5	1.00	-0.3	-0.9
Counties in metro areas of less than 1 million population	733	35.5	3,616	35.6	1.68	0.0	-0.1
Nonmetro counties	1,976	28.4	2,078	26.7	2.04	1.7	6.4
Percentage of population 1+ moved from abroad in the past year (ACS: 2013–2017)							
<0.358	2,207	31.4	4,010	30.9	1.48	0.5	1.5
0.358–0.8207	665	36.9	4,276	36.5	1.37	0.4	1.1
≥0.8207	270	39.2	4,044	39.6	1.71	-0.5	-1.1
Percentage of population aged 5+: speak other language and speak English not at all or not well (ACS: 2013–2017)							
<14.54	1,581	37.4	4,104	40.4	1.70	-3.0	-7.3
14.54–21.24	787	38.3	4,129	36.0	1.47	2.2	6.2
≥21.24	774	31.5	4,097	31.1	1.34	0.4	1.3
Percentage of foreign-born people who entered United States after year 2010 (ACS: 2013–2017)							
<12.08	1,668	33.6	4,170	33.2	1.35	0.4	1.2
12.08–17.86	666	36.1	3,977	33.8	1.45	2.2	6.6
≥17.86	808	39.9	4,183	40.8	1.63	-0.9	-2.3
Percentage of population aged 16+ and didn't work at home: less than 30 minutes to work (ACS: 2013–17)							
<56.15	526	36.6	4,103	37.6	1.57	-1.0	-2.6
56.15–68.69	989	36.2	4,021	34.6	1.46	1.5	4.4
≥68.69	1,627	35.2	4,206	35.3	1.54	-0.1	-0.2

See notes at end of table.

Table C-10. Evaluation of aggregate estimates for numeracy proportion at or above Level 3: 2012/2014/2017—Continued

Subgroup (source: year)	Indirect estimate: Number of counties	Indirect estimate: Weighted estimate	Direct estimate: Sample size	Direct estimate: Estimate	Direct estimate: Standard error	Percentage point difference	Relative difference (percent)
Percentage of population receiving SNAP/Food stamps (ACS: 2013–2017)							
<9.5	883	43.7	4,040	43.3	1.29	0.4	1.0
9.5–13.658	839	35.8	4,139	35.6	1.55	0.2	0.6
≥13.658	1,420	29.5	4,151	26.7	1.35	2.8	10.4
Median household income—ACS (ACS: 2013–2017)							
<52,017	2,021	28.4	4,070	26.0	1.27	2.3	8.9
52,017–62,293	707	34.8	4,120	34.7	1.42	0.1	0.3
≥62,293	414	44.7	4,140	44.6	1.41	0.1	0.2
Percentage of diagnosed diabetes (DDT: 2013)							
<8.7	451	43.1	3,935	44.2	1.68	-1.0	-2.3
8.7–10.5	785	35.8	4,137	34.4	1.23	1.3	3.8
≥10.5	1,906	29.8	4,258	28.2	1.61	1.6	5.7
Average amount of grant and scholarship aid received (IPEDS: 2014–2015)							
<7,030	1,226	33.2	4,041	32.2	1.55	1.0	3.3
7,030–7,996	949	36.7	4,134	35.1	1.35	1.6	4.6
≥7,996	967	38.1	4,155	40.4	1.40	-2.3	-5.7
Graduation rate of postsecondary institutes (IPEDS: 2014–2015)							
<50	1,122	35.5	3,386	36.3	1.91	-0.8	-2.3
50–57	1,551	36.0	4,738	35.6	1.72	0.4	1.2
≥57	469	36.2	4,206	35.8	1.40	0.4	1.2
Infant mortality rate per 1,000 live births (NCHS: 2013)							
<5.68	767	39.5	4,136	39.8	1.23	-0.4	-0.9
5.68–6.69	1,139	34.4	4,026	32.6	1.75	1.8	5.6
≥6.69	1,236	33.0	4,168	34.0	1.59	-0.9	-2.7

NOTE: ACS: American Community Survey; USDA: U.S. Department of Agriculture; DDT: Centers for Disease Control and Prevention's Division of Diabetes Translation; IPEDS: Integrated Postsecondary Education Data System; NCHS: National Center for Health Statistics; SNAP: Supplemental Nutrition Assistance Program.
SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

Table C-11. Evaluation of aggregate estimates for numeracy average: 2012/2014/2017

Subgroup (source: year)	Indirect estimate: Number of counties	Indirect estimate: Weighted estimate	Direct estimate: Sample size	Direct estimate: Estimate	Direct estimate: Standard error	Percentage point difference	Relative difference (percent)
Variables used in the model:							
Percentage of population aged 25+: with education less than high school (ACS: 2013–2017)							
<9.28	862	264.9	4,099	264.8	1.59	0.0	0.0
9.28–13.82	938	250.9	4,168	248.8	1.75	2.1	0.8
≥13.82	1,342	233.1	4,063	232.0	2.18	1.0	0.4
Percentage of population aged 25+: with education more than high school (ACS: 2013–2017)							
<56.57	2,107	236.5	4,079	232.0	1.98	4.4	1.9
56.57–64.27	652	247.9	4,097	250.0	1.87	-2.1	-0.8
≥64.27	383	262.4	4,154	261.7	1.87	0.6	0.2
Percentage of population below 100 percent poverty line (ACS: 2013–2017)							
<11.99	919	262.5	4,102	261.9	1.68	0.6	0.2
11.99–16.78	984	249.1	4,097	248.1	1.68	0.9	0.4
≥16.78	1,239	235.7	4,131	235.0	2.19	0.8	0.3
Percentage of Blacks (ACS: 2013–2017)							
<4.35	1,937	253.0	4,042	250.6	2.06	2.4	0.9
4.35–12.91	523	251.7	4,154	252.7	1.80	-1.0	-0.4
≥12.91	682	243.1	4,134	242.4	2.40	0.7	0.3
Percentage of Hispanics (ACS: 2013–2017)							
<5.16	1,860	252.1	4,093	252.6	2.16	-0.5	-0.2
5.16–16.27	810	255.3	4,117	255.1	2.01	0.2	0.1
≥16.27	472	241.9	4,120	240.3	2.18	1.6	0.7
Percentage of civilian noninstitutionalized population with no health insurance coverage (ACS: 2013–2017)							
<8.34	1,057	260.2	4,091	262.2	1.63	-2.0	-0.8
8.34–12.32	1,001	247.5	4,131	245.3	1.73	2.3	0.9
≥12.32	1,084	237.0	4,108	238.8	2.42	-1.8	-0.8
Percentage of population aged 16+: service occupation (ACS: 2013–2017)							
<16.37	1,054	260.3	4,025	259.1	1.48	1.2	0.5
16.37–18.60	901	249.2	4,150	250.6	1.67	-1.4	-0.6
≥18.60	1,187	239.9	4,155	238.0	1.88	1.9	0.8
Variables not used in the model:							
Total population (ACS: 2013–2017)							
<164,110	2,754	247.0	4,088	245.9	2.41	1.0	0.4
164,110–837,288	319	254.0	4,076	255.6	2.48	-1.6	-0.6
≥837,288	69	245.9	4,166	245.6	1.89	0.3	0.1
Census region (ACS: 2013–2017)							
Northwest	217	251.6	2,140	254.2	1.85	-2.7	-1.0
Midwest	1,055	255.7	2,995	255.4	2.53	0.4	0.1
South	1,422	244.0	5,058	242.6	2.14	1.4	0.6
West	448	249.3	2,137	249.2	1.95	0.1	0.1

See notes at end of table.

Table C-11. Evaluation of aggregate estimates for numeracy average: 2012/2014/2017—Continued

Subgroup (source: year)	Indirect estimate: Number of counties	Indirect estimate: Weighted estimate	Direct estimate: Sample size	Direct estimate: Estimate	Direct estimate: Standard error	Percentage point difference	Relative difference (percent)
Beale codes (USDA: 2013)							
Counties in metro area of 1 million population or more	432	250.4	6,636	251.2	1.44	-0.9	-0.3
Counties in metro areas of less than 1 million population	733	249.5	3,616	248.7	2.11	0.8	0.3
Nonmetro counties	1,976	243.1	2,078	240.7	3.05	2.4	1.0
Percentage of population 1+ moved from abroad in the past year (ACS: 2013–2017)							
<0.358	2,207	247.5	4,010	246.8	2.03	0.6	0.3
0.358–0.8207	665	249.6	4,276	249.6	2.03	0.0	0.0
≥0.8207	270	249.9	4,044	250.1	2.43	-0.2	-0.1
Percentage of population aged 5+: speak other language and speak English not at all or not well (ACS: 2013–17)							
<14.54	1,581	255.2	4,104	258.7	2.03	-3.6	-1.4
14.54–21.24	787	252.6	4,129	250.1	1.58	2.5	1.0
≥21.24	774	239.0	4,097	237.8	1.85	1.2	0.5
Percentage of foreign-born people who entered United States after year 2010 (ACS: 2013–2017)							
<12.08	1,668	245.8	4,170	245.7	1.79	0.1	0.0
12.08–17.86	666	248.6	3,977	244.6	2.10	4.0	1.6
≥17.86	808	255.5	4,183	256.8	1.93	-1.3	-0.5
Percentage of population aged 16+ and didn't work at home: less than 30 minutes to work (ACS: 2013–2017)							
<56.15	526	247.7	4,103	250.0	2.26	-2.3	-0.9
56.15–68.69	989	250.2	4,021	248.2	1.80	2.0	0.8
≥68.69	1,627	249.1	4,206	248.5	2.29	0.7	0.3

See notes at end of table.

Table C-11. Evaluation of aggregate estimates for numeracy average: 2012/2014/2017—Continued

Subgroup (source: year)	Indirect estimate: Number of counties	Indirect estimate: Weighted estimate	Direct estimate: Sample size	Direct estimate: Estimate	Direct estimate: Standard error	Percentage point difference	Relative difference (percent)
Percentage of population receiving SNAP/Food stamps (ACS: 2013–2017)							
<9.5	883	259.6	4,040	258.4	1.75	1.2	0.5
9.5–13.658	839	250.3	4,139	250.3	1.84	0.1	0.0
≥13.658	1,420	239.1	4,151	235.4	2.16	3.6	1.5
Median household income—ACS (ACS: 2013–2017)							
<52,017	2,021	239.5	4,070	236.2	1.96	3.2	1.4
52,017–62,293	707	246.4	4,120	247.3	2.01	-0.9	-0.4
≥62,293	414	261.3	4,140	260.3	1.95	1.0	0.4
Percentage of diagnosed diabetes (DDT: 2013)							
<8.7	451	256.8	3,935	257.0	2.44	-0.2	-0.1
8.7–10.5	785	248.0	4,137	247.9	1.78	0.1	0.0
≥10.5	1,906	243.3	4,258	241.1	2.33	2.3	0.9
Average amount of grant and scholarship aid received (IPEDS: 2014–2015)							
<7,030	1,226	245.0	4,041	243.8	2.20	1.2	0.5
7,030–7,996	949	248.7	4,134	246.5	1.86	2.2	0.9
≥7,996	967	253.9	4,155	256.7	1.73	-2.8	-1.1
Graduation rate of postsecondary institutes (IPEDS: 2014–2015)							
<50	1,122	250.3	3,386	253.1	2.25	-2.8	-1.1
50–57	1,551	249.8	4,738	247.7	2.38	2.1	0.8
≥57	469	247.5	4,206	247.1	2.01	0.3	0.1
Infant mortality rate per 1,000 live births (NCHS: 2013)							
<5.68	767	251.3	4,136	251.7	1.74	-0.5	-0.2
5.68–6.69	1,139	247.0	4,026	244.4	2.56	2.6	1.1
≥6.69	1,236	248.4	4,168	249.7	2.07	-1.4	-0.5

NOTE: ACS: American Community Survey; USDA: U.S. Department of Agriculture; DDT: Centers for Disease Control and Prevention's Division of Diabetes Translation; IPEDS: Integrated Postsecondary Education Data System; NCHS: National Center for Health Statistics; SNAP: Supplemental Nutrition Assistance Program.

SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

Appendix D

Negative Estimates

APPENDIX D

NEGATIVE ESTIMATES

This appendix lists the counties with negative indirect estimates of proportions or negative lower boundary for credible intervals. In the Skills Map website, negative estimates were set to zero. Tables D-1 and D-2 provide lists of counties with negative indirect estimates for the proportion at or above Level 3 for literacy and numeracy, respectively. Table D-3 provides a list of counties with positive indirect estimates but negative lower bound of credible intervals for the literacy proportion at or below Level 1. Tables D-4 and D-5 provide lists of counties with positive indirect estimates but negative lower bound of credible intervals for the proportion at or above Level 3 for literacy and numeracy, respectively.

Table D-1. Counties with negative indirect estimates for literacy proportion at or above Level 3: 2012/2014/2017

County	State	Indirect estimate (percent)
Kenedy County	Texas	-5.9

SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

Table D-2. Counties with negative indirect estimates for numeracy proportion at or above Level 3: 2012/2014/2017

County	State	Indirect estimate (percent)
Taliaferro County	Georgia	-4.5
Telfair County	Georgia	-0.1
Wheeler County	Georgia	-0.4
Tensas Parish	Louisiana	-0.1
Holmes County	Mississippi	-0.4
Kenedy County	Texas	-4.4
Starr County	Texas	-0.1
Willacy County	Texas	-1.5

SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

Table D-3. Counties with positive indirect estimate and negative lower bound of credible intervals for literacy proportion at or below Level 1: 2012/2014/2017

County	State	Indirect estimate (percent)	95 percent credible interval	
			Lower bound	Upper bound
Petroleum County	Montana	8.1	-0.4	16.9

SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

Table D-4. Counties with positive indirect estimate and negative lower bound of credible intervals for literacy proportion at or above Level 3: 2012/2014/2017

County	State	Indirect estimate (percent)	95 percent credible interval	
			Lower bound	Upper bound
Hancock County	Georgia	6.2	-2.7	15.2
Taliaferro County	Georgia	3.4	-2.7	9.5
East Carroll Parish	Louisiana	7.1	-0.9	15.5
Holmes County	Mississippi	7.4	-1.4	15.8
Issaquena County	Mississippi	4.6	-2.8	12.3
Jefferson County	Mississippi	6.9	-1.8	15.7
Culberson County	Texas	8.1	-0.6	16.8
Hudspeth County	Texas	1.0	-9.2	11.3
Starr County	Texas	0.2	-8.1	8.7
Willacy County	Texas	6.5	-0.4	13.2
Zapata County	Texas	5.9	-1.1	13.2

SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.

Table D-5. Counties with positive indirect estimates and negative lower bound of credible intervals for numeracy proportion at or above Level 3: 2012/2014/2017

County	State	Indirect estimate (percent)	95 percent credible interval	
			Lower bound	Upper bound
Conecuh County	Alabama	4.5	-2.4	11.2
Greene County	Alabama	7.6	-0.7	15.7
Monroe County	Alabama	7.7	0.0	15.5
Perry County	Alabama	4.8	-3.5	13.4
Sumter County	Alabama	7.1	-0.8	14.8
Wilcox County	Alabama	4.1	-3.3	11.6
Aleutians East Borough	Alaska	13.5	-1.9	29.4
Kusilvak Census Area	Alaska	5.3	-4.3	15.7
Lincoln County	Arkansas	7.1	-0.4	14.6
Hardee County	Florida	5.9	-0.7	12.4
Atkinson County	Georgia	7.5	0.0	15.2
Calhoun County	Georgia	4.6	-1.4	10.7
Clay County	Georgia	0.4	-9.0	9.9
Hancock County	Georgia	0.7	-8.5	10.1
Stewart County	Georgia	2.4	-5.2	10.0

See notes at end of table.

Table D-5. Counties with positive indirect estimates and negative lower bound of credible intervals for numeracy proportion at or above Level 3: 2012/2014/2017—Continued

County	State	Indirect estimate (percent)	95 percent credible interval	
			Lower bound	Upper bound
Warren County	Georgia	4.8	-2.2	12.1
Webster County	Georgia	6.9	-0.4	13.9
Wilcox County	Georgia	4.2	-2.2	10.6
Wilkinson County	Georgia	5.8	-1.3	12.8
East Carroll Parish	Louisiana	3.4	-5.0	12.2
Madison Parish	Louisiana	3.7	-3.8	11.2
Issaquena County	Mississippi	2.2	-5.7	10.2
Jefferson County	Mississippi	1.7	-7.5	11.1
Quitman County	Mississippi	7.6	-0.1	15.4
Walthall County	Mississippi	7.6	-0.2	15.2
Wilkinson County	Mississippi	5.1	-2.7	12.8
Glacier County	Montana	13.3	-1.1	27.9
Guadalupe County	New Mexico	11.2	-1.5	23.7
Holmes County	Ohio	12.8	-2.9	28.9
Forest County	Pennsylvania	6.9	-1.3	15.1
Allendale County	South Carolina	5.0	-2.4	12.7
Buffalo County	South Dakota	7.4	-7.6	23.1
Lake County	Tennessee	6.3	-0.2	12.9
Brooks County	Texas	8.3	-1.3	17.6
Culberson County	Texas	3.6	-5.3	12.8
Dimmit County	Texas	5.6	-1.2	12.0
Duval County	Texas	4.8	-2.3	11.8
Frio County	Texas	5.3	-1.7	12.4
Hudspeth County	Texas	1.2	-9.2	11.8
La Salle County	Texas	3.8	-5.4	13.3
Zapata County	Texas	3.4	-4.5	11.4
Zavala County	Texas	5.1	-3.4	13.1
Greensville County	Virginia	6.5	-1.5	14.7
McDowell County	West Virginia	8.1	-0.5	16.9

SOURCE: U.S. Department of Education, National Center for Education Statistics, U.S. Program for the International Assessment of Adult Competencies (PIAAC), 2012/2014/2017.