United States Department of Agriculture

The National School Lunch Program Direct Certification Improvement Study: Analysis of Unmatched Records

Nutrition Assistance Program Report **August 2014**
Food and Nutrition Service
Office of Policy Support

USDA United States Department of Agriculture

Food and Nutrition Service, Office of Policy Support *August 2014*

The National School Lunch Program Direct Certification Improvement Study: Analysis of Unmatched Records

Authors:
Andrew Gothro
Quinn Moore
Kevin Conway
Brandon Kyler

Submitted by:
Mathematica Policy Research
P.O. Box 2393
Princeton, NJ 08543

Project Director:
Kevin Conway

Submitted to:
Office of Policy Support
Food and Nutrition Service
3101 Park Center Drive
Alexandria, VA 22302

Project Officer(s):
Allison Magness
Joseph F. Robare

This study was conducted under Contract number AG-3198-D-10-0074 with the Food and Nutrition Service, United States Department of Agriculture.

This report is available on the Food and Nutrition website: http://www.fns.usda.gov/research-and-analysis

Suggested Citation:
Moore, Q., Gothro, A., Conway, K., Kyler B., (2014). National School Lunch Program Direct Certification Improvement Study: Analysis of Unmatched Records. Prepared by Mathematica Policy Research. Alexandria, VA: U.S. Department of Agriculture, Food and Nutrition Service, Office of Policy Support, Project Officers: Allison Magness and Joseph F. Robare. Available online at: www.fns.usda.gov/research-and-analysis.

ACKNOWLEDGMENTS

Many people made important contributions to this study. We thank Joseph Robare, Vivian Lees, Patty Von Reyn, and Allison Magness at the U.S. Department of Agriculture (USDA) Food and Nutrition Service (FNS) for their support, guidance, and feedback, all of which were necessary to the successful completion of this report.

We are very grateful to the State agencies that participated in this study: the Department of Human Resources and State Department of Education in Alabama; the Department of Economic Security and Department of Education in Arizona; the Department of Social Services and State Department of Education in Connecticut; the Department of Education and Family Social Services Administration in Indiana; the Department of Education and the Department of Health and Human Services in Nebraska; the Department of Agriculture, the Education Agency, and the Health and Human Services Commission in Texas; and the Department of Education and Department of Health and Human Services in West Virginia. Child nutrition staff members in 24 school districts within these States also provided valuable NSLP application data for our analysis.

Many Mathematica colleagues provided invaluable research assistance and advice. Larry Vittoriano, Taylor Calise, and Andrea Sleeper managed the application data collection effort. Ryan McInerney contributed to qualitative data collection. Lara Hulsey provided quality assurance. Kathryn Gonzalez performed data analysis. John Kennedy and Patricia Ciaccio carefully edited the report. Dawn Patterson provided production support.

GLOSSARY OF ACRONYMS

CEP	Community Eligibility Provision
HH	household
NSLP	National School Lunch Program
PDF	portable document format
POS	point-of-sale
SNAP	Supplemental Nutrition Assistance Program
SSN	Social Security number
SY	school year
TANF	Temporary Assistance for Needy Families

CONTENTS

ACKNOWLEDGEMENTS iii

GLOSSARY OF ACRONYMS v

I INTRODUCTION 1

A. Overview of Approach 1

B. Study Sample 2

C. Data Collection Procedures 3

1. State and District Data 3
2. Data from Publicly Available Sources 3

D. Methods for Independent Matching and Analysis 6

1. Deterministic Match 7
2. Probabilistic Match 7
3. Implications of Using Application Data for Matching 9

E. Organization of Rest of Report 10

II APPROACH TO DATA MATCHING IN SELECTED STATES 11

A. Overview of Current Data Matching Practices and Procedures, by State 11

B. Common Challenges to Data Matching 13

1. Technological Challenges 13
2. Data Challenges 13

III DESCRIPTIVE ANALYSIS OF SNAP RECORDS 15

IV INDEPENDENT MATCH OF SNAP RECORDS TO NSLP APPLICATIONS 19

A. NSLP Application Data 19

B. Matching Results 20

1. Independent Matching Results for Alabama 23
2. Independent Matching Results for Arizona 26
3. Independent Matching Results for Connecticut 29
4. Independent Matching Results for Indiana 30
5. Independent Matching Results for Nebraska 33

6. Independent Matching Results for Texas 36
7. Independent Matching Results for West Virginia 38

C. Cross-State Themes 39

V CONCLUSION 43

APPENDIX A: IN-DEPTH CASE STUDY NSLP DIRECT CERTIFICATION PROFILES

TABLES

I.1 Characteristics of State SNAP Participation Data and Sampled District Data on Students Approved for NSLP Benefits Based on Categorical Eligibility by Application 4

II.1 Characteristics of the Direct Certification Matching Processes in Select States, SY 2012–2013 12

III.1 Average Characteristics of School-Age Children with SNAP Records in Arizona, by Whether Matched to School Enrollment Data (percentage unless otherwise noted) 15

III.2 Average Characteristics of School-Age Children with SNAP Records in West Virginia, by Whether Matched to School Enrollment Data (percentage unless otherwise noted) 16

IV.1 Descriptions of Sampled School Districts, by State 19

IV.2 Average Characteristics of Students Certified for NSLP Benefits Based on Categorical Eligibility by Application, by State (percentages unless otherwise noted) 21

IV.3 Average Characteristics of Students Certified for NSLP Benefits Based on Categorical Eligibility by Application in Alabama, by Whether Matched to State SNAP Data (percentages unless otherwise noted) 25

IV.4 Average Characteristics of Students Certified for NSLP Benefits Based on Categorical Eligibility by Application in Arizona, by Whether Matched to State SNAP Data (percentages unless otherwise noted) 28

IV.5 Average Characteristics of Students Certified for NSLP Benefits Based on Categorical Eligibility by Application in Connecticut, by Whether Matched to State SNAP Data (percentages unless otherwise noted) 30

IV.6 Average Characteristics of Students Certified for NSLP Benefits Based on Categorical Eligibility by Application in Indiana, by Whether Matched to State SNAP Data (percentages unless otherwise noted) 32

IV.7 Average Characteristics of Students Certified for NSLP Benefits Based on Categorical Eligibility by Application in Nebraska, by Whether Matched to State SNAP Data (percentages unless otherwise noted) 35

IV.8 Average Characteristics of Students Certified for NSLP Benefits Based on Categorical Eligibility by Application in Texas, by Whether Matched to State SNAP Data (percentages unless otherwise noted) 37

IV.9 Average Characteristics of Students Certified for NSLP Benefits Based on Categorical Eligibility by Application in a Rural, Medium-Sized Texas District, by Whether Matched to State SNAP Data (percentages unless otherwise noted) 40

FIGURES

I.1 Percentage of Students Approved for NSLP Benefits Based on Categorical Eligibility Through Application, for Selected Districts 9

IV.1 Analysis Matching Rate, by State 23

IV.2 Alabama Matching Results 24

IV.3 Arizona Matching Results 27

IV.4 Connecticut Matching Results 29

IV.5 Indiana Matching Results 31

IV.6 Nebraska Matching Results 34

IV.7 Texas Matching Results 36

IV.8 West Virginia Matching Results 39

I. INTRODUCTION

To increase access to nutritious meals and reduce burden on school districts, the federal government allows students to be certified to receive free school meals without application based on participation in programs that confer categorical eligibility.[1] Directly certifying students who are categorically eligible for free meals involves matching lists of enrolled students to lists of program participants. However, the specific procedures used vary widely across States. In most cases, States use a central process for direct certification matching, in which a State agency is responsible for developing and maintaining the system that conducts direct certification matching. Other States use local matching systems, in which school districts have that responsibility. All States use computer data-matching techniques to perform direct certification, but the timing and frequency of the matching, as well as the methods used to transmit data, all vary. In all cases, however, effective direct certification relies on accurate, complete, and timely data.

It would be very difficult for any matching system to identify every student categorically eligible for free school meals. Data quality problems, such as incomplete data or misspelled or inconsistent names, can hinder effective direct certification and leave children categorically eligible for benefits uncertified. At best, these children would be required to complete an application to receive benefits, creating unnecessary burden on their families and their schools' and districts' administrative staff. At worst, some eligible children might go without National School Lunch Program (NSLP) benefits, increasing financial strain for families and possibly leading to diminished nutrition for their children. In addition, failing to certify all eligible children could increase debt for district nutrition programs if students eat school lunches but cannot afford to pay for them; it can also hinder schools' ability to qualify for Community Eligibility Provision (CEP) status.

The purpose of this report is to gain a better understanding of the categorically eligible children who are not matched in the direct certification process and to identify potential matching process improvements that might capture more of them. The analysis described in the report has two components. First, we present a descriptive analysis of the characteristics of children with Supplemental Nutrition Assistance Program (SNAP) records who are not matched to enrollment data. This provides insight into the types of students who might be more difficult to match for direct certification, such as those with longer or less common names and those for whom complete data are not available. Second, we present the results of an independent match of sampled categorically approved NSLP applications. These results are highly relevant to the efficacy of direct certification processes because students certified by application based on categorical eligibility represent a population that could have been directly certified but was not. Therefore, the results provide insight into ways in which current matching methods could be strengthened.

A. Overview of Approach

To understand more about eligible children not matched in direct certification processes, we analyzed SNAP participation data for selected States that were able to provide an indicator for

[1] For more details on the history and implementation of direct certification, please refer to the Direct Certification Study's main report: Moore, Quinn, Andrew Gothro, Kevin Conway, and Brandon Kyler. "National School Lunch Program Direct Certification Improvement Study: Main Report." Alexandria, VA: U.S. Department of Agriculture, Food and Nutrition Service, Office of Policy Support, 2014.

whether participants were matched to student enrollment lists in the direct certification process. We compared SNAP participant children who were and were not matched in terms of their age, first and last name characteristics, missing data patterns, and local area school and economic characteristics. These comparisons allow for an assessment of the characteristics associated with greater or lower probability of successful direct certification matching.

To assess ways in which direct certification procedures could be improved, we examined applications for school meal benefits from categorically eligible students in selected districts within States participating in the study (described below). These applications provide an efficient way to identify students who were not directly certified for free school meals despite their categorical eligibility. We assess the efficacy of a two-stage approach to matching data from the school meal applications to State SNAP participation data. In the first stage of this analysis, we used a deterministic matching method, requiring exact matches for multiple data elements. This method is similar to the direct certification matching approach used in 41 States and districts in school year (SY) 2012–2013, nationally and for the States studied in this report.[2] In the second stage, we conducted a probabilistic match between the application data and SNAP participation data for cases that were not matched deterministically. This approach, which was implemented using off-the-shelf matching software, allowed inexact or near matches for included data elements and generated a score indicating the likelihood of a legitimate match. For each State in the study, we assess the extent to which we can match students certified categorically by application using this method. We describe the differences in the number of matches identified with deterministic and probabilistic matching. Furthermore, we compare the characteristics of students who were and were not matched. This analysis highlights the extent to which students who are interested in receiving school meal benefits and eligible for free school meals categorically can be identified in State SNAP records. It also provides insight into the potential usefulness of probabilistic matching in direct certification.

B. Study Sample

This report's analysis focuses on seven States: Alabama, Arizona, Connecticut, Indiana, Nebraska, Texas, and West Virginia. The States vary geographically and in student population size. All States except Connecticut used central matching systems in SY 2012–2013.[3]

Within each participating State, the study team randomly sampled four school districts for inclusion in the categorically eligible application matching analysis. The study's sampling and weighting strategy was designed to yield results that will be representative of each participating State. However, in two States (Alabama and Indiana), half of the sampled districts could not provide data suitable for the study. Therefore, results for these two States might not be representative of the entire State.

---

[2] Moore, Quinn, Andrew Gothro, Kevin Conway, and Brandon Kyler. "National School Lunch Program Direct Certification Improvement Study: Main Report." Alexandria, VA: U.S. Department of Agriculture, Food and Nutrition Service, Office of Policy Support, 2014.

[3] The States in this study also served as in-depth case study States in a related report. For more detail on direct certification in these States, refer to the Direct Certification Improvement Study's main report.

C. Data Collection Procedures

The study's analysis required collection of State SNAP participant lists, district information on applications certified for free school meals based on categorical eligibility, and descriptive information drawn from publicly available sources.

1. State and District Data

Participating States and districts provided data files the study team used to conduct the matching analysis. Each State provided the study team with the statewide lists of school-age SNAP participants used for the initial direct certification match in SY 2012–2013. The data files contain many of the variables used in the State direct certification matching algorithms. Two of the States—Arizona and West Virginia—provided a matching flag in the data indicating whether a child was matched in the States' initial direct certification match for SY 2012–2013. Due to data limitations, other States in this study were unable to provide information on which school-age SNAP participants were matched in the direct certification process. Therefore, these States could not be included in the comparison of records that were and were not matched in the direct certification process.

Participating school districts provided data on applications for NSLP benefits from categorically eligible students (generally, members of households receiving SNAP or Temporary Assistance for Needy Families [TANF] benefits) from SY 2012–2013.[4] To reduce computer processing time, we limited the number of applicants to 300 per district. For districts submitting data on more than 300 applicants, we randomly sampled 300 for inclusion in the study. In these cases, we weighted the results to account for the random sampling when aggregating the results to the State level. Table I.1 presents characteristics of the SNAP and application data used in the analysis.

2. Data from Publicly Available Sources

To characterize SNAP participant children and NSLP applicants who were or were not matched, we obtained data from external sources on characteristics that might be associated with successful matching.

a. Name Commonality

Name commonality might be associated with direct certification matching success. It is possible that having very common names can lead to less successful matching due to the likelihood of duplicate matches. Conversely, uncommon names could be more likely to generate spelling errors, impeding successful matching. We obtained data on first name commonality using Social Security Administration lists of all first names given to at least five children in a single year in the United

[4] Some districts in West Virginia include CEP schools. Under this policy, schools would not collect applications for NSLP benefits. None of the districts sampled for inclusion in this study consisted exclusively of CEP schools. One district included some CEP schools. In that district, applications were drawn only from the non-CEP schools.

Table I.1. Characteristics of State SNAP Participation Data and Sampled District Data on Students Approved for NSLP Benefits Based on Categorical Eligibility by Application

	SNAP Data File		District Application Files		
	Participants	Variables	Number of Districts	Total Categorically Eligible Applicants	Variables
Alabama	326,855	First name	2	110	First name
		Middle name			Middle initial
		Last name			Last name
		SSN			SSN
		Date of birth			
		Address			Address
		City			City
		State			State
		Zip code			Zip code
		HH first name			Parent first name
		HH middle name			Parent middle initial
		HH last name			Parent last name
Arizona	626,186	First name	4	832[a]	First name
		Middle initial			Middle initial
		Last name			Last name
		Date of birth			Date of birth
		Gender			Gender
		SSN			
		Address			Address
		City			City
		State			State
		Zip code			Zip code
		SNAP case number			SNAP case number
		Parent first name			Parent first name
		Parent middle initial			Parent middle initial
		Parent last name			Parent last name
		Parent SSN			
Connecticut	143,677	First name	4	232	First name
		Middle initial			Middle initial
		Last name			Last name
		Date of birth			Date of birth
		Address			Address
		City			City
		State			State
		Zip code			Zip code
		SNAP case number			SNAP case number
		Parent first name			
		Parent middle initial			
		Parent last name			
Indiana	3,839,878	First name	2	512[a]	First name
		Middle initial			
		Last name			Last name
		Date of birth			Date of birth
		SSN			
		Address			Address
		City			City
		State			State
		Zip code			Zip code

	SNAP Data File		District Application Files		
	Participants	Variables	Number of Districts	Total Categorically Eligible Applicants	Variables
Nebraska	160,888	First name Last name Date of birth Gender Address City State Zip code SNAP case number HH first name HH last name	4	366[a]	First name Last name Date of birth Gender Address City State Zip code SNAP case number Parent first name Parent last name
Texas	1,452,913	First name Middle name Last name Date of birth Gender Ethnicity Grade School name District name SSN Address City State Zip code SNAP case number	4	893[a]	First name Middle name Last name Date of birth Gender Ethnicity Grade District name Address City State Zip code SNAP case number
West Virginia	206,413	First name Middle initial Last name Date of birth Gender SSN Address City State Zip code SNAP case number Parent first name Parent middle initial Parent last name Parent SSN	4	78	First name Middle initial Last name Date of birth Address City State Zip code SNAP case number Parent first name Parent middle initial Parent last name Parent SSN

Sources: Records of Matched and Unmatched SNAP Participants from the Alabama Department of Human Resources, the Arizona Department of Economic Security, the Connecticut Department of Social Services, the Indiana Family and Social Services Administration, the Nebraska Department of Health and Human Services, the Texas Health and Human Services Commission, and the West Virginia Department of Health and Human Services. NSLP application data from sampled school districts.

Notes: Variables listed under District Files include only variables that are also included in the State SNAP data file, that is, variables that would be useful for the independent match. Moreover, these variables are those that were provided by any district; some listed variables are not available for all districts.

[a]Some districts in these states provided data on more than 300 applicants. In these cases we randomly selected data from 300 applicants to include in the study. The observation counts reflect the sample used in the analysis.

HH = household; NSLP = National School Lunch Program; SNAP = Supplemental Nutrition Assistance Program; SSN = Social Security number.

States. To match our sample of children who were of school age in 2012, we used lists of children born in each year from 1994 to 2007.[5] We obtained data on the commonality of last names using 2000 Decennial Census data.[6] The list contains all last names that appeared at least 100 times in that year's census. We used these lists to calculate national-level commonality percentiles for both first and last names and applied them to the children in our analysis tables. For example, a child in the 80th first name commonality percentile has a more common first name than 80 percent of people nationally born from 1994 to 2007.

b. Private School Statistics

Although private schools that participate in the NSLP are expected to participate in direct certification, they often are less integrated than public schools in statewide data systems typically used in the matching process. Therefore, the presence of large numbers of private schools or a high percentage of private school students might be associated with less successful matching. We obtained county-level data on private schools and private school students from the SY 2009–2010 Private School Universe Survey, the most recent data available.[7] We obtained data on the number of public school students in each county using Common Core of Data survey data from the same school year. We used these statistics to calculate the percentage of total students in each county who attended private school that year.

c. Economic and Geographic Indicators

County- or zip code-level data on economic and geographic characteristics can be associated with successful matching. We obtained county-level unemployment rate data from the Bureau of Labor Statistics[8] and county-level poverty rate statistics from the Census Bureau's Small Area Income and Poverty Estimates.[9] We also obtained zip code-level measures of urban and rural classifications from the Census Bureau.[10]

D. Methods for Independent Matching and Analysis

Categorically eligible students who were certified for free meals by application should appear on the State SNAP participation lists, assuming the application contains accurate information and the SNAP participation list is complete. As noted earlier, we compared applications approved based on categorical eligibility to State SNAP participation records using deterministic algorithms to identify

[5] Social Security Administration. "Beyond the Top 1000 Names." Available at http://www.ssa.gov/oact/babynames/limits.html.

[6] U.S. Census Bureau. "Genealogy Data: Frequently Occurring Surnames from Census 2000." Available at http://www.census.gov/genealogy/www/data/2000surnames/index.html.

[7] Institute of Education Sciences, National Center for Education Statisitics. "Private School Universe Survey." Available at https://nces.ed.gov/surveys/pss/pssdata.asp.

[8] This analysis used unemployment rate data from August 2013, the most recent data available at the time: http://data.bls.gov/map/MapToolServlet?survey=la&map=county&seasonal=u.

[9] This analysis used poverty rate data from 2012, the most recent data available at the time: http://www.census.gov/did/www/saipe/data/index.html.

[10] Urban and rural classifications are based on Census Bureau Zip Code Tabulation Areas, which overlap substantially with zip codes: http://www.census.gov/geo/maps-data/data/ua_rel_download.html.

obvious exact matches. We then compared the same files using probabilistic matching methods to identify legitimate matches that require a more flexible matching strategy.

We used consistent matching methods across the seven study States to the extent possible, given the data elements available. In most cases, we required matches on four data elements for a match. Algorithms drew from the following data elements: first name, last name, date of birth, address, Social Security number (SSN), parent name, or SNAP or TANF case number. Specific matching algorithms varied by State according to which data fields were available.

1. Deterministic Match

In the deterministic match, we required four exact matches among the data elements listed previously (or three exact matches if SSN was one of the three matching elements). In most States, we required these matches to include first name, last name, and date of birth (the exception was Alabama, which did not provide date of birth but did include SSN—see page 24 for the specific algorithm used for that State). In the deterministic matching step, we conducted manual review to ensure the matching algorithm worked properly, but did not alter the matching results. Variations in spelling, truncated values, and other close but inexact matches precluded deterministic matches. However, we did not impose a penalty for conflicting values in one or more data fields, provided there were exact matches in at least four fields or an exact match on SSN. For example, a pair of observations with conflicting addresses would still be a deterministic match if they matched exactly on first name, last name, date of birth, and SNAP case number.

2. Probabilistic Match

In the second stage of the matching analysis, we compared data fields available in both the application data and in the State SNAP data in each State. In most cases, we required four data elements to match to identify probabilistic matches. Unlike in the deterministic process, however, we allowed inexact and exact matches. LinkageWiz (described in the next paragraph) compared the data sets and compiled a list of the most likely pairs. We manually reviewed pairs that did not appear in the deterministic results and accepted as matches those that had exact or inexact matches on at least four data fields. The exception to this process was Alabama, the only State in the study to include SSNs in both data sets. Because of the unique reliability of SSNs, if a student matched exactly on this field, we required matches on only two additional fields, rather than three.

a. Probabilistic Matching Software

The study team conducted probabilistic matching using an off-the-shelf software tool, LinkageWiz. It is one of several such matching software tools available for purchase that States and districts could use to conduct direct certification. These programs compare data sets using such fields as name, date of birth, or address. Users can easily include additional data fields according to their needs and data availability. This software calculates a score indicating the likelihood of a match given the information available.

The calculated score accounts for incomplete information and data fields that are close matches because of misspellings, inverted dates, and other data errors. The score is based on bonuses applied for fields that match (or nearly match) and penalties applied for fields that do not match. Near matches receive smaller bonuses than exact matches. The relative size of the bonuses is proportional to the ability of a data field to uniquely identify matches, whereas the size of the penalties is inversely proportional to the likelihood that the variable might differ even for legitimate matches. Based on these scores, the software identifies the record from the comparison file that most closely matches each record from the original source file.

In typical use of this software, scores above an upper threshold are designated as matches. Scores between a lower and an upper threshold are designated as matches (or nonmatches) based on a case-by-case manual review. Scores below a lower threshold are designated as nonmatches. The upper and lower thresholds are determined based on a preliminary manual review of all potential matches sorted by match score and are selected to ensure a detailed manual review of any questionable cases. Thus, the upper threshold is selected to be sufficiently high to have a very high degree of confidence that all scores above the threshold are true matches. Similarly, the lower threshold is selected to be sufficiently low to have a high degree of confidence that all scores below the threshold are not matches.

In conducting the probabilistic matching analysis, the study team did not make any advanced modifications to the software. The software requires familiarity with computers but does not require programming or other specialized skills.

b. Process for Probabilistic Matching

The probabilistic matching software generated a list of the best match available from the State SNAP file for each sampled application, along with the matching confidence score associated with each match. We conducted a manual review of these results by viewing the output in Excel and sorting the results by the confidence score.[11] Ignoring all pairs that had already been matched in the deterministic process, we reviewed each potential match to see if it matched on four elements, allowing inexact matches on any data element.

Identifying inexact matches manually requires reviewer discretion. However, we applied consistent standards across observations and States. Most inexact matches resulted from obvious spelling variations, such as *Oak Street* versus *Oak St*. We also accepted spelling variations such as *Stephen* versus *Steven*. Similarly, many names contained suffixes in one data source but not the other. Obvious name variations, such as *Jon* and *Jonathon*, frequently led to inexact matches. Variable truncation caused many inexact matches, particularly in long last names. For compound last names, we accepted as a match any comparison in which one source contained only one portion of a compound name (for example, *Smith-Jones* and *Jones*). For dates of birth, we accepted as an inexact match any comparison in which two of the three components matched. For example, we accepted 2/05/2003 as an inexact match for 2/16/2003.[12]

---

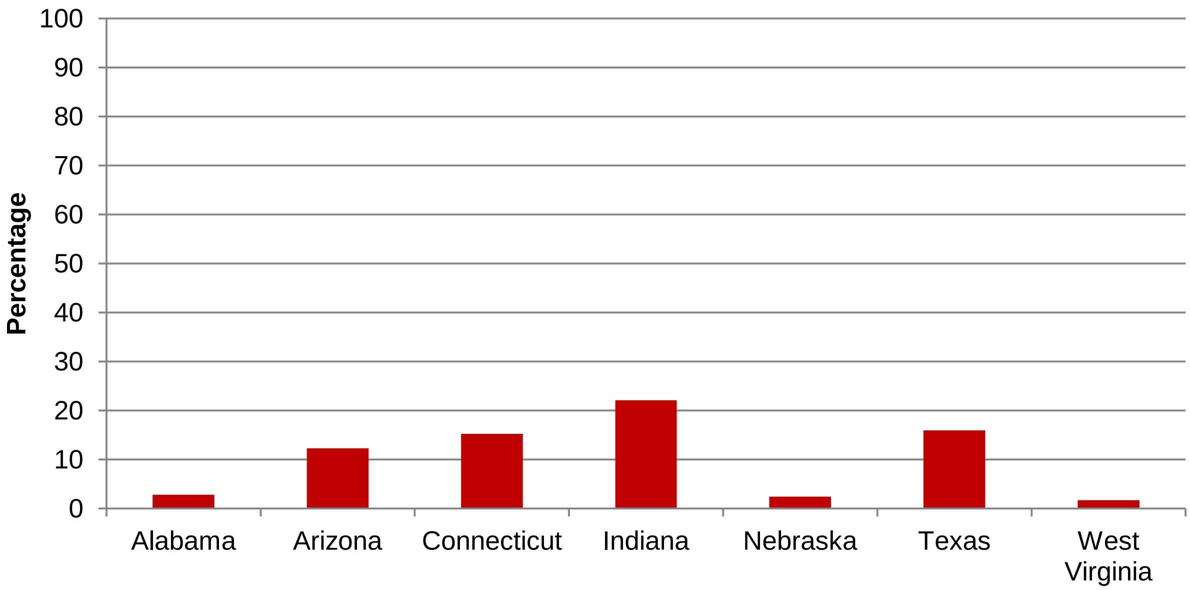
[11] When conducting probabilistic matching, users are not required to review the entire set of LinkageWiz results manually. Users can set upper and lower confidence score thresholds to delineate matches and nonmatches, and manually review only the pairs between these thresholds. However, because the matching results varied so widely from State to State and because we conducted only a single probabilistic matching iteration in each State, we manually reviewed the entire set of results in all States. This was feasible because of the relatively small sample sizes in our analysis compared with full State-level direct certification processes.

[12] This could be a more lenient approach to matching dates of birth than staff would use when manually reviewing actual direct certification results. In reality, not all children born in February 2003 should be considered near matches on that data element. Manual reviewers consider each data element in context of all other data available. A hypothetical Joe Stevens born 2/05/2003 would not be considered to have a near-matching date of birth as a hypothetical Jane Smith born 2/16/2003 living in a different city. However, Joe Stevens born 2/05/2003 living at the same address as Joseph Stevens born 2/16/2003 should likely be considered a match. This leniency was required to identify such likely matches. Because the purpose of our analysis was to describe the results yielded by specific matching approaches with the same data—and because few data elements were available in our analysis for many States—we applied this lenient approach to dates of birth to all potential matches. The more stringent requirements in other data fields mitigated the risks of false positives.

3. Implications of Using Application Data for Matching

The categorically eligible students used in our data matching processes represent a small portion of the total students certified for NSLP benefits based on categorical eligibility. Figure I.1 depicts the proportion of all certified categorically eligible students who were certified based on application. These proportions apply to the districts participating in this study, aggregated by State. In districts across all study States, most categorically eligible students are identified through direct certification. However, wide variation exists on the proportion of certified categorically eligible students identified through direct certification. Districts in Nebraska and West Virginia, which both use probabilistic matching for direct certification, directly certify the most students, with only 2 percent certified through applications. At the other extreme, more than 20 percent of categorically eligible students in the participating districts in Indiana are certified by application. Our matching analysis draws from these pools of students in the participating districts.

Figure I.1. Percentage of Students Approved for NSLP Benefits Based on Categorical Eligibility Through Application, for Selected Districts



Source: FNS Verification and Summary report data.

Although school meal benefit applications provide a convenient source of categorically eligible students not matched in the direct certification process, they also present some data challenges. The primary drawback to these data is the limited range of data elements they contain. School enrollment data used in actual direct certification matching processes often contain many more data fields that can be used to identify matches. In particular, the application data for six of the seven States did not contain SSNs, a particularly effective matching element. Due to this limitation, the matching results presented here might understate the matching rates that would be available with richer student identification information.

The second challenge in using school meal benefit application data for matching is that they can be of lower quality than school enrollment data. Many applications used in this analysis were incomplete, lacking data for such fields as date of birth or parent name. Districts might not conduct the same quality assurance processes on application data that they use for school enrollment data. Because applicants often complete school meal benefit applications on paper, rather than electronically, application illegibility might also have led to errors. Such errors could have occurred when school officials entered data from applications into their data systems. Likewise, some of the districts selected for participation in this study submitted scanned portable document format (PDF) files of applications; illegibility might have caused data entry errors when we processed the data.

Finally, some categorically eligible students certified by application eventually might have been matched as a part of a central or local direct certification process sometime after submitting an application. In these cases, applications certified based on categorical eligibility do not represent students who were not identified by the direct certification process; rather, they represent students who should have been identified earlier. This situation might be unlikely in States that conduct frequent matches with updated student enrollment and program participation data because categorically eligible students would be directly certified soon after beginning participation in a program that confers categorical eligibility.

Some districts reclassify students as being directly certified when they are originally certified by application and later directly certified. This process is automatic for definite matches in West Virginia's direct certification process. We do not have information on which districts in other States use this strategy. For these districts, however, remaining applications certified based on categorical eligibility do represent students who should have been identified by the direct certification process.

E. Organization of Rest of Report

The rest of this report describes the direct certification process in each State in the study and provides the results of the study's analysis. Chapter II contains information on the direct certification procedures used in SY 2012–2013 in the seven States selected for this study. We describe the data sources and matching methods used in each State and discuss the primary challenges they faced. In Chapter III, we present descriptive analysis comparing school-age SNAP participants who were matched in the direct certification process to those who were not. In Chapter IV, we describe the results of our independent matching analysis using NSLP benefit application data. In Chapter V, we synthesize our findings, identifying apparent strengths and limitations of our analysis methods and possible improvements to direct certification suggested by the results.

II. APPROACH TO DATA MATCHING IN SELECTED STATES

State direct certification procedures determine which categorically eligible students are certified without an application; students not certified in this process must submit applications in order to receive benefits. Thus, these procedures directly inform the analysis comparing the characteristics of matched and unmatched students. They also determine the sample for the analysis matching students certified by application to State SNAP records. Therefore, understanding the direct certification procedures of the States in this study provides context for interpreting the results of the analysis.

The seven States in this study represent a range of direct certification approaches. States used different technology and data sources in the matching process. The administrative structures varied, as did the role of districts in the process. Different strategies and varying State contexts led to different challenges in completing direct certification. In this chapter, we describe the process used in each State, including the data sources, matching process, and matching algorithms in place. We also discuss common challenges States face in completing data matching. Information from this chapter is based on responses to the National Survey of Direct Certification Practices and case study visits conducted in each of the States. For more information on the data collection procedures and a more detailed description of State procedures, please refer to the Direct Certification Improvement Study's main report.[1]

A. Overview of Current Data Matching Practices and Procedures, by State

The matching procedures in place in SY 2012–2013 varied across the States in this study. One State—Connecticut—used a local matching system, and the specific State and district roles differed greatly across States using central matching systems. The frequency of student enrollment data updates varied greatly, from only once annually in Texas, to real-time updates in West Virginia's statewide enrollment data system. Similarly, States conducted the matching with varying frequency, from daily to the required minimum frequency of three times per year.[2] Finally, States in the study used a range of data matching algorithms, incorporating different data elements and using different criteria to identify matches. Two States—Nebraska and West Virginia—incorporated probabilistic matching into their direct certification process. The rest of this section summarizes the direct certification procedures for each State in this analysis, including data sources available for direct certification, the matching process, and the matching algorithm.

We summarize the direct certification procedures used by the seven study States in Table II.1. For more detailed information on these procedures, see Appendix A.

[1] Moore, Quinn, Andrew Gothro, Kevin Conway, and Brandon Kyler. "National School Lunch Program Direct Certification Improvement Study: Main Report." Alexandria, VA: U.S. Department of Agriculture, Food and Nutrition Service, Office of Policy Support, 2014.

[2] Because Texas updated its enrollment data only annually and matched monthly, it used the same enrollment data for each match conducted during a 12-month period (March of one year to February of the next). The matching yielded different results each time because Texas used updated program participation data each month.

Table II.1. Characteristics of the Direct Certification Matching Processes in Select States, SY 2012–2013

State	Type of Matching System	How Does Direct Certification Work?	Approach for Unmatched Students?	Frequency of Direct Certification
Alabama	Central	The matching process produced a list of directly certified students that districts retrieve from a secure website. Districts could directly certify students in their local systems by comparing their enrollment files with the State's matched list. Alternatively, districts could compare their local files against the statewide program data files.	Optional at the district level. The State did not investigate unmatched records.	Monthly
Arizona	Central	Arizona conducted direct certification matching on a central State server. Districts triggered the matching process and could upload updated enrollment data or match against data already on the server. The State system produced matched and unmatched lists for districts to view or download.	There was no process for reviewing unmatched records.	At least three times per year
Connecticut	Local	Districts received program data three times per year. They compared those data against their local enrollment files to identify directly certified students, using any algorithm they wished.	Varied by district.	District discretion
Indiana	Central	Districts initiated the matching process. For the initial match of the school year, districts uploaded a current school enrollment file. During the school year, student enrollment data on the State server are updated in real time. The State system matched enrollment data against the current program participation data and produced lists of matched, partially matched, and unmatched students.	Districts could attempt to match unmatched records using State-generated lists.	At least three times per year
Nebraska	Central	The State conducted central matching using a probabilistic algorithm. The matching was automated and conducted nightly. The system produced lists of matched and partially matched students. Districts downloaded the results from the State server as often as they wished.	There was no process for reviewing unmatched records.	Daily
Texas	Central	State staff matched the State enrollment file with SNAP and TANF program data. Each district received a list containing only the students that appeared to attend schools in that district. District staff then matched the State list with their local enrollment files in their point-of-sale (POS) systems.	Beginning in SY 2013–2014, districts will be able to attempt to match these students.	Monthly
West Virginia	Central	West Virginia conducted matching daily using a probabilistic algorithm. The State Department of Education matched program data against the statewide school enrollment data and made matched, unmatched, and partially matched lists available to each district. Districts viewed matched, unmatched, and partially matched results through their local POS systems.	Districts attempted to match unmatched students.	Daily

Source: Direct certification case study interviews.

B. Common Challenges to Data Matching

States in this study reported some challenges in their direct certification matching processes. These ranged from technological problems, such as bandwidth constraints and system performance limitations, to data problems, such as the timeliness and accuracy of school enrollment and program participation data.

1. Technological Challenges

System limitations and other technological challenges inhibited some aspects of direct certification. Staff in Alabama expressed a desire to match more frequently than monthly, but reported that system and financial resource constraints prevented them from doing so. Bandwidth limitations slowed the direct certification process in West Virginia. Five States (all except Indiana and West Virginia) reported that their State systems had insufficient information to automate the process of extending categorical eligibility to all children in households with directly certified students. State staff in Indiana said that direct certification performance varied across the State according to district system type. Some districts' systems automatically integrated State direct certification data; others used simpler point-of-sale (POS) systems and processed State data manually. In addition, some districts did not effectively use all the technology tools the State made available.

2. Data Challenges

Most data challenges in direct certification consisted of issues with the accuracy or timeliness of school enrollment or program participation data. Staff in several States also reported concerns about data security, data handling procedures, or communication with data partners.

States reported timeliness problems with both school enrollment and program participation data. Timely submission of data is an important component to effective direct certification. If either data source is out of date, fewer matches can be identified. In States dividing matching results by county or school district, results could be sent to the wrong place for students who recently moved. In Arizona and Indiana, districts triggered the matching process by uploading data elements for their current rosters of students. Districts in Arizona sometimes did not keep their local enrollment records up to date, resulting in fewer successful matches against State program data. Districts in Indiana often did not initiate matches more frequently than three times per year, as the State would prefer.

A particular concern for several States was including newly enrolled students in the initial match of a school year, which usually occurred before the start of school. Staff in Indiana, Nebraska, and Texas reported that newly enrolled students did not appear in their statewide school enrollment data normally used for direct certification. In Indiana and Nebraska, districts could upload current school enrollment files to be used for the States' initial match to mitigate this problem, although staff in Nebraska reported that not all districts completed this step. In Texas, enrollment data timeliness extended beyond newly enrolled students. State staff updated student enrollment data only once per year and did not release them for use in direct certification until March each year, after a long review process. Therefore, direct certification matching conducted before March was based on enrollment data from the previous school year, and matching conducted after March was based on data that might be out of date by the time they are used for matching.

Several States also cited data accuracy as a direct certification challenge. State staff in Alabama reported that data entry error in the school enrollment data affected direct certification matching. Indiana staff reported that enrollment data for charter and parochial schools were more likely to contain errors than other school enrollment data.

In other data challenges, staff in Texas reported that they would prefer not to use SSNs as a matching element, due to concerns about data sensitivity, but no other unique identifier existed in both enrollment and program data. Districts in Connecticut did not have a clear State point of contact for program participation data and had difficulty requesting changes that would improve their local matching processes.

III. DESCRIPTIVE ANALYSIS OF SNAP RECORDS

Comparing the characteristics of school-age SNAP participants who were matched in the direct certification process to the characteristics of participants who were not directly certified provides insight into the types of categorically eligible students who are less likely to be directly certified and the student characteristics that can make direct certification more challenging. Two States—Arizona and West Virginia—provided SNAP participant data that contained variables identifying directly certified children; similar data from the other study States were not available. For these two States, we compared SNAP participants who were and were not directly certified in terms of age, name characteristics, local private school concentration, and local economic conditions. In Arizona we also examined the frequency of direct certification by gender. Tables III.1 and III.2 and the text in this chapter present findings from this analysis.

Table III.1. Average Characteristics of School-Age Children with SNAP Records in Arizona, by Whether Matched to School Enrollment Data (percentage unless otherwise noted)

Characteristic	Directly Certified	Not Directly Certified
Student Characteristics		
Age[a]	●●●	
5	9.3	9.0
6–17	89.4	81.7
18	1.3	9.3
Mean (years)	11.0***	11.8
Female	49.5	49.8
First Name Commonality[b]		
Average percentile	41.0***	39.3
Average name length (number of letters)	6.17***	6.20
Last Name Commonality[c]		
Average percentile	52.0***	43.8
Average name length (number of letters)	6.53***	7.44
Local Characteristics		
Percentage of Students in County Attending Private School	4.5***	4.6
County Unemployment Rate	9.5***	9.1
County Poverty Rate	19.6***	19.2
Urbanicity[a]	●●●	
Urbanized area	76.1	79.3
Urban cluster	15.4	12.8
Rural	8.0	7.5
Missing	0.5	0.4
Sample Size (SNAP records)	**240,132**	**247,409**

Source: Arizona Department of Economic Security, Records of Matched and Unmatched SNAP Participants.

[a] Differences between group distributions were tested using a chi-squared test.

[b] Based on Social Security Administration records from 1994 to 2007.

[c] Based on 2000 Decennial Census data.

*/**/*** Mean is significantly different from mean of unmatched records at the .10/.05/.01 levels, respectively.

●/●●/●●● Distribution is significantly different from distribution of unmatched records at the .10/.05/.01 levels, respectively.

Table III.2. Average Characteristics of School-Age Children with SNAP Records in West Virginia, by Whether Matched to School Enrollment Data (percentage unless otherwise noted)

Characteristic	Directly Certified	Not Directly Certified
Student Characteristics		
Age[a]	●●●	
5	8.3	24.0
6–17	87.0	67.7
18	4.6	8.3
Mean (years)	11.4***	10.6
First Name Commonality[b]		
Average percentile	48.9***	46.4
Average name length (number of letters)	6.13***	6.21
Last Name Commonality[c]		
Average percentile	53.0***	50.5
Average name length (number of letters)	6.28***	6.43
Local Characteristics		
Percentage of Students Attending Private School	4.1***	4.7
County Unemployment Rate	6.5***	6.4
County Poverty Rate	19.3	19.3
Urbanicity[a]	●●●	
Urbanized area	34.2	37.8
Urban cluster	16.2	15.2
Rural	47.3	44.6
Missing	2.2	2.4
Sample Size (SNAP records)	**165,974**	**9,102**

Source: West Virginia Department of Health and Human Services, Records of Matched and Unmatched SNAP Participants.

[a] Differences between group distributions were tested using a chi-squared test.

[b] Based on Social Security Administration records from 1994 to 2007.

[c] Based on 2000 Decennial Census data.

*/**/*** Mean is significantly different from mean of unmatched records at the .10/.05/.01 levels, respectively.

●/●●/●●● Distribution is significantly different from distribution of unmatched records at the .10/.05/.01 levels, respectively.

- **Patterns in students' age distribution varied for SNAP participants by direct certification status, but were not consistent across the two States.**

We examined differences by direct certification status in whether SNAP participants were ages 6 to 17—clearly school-age—and whether they were ages 5 or 18—ages more likely to include children who either have not started or already left school. In both Arizona and West Virginia, the age distribution of directly certified school-age SNAP participants differed significantly from that of other school-age SNAP participants: directly certified SNAP participants were more likely to be ages 6 to 17 years old than those not certified. However, patterns for children who were ages 5 or 18 differed between the States. In Arizona, the difference occurred almost completely in 18-year-olds. Only 1.3 percent of directly certified children were 18 years old compared to 9.3 percent of children not directly certified (Table III.1). By contrast, the differences in West Virginia's age distribution occurred mostly in 5-year-olds. Only 8.3 percent of directly certified children were 5 years old, compared to 24 percent of children not directly certified (Table III.2). These conflicting patterns could be due in part to differences in the high school dropout rates between the two States. Arizona's drop-out rate was nearly twice as high as West Virginia's in SY 2009–2010, the most recent data available (7.8 versus 4.0 percent).[1] Children who have dropped out of school would not appear on the school enrollment lists and thus would not be directly certified. In West Virginia's case, the 5-year-olds may not have been enrolled in school yet, or, if they were, the enrollment data may not have included them at the time of the initial direct certification match.

- **There were no important differences in gender by direct certification status.**

About half of school-age SNAP participants in Arizona were female, both among participants who were directly certified and those who were not (Table III.1). Information on the gender of SNAP participants was not available for West Virginia.

- **SNAP participants who were not directly certified tended to have longer, less common names than students who were directly certified.**

In Arizona and West Virginia, both the first and last names of directly certified school-age SNAP participants were significantly more common than those of SNAP participants who were not directly certified. These differences were particularly large for last names. In Arizona, the average last name was at percentile 52.0 for directly certified SNAP participants and percentile 43.8 for other SNAP participants (Table III.1); in West Virginia, these percentiles were 53.0 and 50.5, respectively (Table III.2). Similarly, in both States, the first and last names of SNAP participants who were not directly certified were significantly longer than those of SNAP participants who were directly certified. These findings indicate that students with longer, less common names are more difficult to match in direct certification processes. Difficulty in matching longer, less common names could be related to misspellings and errors in recording such names. It is also possible that more common names are more likely to lead to false positive matches.

- **Directly certified school-age SNAP participants are less likely than other school-age SNAP participants to live in counties with higher private school enrollment.**

In both States, directly certified school-age SNAP participants lived in counties with significantly lower average private school enrollment rates than other SNAP participants (Tables III.1 and III.2). Categorically eligible students attending private school could be less likely to be matched because private schools are less likely than public schools to participate in the NSLP. Among those that do, some do not conduct direct certification or do so in a less integrated way than do public schools. For example, private schools might not be included in statewide student information systems, or they could be more likely to use manual, less effective processes to identify eligible students. This is true in Arizona and West Virginia: in both States, private schools do not participate in the statewide school enrollment systems and supply enrollment data to the State less frequently than public schools do.

---

[1] U.S. Department of Education, National Center for Education Statistics; available at http://nces.ed.gov/ccd/drpcompstatelvl.asp.

- **Both Arizona and West Virginia exhibit differences by direct certification status in the average local economic conditions and urbanicity of school-age SNAP participants.**

In both States, directly certified SNAP participants lived in counties with significantly higher unemployment rates, on average, than other SNAP participants (Tables III.1 and III.2). In Arizona, they also lived in counties with significantly higher poverty rates. In addition, directly certified SNAP participants in both States were less likely than other SNAP participants to live in urban areas.

It is not clear why these patterns of local characteristics emerged. One possibility is that these patterns are related to the relative presence of categorically eligible students in certain types of districts and district incentives to perform their direct certification roles. In both Arizona and West Virginia, districts play key roles in the direct certification process. In Arizona, districts are responsible for triggering direct certification matches; in West Virginia, districts are responsible for processing lists of potential and unmatched students. Therefore, it is possible that districts with relatively worse economic conditions or in more urban areas (where a greater percentage of students are likely to be categorically eligible for school meal benefits) are more diligent in fulfilling these responsibilities and thus have relatively more success in direct certification.

IV. INDEPENDENT MATCH OF SNAP RECORDS TO NSLP APPLICATIONS

To assess the extent to which categorically eligible students who were not directly certified can be matched to State program participation records, we conducted a two-stage matching analysis between school meal benefit application data and SNAP caseload data for each of the seven States in this study. We examine the different overall matching rates among States, as well as the different results achieved with deterministic versus probabilistic matching. We compare the characteristics of categorically eligible children by match results to assess the factors associated with successful data matching. In the rest of this chapter, we describe the application data sample, the independent matching algorithms for each State, the results for each State, and cross-State themes.

A. NSLP Application Data

The application data used in this study consisted of student information available for applications certified for school meal benefits based on categorical eligibility.

We collected these data from randomly sampled districts within each of the participating States. Table IV.1 contains brief descriptions of the sampled districts. We received application data from four districts in each State, except for Alabama and Indiana. Although four districts were selected in these two States, only two districts provided data.

Table IV.1. Descriptions of Sampled School Districts, by State

State	Sampled Districts	Sample Size
Alabama	Small rural district	22
	Medium rural district	88
Arizona	Small rural district	10
	Small urban district	277
	Large urban district 1	246
	Large urban district 2	300
Connecticut	Medium urban district 1	59
	Medium urban district 2	22
	Medium urban district 3	1
	Large urban district	150
Indiana	Small rural district	212
	Large urban district	300
Nebraska	Small rural district 1	10
	Small rural district 2	48
	Small urban district	8
	Large urban district	300
Texas	Small rural district	18
	Small urban district	300
	Medium rural district	280
	Large urban district	300
West Virginia	Medium rural district 1	19
	Medium rural district 2	8
	Large urban district 1	19
	Large urban district 2	32

Although the student data used in this analysis were drawn from applications, note that the unit of analysis is the student, not the application. We limited the number of categorically eligible students in the analysis to 300 per district, randomly selecting 300 for districts with more than that number. The sample sizes used in the matching ranged from 78 in West Virginia to 893 in Texas. All of these categorically eligible applicants were included in both rounds of the independent matching procedure. For presentation, however, we restricted this sample to applicants with at least four nonmissing matching elements. We applied this exclusion because applicants with fewer nonmissing matching elements cannot be identified as a match using our matching algorithms, and including these students in the presentation obscures patterns in the characteristics of categorically eligible applicants who were and were not independently matched. Sample sizes included in this chapter range from 39 categorically eligible applicants in West Virginia to 833 in Arizona.

As Table IV.2 shows, characteristics of applicants varied across States. The average age of applicants ranged from 9.4 years in Texas to 12.2 in Indiana. Name commonality varied as well. Among first names, applicants ranged from percentile 30.9 in Nebraska to percentile 43.8 in West Virginia. Last name commonality spanned from percentile 31.5 in Arizona to percentile 63.0 in West Virginia. Average last name length varied substantially, with the States with large Hispanic populations having the longest names: 7.7 letters in Texas and 9.1 in Arizona.

Average characteristics of the counties and local areas in which applicants resided also varied among sampled districts. The average county rate of private school enrollment ranged from less than 5 percent in Arizona to more than 13 percent in Nebraska. County unemployment rates spanned from 4 percent in Nebraska to more than 9 percent in Indiana. Applicants in Connecticut lived in counties with the lowest average poverty rate, at less than 13 percent. Applicants in Alabama lived in counties with the highest average rate, at more than 23 percent.

Missing data patterns varied greatly across the study States. However, this was partly the result of our restricting the samples presented in the tables to those observations with data in at least four matching fields or an SSN. For States with only four fields available, we therefore restricted the tables to observations with no missing data (because missing a single element would make matching impossible using our algorithms). Thus, the presence of missing data in this sample indicates both that a State had more than four data fields available and that its applications had incomplete data in them.

B. Matching Results

The number of categorically eligible applicants identifiable in State SNAP records through the independent match process varied widely by State. Including both deterministic and probabilistic matches, our match rates ranged from less than 8 percent of the sample in West Virginia to 81 percent in Nebraska (Figure IV.1).

Table IV.2. Average Characteristics of Students Certified for NSLP Benefits Based on Categorical Eligibility by Application, by State (percentages unless otherwise noted)

Characteristic	Alabama	Arizona	Connecticut	Indiana	Nebraska	Texas	West Virginia
Student Characteristics							
Age							
Younger than 5	NA	1.0	4.0	1.6	6.3	15.6	0.0
5–9	NA	33.9	37.0	34.1	44.1	46.3	35.9
10–14	NA	26.7	39.9	35.1	32.3	27.5	41.0
15–18	NA	5.0	17.9	28.2	16.2	10.3	20.5
19 and older	NA	0.0	1.2	1.0	0.5	0.3	2.6
Missing	NA	33.5	0.0	0.0	0.5	0.0	0.0
Mean (years)	NA	10.0	11.2	12.2	10.4	9.4	11.8
Gender							
Male	NA	NA	NA	NA	57.4	48.6	58.1
Female	NA	NA	NA	NA	42.6	51.4	41.9
First Name Commonality							
Average percentile	40.1	36.3	35.8	34.1	30.9	38.8	43.8
Average name length (number of letters)	6.23	6.36	6.44	6.22	6.02	6.18	6.33
Last Name Commonality							
Average percentile	51.0	31.5	46.3	57.8	39.3	46.7	63.0
Average name length (number of letters)	6.17	9.13	6.77	6.32	6.44	7.65	6.03
Missing Data on:							
First name	0.0	0.4	0.0[a]	0.0[a]	0.0	0.0[a]	0.0
Last name	0.0	0.4	0.0[a]	0.0[a]	0.0	0.0[a]	0.0
Date of birth	NA	33.5	0.0[a]	0.0[a]	0.3	0.0[a]	0.0
SSN	78.2	NA	NA	NA	NA	NA	NA
Address	3.0	1.2	0.0[a]	0.0[a]	0.0	0.0[a]	0.0
Parent name	0.0	65.5	NA	NA	0.0	NA	79.5
SNAP case number	NA	5.5	NA	NA	0.0	NA	100.0
Any element	81.2	100.0	0.0[a]	0.0[a]	0.3	0.0[a]	100.0
Multiple elements	0.0	5.8	0.0[a]	0.0[a]	0.0	0.0[a]	79.5

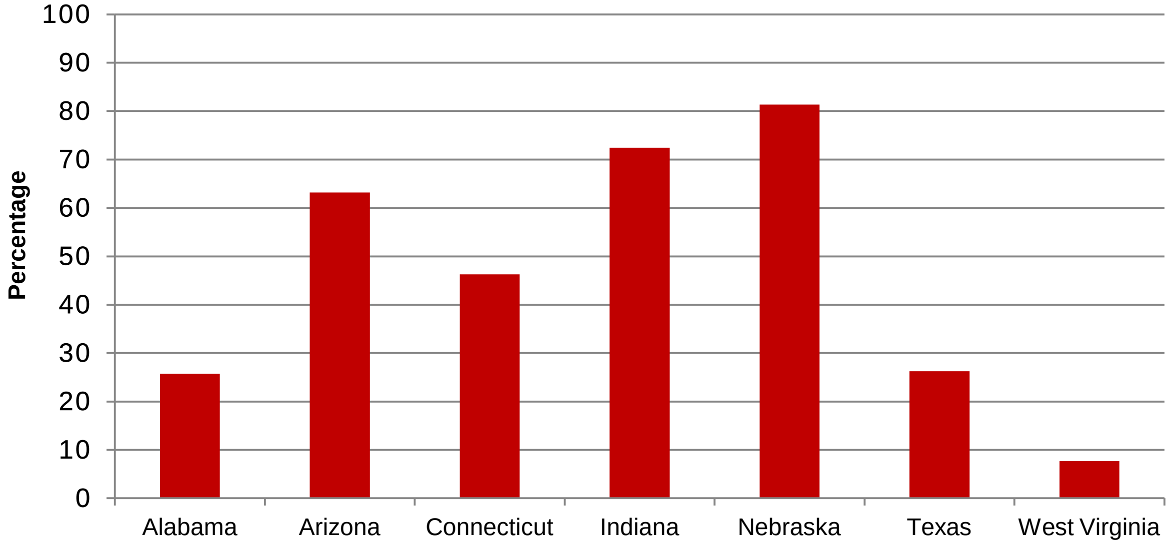
Characteristic	Alabama	Arizona	Connecticut	Indiana	Nebraska	Texas	West Virginia
Local Characteristics							
Students in County Attending Private School	5.2	4.7	9.2	6.7	13.5	5.1	5.0
County Unemployment Rate	7.7	7.9	8.7	9.3	4.1	5.7	6.1
County Poverty Rate	23.4	18.0	12.5	18.0	14.2	17.2	17.1
Urbanicity							
Urbanized area	0.0	64.2	NA	0.0	81.1	50.8	48.7
Urban cluster	0.0	29.7	NA	26.3	2.2	0.0	2.6
Rural	97.0	3.0	NA	15.3	2.7	49.2	46.2
Missing	3.0	3.1	NA	58.4	14.0	0.0	2.6
Sample Size	**101**	**833**	**173**	**510**	**365**	**590**	**39**

Source: NSLP application data from sampled school districts.

[a] Connecticut, Indiana, and Texas submitted data with only four variables suitable for matching across the State SNAP and district NSLP application files. Because we only display results for individuals with data for at least four data elements, we excluded individuals in these States with any missing data from the tables in this chapter.

NA = not available.

Figure IV.1. Analysis Matching Rate, by State



Source: Mathematica matching analysis of NSLP application and State SNAP participation data.

In all but two States in the study, we found more matches using probabilistic matching than we did using deterministic matching. This was expected for two reasons: (1) more straightforward matches are likely to have been certified through direct certification rather than by application; and (2) our probabilistic algorithms allowed more flexible matching, accepting inexact matches and a wider range of data element combinations. The two States in which this pattern did not hold—Indiana and Nebraska—were also the States with the highest overall match rates in our study. That we were able to identify so many matches using simple exact match algorithms—more than 40 percent of the sample in each State—could suggest the States failed to certify directly fairly easy-to-match students. Alternatively, it could indicate that the districts sampled in those States did not reclassify students initially certified by application who were later directly certified. If this is the case, we might have included students in our analysis who were indeed directly certified, inflating the match rate in those districts compared with districts that do reclassify such students.

The matching results in the States with more probabilistic than deterministic matches illustrate the potential advantages of incorporating more flexible matching into direct certification systems. Probabilistic matching can be particularly valuable for matching students with long or uncommon last names. It also can be an effective way to match students with missing data. Next, we explore these themes in greater detail and present the individual matching algorithms and State results.

1. Independent Matching Results for Alabama

The data elements available in both the SNAP program data and NSLP applications in Alabama included first name, last name, address, parent name, and (for 20 percent of observations), SSN. We used similar algorithms for the deterministic and probabilistic matching processes. Because only two of the four sampled districts were able to provide application data for this study, these results are unweighted and might not be representative of the entire State.

Independent Matching Algorithms for Alabama

Deterministic Match

Must exactly match all of the following:
- First name
- Last name
- Address
- Parent's full name

OR

Must exactly match all of the following:
- First name
- Last name
- SSN

Probabilistic Match

Must closely match all of the following:
- First name
- Last name

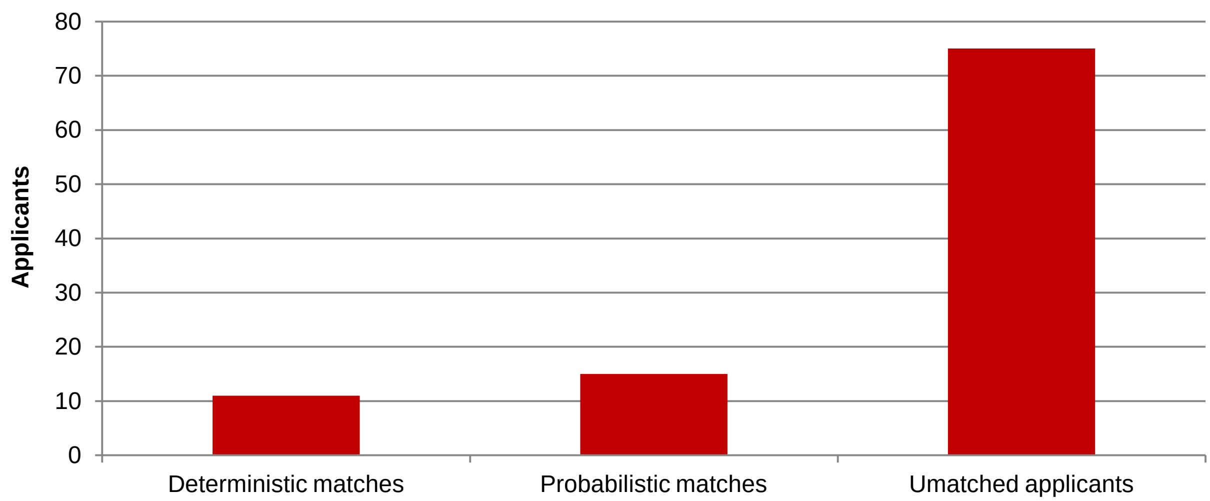
And closely match at least two of the following:
- Address
- SSN
- Parent's full name

OR

Must match all of the following:
- SSN (exact match)
- First name (close match)

And closely match at least one of the following:
- Last name
- Address
- Parent's full name

Figure IV.2. Alabama Matching Results



Source: Mathematica matching analysis of Alabama Department of Human Resources records of matched and unmatched SNAP participants and NSLP application data from sampled school districts.

The relatively low matching rate for the Alabama analysis might be related to the limited data elements available. Date of birth was not available in the student application data. Although SSNs were included in data, they were available for only a small portion of the caseload. Therefore, most student matching was based on first name, last name, address, and parent's full name.

Table IV.3 compares the characteristics of categorically eligible applicants who were (1) deterministically matched, (2) probabilistically matched, or (3) not independently matched. These results show important differences in name commonality and missing data patterns. Categorically eligible applicants who were matched deterministically had much more common first names than those who were unmatched (percentile 54.7 versus 37.7 of name commonality); although large in magnitude, this difference is not statistically significant due to the small sample size.

Categorically eligible applicants who were matched deterministically had much more complete data than the other two groups. Only 36.4 percent of deterministic matches were missing any data element, compared with 80.0 percent for probabilistic matches and 88.0 percent for unmatched applications. Additionally, more than four-fifths of deterministic matches included SSNs, compared with one-fifth of probabilistic matches and less than 15 percent of unmatched applications (Table IV.3). This finding highlights the fact that students with higher quality data are easier to match successfully with the deterministic matching approach. The fact that the rates of missing data were relatively high for probabilistic matches points to the role probabilistic matching can play in identifying matches for cases with incomplete data.

Table IV.3. Average Characteristics of Students Certified for NSLP Benefits Based on Categorical Eligibility by Application in Alabama, by whether Matched to State SNAP Data (percentages unless otherwise noted)

Characteristic	Matched Through Deterministic Process	Matched Through Probabilistic Process Only	Unmatched Applications
First Name Commonality[a]			
Average percentile	54.7	41.2	37.7
Average name length (number of letters)	6.00	6.13	6.28
Last Name Commonality[b]			
Average percentile	62.8	45.9	49.8
Average name length (number of letters)	5.82	6.07	6.21
Missing Data on:			
First name	0.0	0.0	0.0
Last name	0.0	0.0	0.0
SSN	18.2*	80.0††	86.7
Address	18.2	0.0	1.3
Parent name	0.0	0.0	0.0
Any element	36.4**	80.0	88.0
Multiple elements	0.0	0.0	0.0
Sample Size	**11**	**15**	**75**

Source: Alabama Department of Human Resources, records of matched and unmatched SNAP participants. NSLP application data from sampled school districts.

Note: Information on age is not available for Alabama.

[a] Based on Social Security Administration records from 1994 to 2007.

[b] Based on 2000 Decennial Census data.

*/**/*** Mean is significantly different from mean of unmatched records at the .10/.05/.01 levels, respectively.

†/††/††† Mean is significantly different from mean of deterministic matches at the .10/.05/.01 levels, respectively.

2. Independent Matching Results for Arizona

The data elements available in both the SNAP program data and NSLP applications in Arizona included first name, last name, date of birth, address, parent's name, and SNAP or TANF case number.

Although about two-thirds of categorically eligible applicants in the Arizona sample were matched to State SNAP records, we identified few matches through the deterministic process. Of the 832 applicants included in the analysis, only 7 percent (60 applicants) were matched deterministically, whereas 56 percent (464 applicants) were matched probabilistically. We identified far more probabilistic matches in Arizona, in raw numbers and as a share of the State sample, than in any other State.

Independent Matching Algorithms for Arizona

Deterministic Match

Must exactly match all of the following:
- First name
- Last name
- Date of birth

And at least one of the following:
- Address
- SNAP or TANF case number
- Parent's full name

Probabilistic Match

Must closely match all of the following:
- First name
- Last name

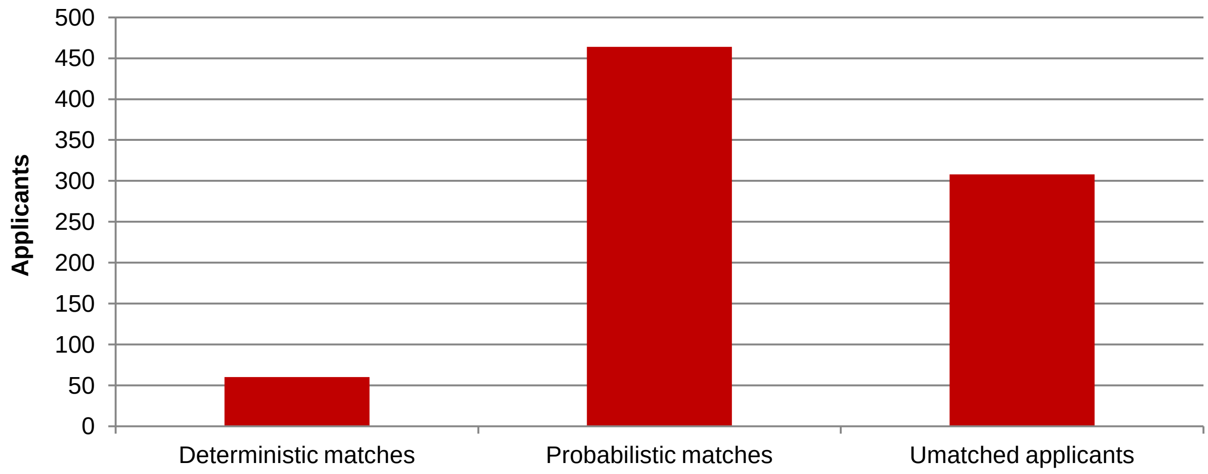
And closely match at least two of the following:
- Address
- Date of birth
- SNAP or TANF case number
- Parent's full name

OR

Must closely match:
- First name

And closely match at least three of the following:
- Last name
- Date of birth
- Address
- SNAP or TANF case number
- Parent's full name

Figure IV.3. Arizona Matching Results



Source: Mathematica matching analysis of Arizona Department of Economic Security records of matched and unmatched SNAP participants and NSLP application data from sampled school districts.

The prevalence of longer, often compound, last names might explain the small number of deterministic matches and large number of probabilistic matches in Arizona. Longer, less common names could have a greater potential for data errors and inconsistencies, hindering effective matching with deterministic systems. The manual review process revealed many cases of close but inexact matches among last names—more than half of probabilistic matches in Arizona relied on inexact last name matches. Consistent with this hypothesis, probabilistically matched applicants had last names almost twice as long, on average, as those of deterministically matched applicants (Table IV.4). Probabilistically matched students also had much less common last names than deterministically matched applicants (percentile 17.2 versus 46.5 of last name commonality). Compound last names and common Hispanic last names represented a large share of the inexactly matched last names.

Date of birth proved to be an important matching element in Arizona. Our deterministic algorithm required it, so no deterministic matches lacked it. However, even under the more flexible probabilistic algorithm, only 17 percent of students lacked date of birth, compared with half of unmatched students.

Table IV.4. Average Characteristics of Students Certified for NSLP Benefits Based on Categorical Eligibility by Application in Arizona, by Whether Matched to State SNAP Data (percentages unless otherwise noted)

Characteristic	Matched Through Deterministic Process	Matched Through Probabilistic Process Only	Unmatched Applications
Age[a]		●●	
Younger than 5	1.3	1.2	0.3
5–9	45.1	47.8	17.8
10–14	46.8	28.4	23.0
15–18	6.8	5.6	8.9
19 or older	0.0	0.0	0.0
Missing	0.0	17.0	50.0
Mean (years)	10.0	9.6	11.3
First Name Commonality[b]			
Average percentile	36.5	38.1	35.4
Average name length (number of letters)	6.03	6.44	6.18
Last Name Commonality[c]			
Average percentile	46.5	17.2††**	43.4
Average name length (number of letters)	6.66	11.18†**	7.40
Missing Data on:			
First name	0.0	0.0	0.0
Last name	0.0	0.0	0.0
Date of birth	0.0	17.0	50.0
Address	0.0	0.3	1.4
Parent name	96.0	82.1	50.3
SNAP case number	4.0	2.2	8.4
Any element	100.0	100.0	100.0
Multiple elements	0.0	1.6	10.1
Sample Size	**60**	**464**	**309**

Source: Arizona Department of Economic Security, records of matched and unmatched SNAP participants. NSLP application data from sampled school districts.

[a] Differences between group distributions were tested using a chi-squared test.
[b] Based on Social Security Administration records from 1994 to 2007.
[c] Based on 2000 Decennial Census data.

*/**/*** Mean is significantly different from mean of unmatched records at the .10/.05/.01 levels, respectively.
†/††/††† Mean is significantly different from mean of deterministic matches at the .10/.05/.01 level, respectively.
●/●●/●●● Distribution is significantly different from distribution of unmatched records at the .10/.05/.01 levels, respectively.
○/○○/○○○ Distribution is significantly different from distribution of deterministic matches at the .10/.05/.01 levels, respectively.

3. Independent Matching Results for Connecticut

The data elements available in both the SNAP program data and NSLP applications in Connecticut included first and last names, date of birth, and address. Because the independent match in Connecticut is based on only four data elements, students missing any data elements cannot be matched; about 25 percent of the 232 categorically eligible applicants in the original Connecticut sample were missing at least one data element.

Independent Matching Algorithms for Connecticut

Deterministic Match

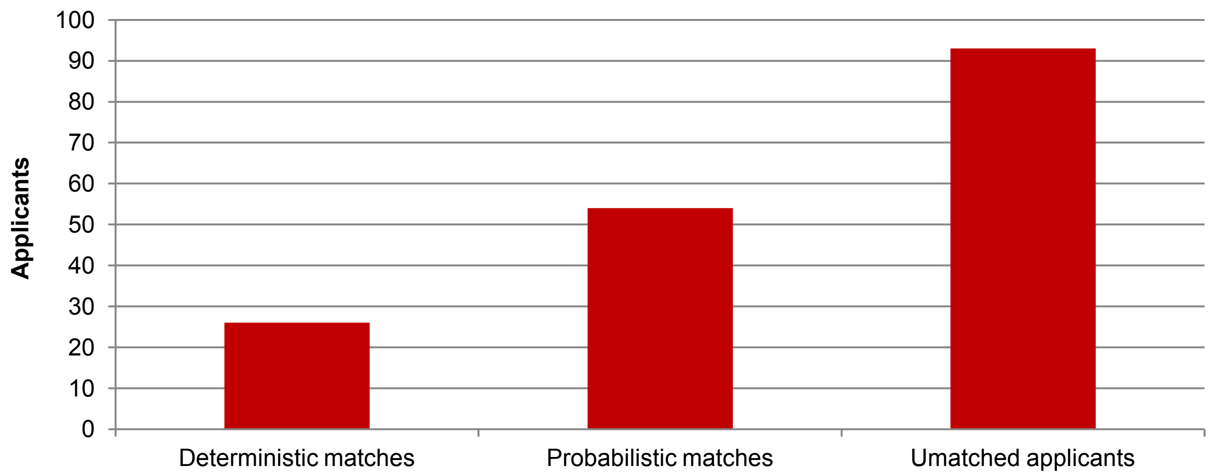
Must exactly match all of the following:
- First name
- Last name
- Date of birth
- Address

Probabilistic Match

Must closely match all of the following:
- First name
- Last name
- Address
- Date of birth

Among the 173 categorically eligible applicants with nonmissing values in all four data elements, 46 percent (80 applicants) were matched to State SNAP records: 15 percent (26 applicants) were deterministic matches and 31 percent (54 applicants) were probabilistic matches.

Figure IV.4. Connecticut Matching Results



Source: Mathematica matching analysis of Connecticut Department of Social Services records of matched and unmatched SNAP participants and NSLP application data from sampled school districts.

Table IV.5 compares the characteristics of categorically eligible applicants' independent match status. The findings related to name commonality follow a different pattern than those of most other States. Probabilistically matched applicants have less common first names, on average, than those of deterministically matched applicants. However, probabilistically matched applicants have more common last names, on average, than those of deterministically matched applicants. Potentially relevant to this latter finding is that last names were not what prevented probabilistically matched students from matching deterministically; all but three probabilistic matches matched exactly on last name. Rather, variation in the address field distinguished probabilistic from deterministic matches—49 of the 54 probabilistic matches relied on inexact matches in address.

Table IV.5. Average Characteristics of Students Certified for NSLP Benefits Based on Categorical Eligibility by Application in Connecticut, by whether Matched to State SNAP Data (percentages unless otherwise noted)

Characteristic	Matched Through Deterministic Process	Matched Through Probabilistic Process Only	Unmatched Applications
Age[a]			
Younger than 5	0.0	5.6	4.3
5–9	26.9	51.9	31.2
10–14	46.2	37.0	39.8
15–18	26.9	5.6	22.6
19 or older	0.0	0.0	2.2
Missing	0.0	0.0	0.0
Mean (years)	12.4*	9.8	11.7
First Name Commonality[b]			
Average percentile	56.4**	45.4††*	52.8
Average name length (number of letters)	6.46	6.28	6.53
Last Name Commonality[c]			
Average percentile	44.7	57.8††**	47.6
Average name length (number of letters)	6.54	6.41	7.04
Sample Size	**26**	**54**	**93**

Source: Connecticut Department of Social Services, records of matched and unmatched SNAP participants. NSLP application data from sampled school districts.

[a] Differences between group distributions were tested using a chi-squared test.
[b] Based on Social Security Administration records from 1994 to 2007.
[c] Based on 2000 Decennial Census data.

*/**/*** Mean is significantly different from mean of unmatched records at the .10/.05/.01 levels, respectively.
†/††/††† Mean is significantly different from mean of deterministic matches at the .10/.05/.01 levels, respectively.
●/●●/●●● Distribution is significantly different from distribution of unmatched records at the .10/.05/.01 levels, respectively.
○/○○/○○○ Distribution is significantly different from distribution of deterministic matches at the .10/.05/.01 levels, respectively.

4. Independent Matching Results for Indiana

The data elements available in both the SNAP program data and NSLP applications in Indiana included first and last names, date of birth, and address. Because only two of the four sampled districts were able to provide application data for this study, these results are unweighted and might not be representative of the entire State. Because the independent match in Indiana is based on only four data elements, students missing any data elements cannot be matched; however, no categorically eligible applicants in the Indiana sample were missing any data elements.

Independent Matching Algorithms for Indiana

Deterministic Match

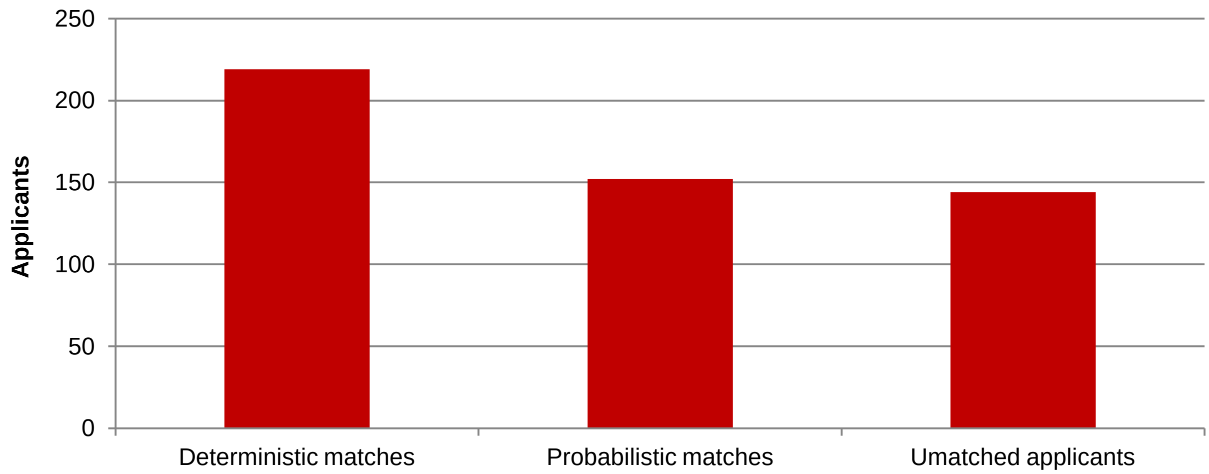
Must exactly match all of the following:
- First name
- Last name
- Date of birth
- Address

Probabilistic Match

Must closely match all of the following:
- First name
- Last name
- Address
- Date of birth

Despite the limited range of matching elements available, we achieved our second highest overall match rate (72 percent) and third highest probabilistic match rate (30 percent) with Indiana. Nearly three-quarters of the categorically eligible applicants in the Indiana sample were matched to State SNAP records. Of the 515 applications, we identified 219 deterministic matches and 152 probabilistic matches.

Figure IV.5. Indiana Matching Results



Source: Mathematica matching analysis of Indiana Family and Social Services Administration records of matched and unmatched SNAP participants and NSLP application data from sampled school districts.

Indiana's probabilistic matches had less common first names than the deterministic matches (Table IV.6). This was not a significant barrier to their matching, however; more than 90 percent of probabilistic matches matched exactly on first name. As in Connecticut, nearly all probabilistic matches relied on inexact matches in addresses—146 out of 152 matches.

Table IV.6. Average Characteristics of Students Certified for NSLP Benefits Based on Categorical Eligibility by Application in Indiana, by whether Matched to State SNAP Data (percentages unless otherwise noted)

Characteristic	Matched Through Deterministic Process	Matched Through Probabilistic Process Only	Unmatched Applications
Age[a]			
Younger than 5	1.4	3.3	0.0
5–9	24.7	46.7	35.3
10–14	34.7	30.3	41.0
15–18	37.9	19.7	22.3
19 or older	1.4	0.0	1.4
Missing	0.0	0.0	0.0
Mean (years)	13.1***	11.1†††**	12.0
First Name Commonality[b]			
Average percentile	38.4***	34.3**	27.2
Average name length (number of letters)	6.26	6.14	6.24
Last Name Commonality[c]			
Average percentile	57.9	58.8	56.6
Average name length (number of letters)	6.27	6.36	6.37
Sample Size	**219**	**152**	**139**

Source: Indiana Family and Social Services Administration, records of matched and unmatched SNAP participants. NSLP application data from sampled school districts.

[a] Differences between group distributions were tested using a chi-squared test.
[b] Based on Social Security Administration records from 1994 to 2007.
[c] Based on 2000 Decennial Census data.

*/**/*** Mean is significantly different from mean of unmatched records at the .10/.05/.01 levels, respectively.
†/††/††† Mean is significantly different from mean of deterministic matches at the .10/.05/.01 levels, respectively.
●/●●/●●● Distribution is significantly different from distribution of unmatched records at the .10/.05/.01 levels, respectively.
○/○○/○○○ Distribution is significantly different from distribution of deterministic matches at the .10/.05/.01 levels, respectively.

5. Independent Matching Results for Nebraska

The data elements available in both the SNAP program data and NSLP applications in Nebraska included first and last names, date of birth, address, parent's name, and SNAP or TANF case number.

Independent Matching Algorithms for Nebraska

Deterministic Match

Must exactly match all of the following:
- First name
- Last name
- Date of birth

And at least one of the following:
- Address
- SNAP or TANF case number
- Parent's full name

Probabilistic Match

Must closely match all of the following:
- First name
- Last name

And closely match at least two of the following:
- Address
- Date of birth
- SNAP or TANF case number
- Parent's full name

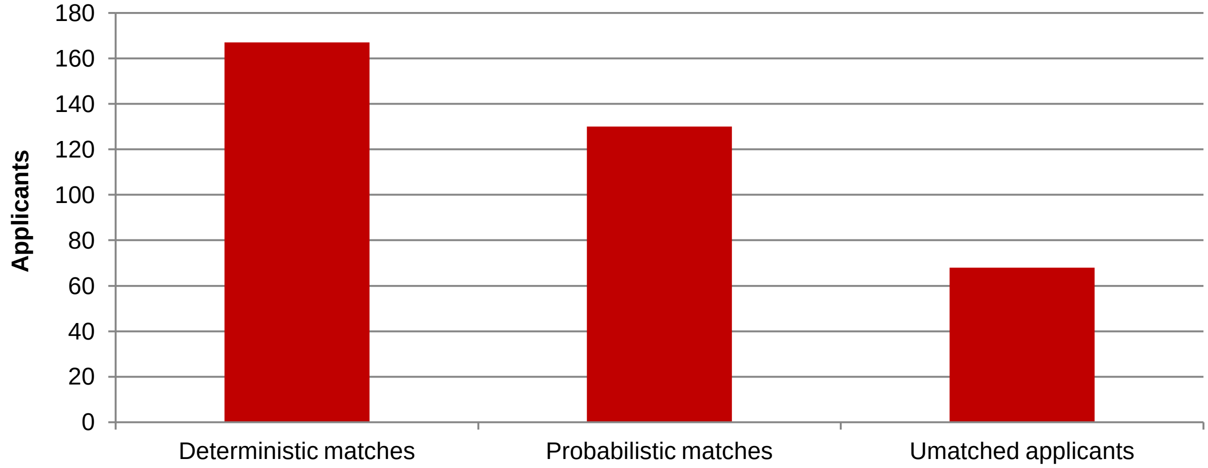
OR

Must closely match:
- First name

And closely match at least three of the following:
- Last name
- Date of birth
- Address
- SNAP or TANF case number
- Parent's full name

Independent matching for Nebraska yielded the highest match rate in the study. Of the 365 applications from Nebraska, 54 percent (167 applicants) were deterministic matches and 36 percent (130 applicants) were probabilistic matches, accounting for 81 percent of the categorically eligible applicant sample for Nebraska.

Figure IV.6. Nebraska Matching Results



Source: Mathematica matching analysis of Nebraska Department of Health and Human Services records of matched and unmatched SNAP participants and NSLP application data from sampled school districts.

The high match rate in Nebraska is surprising because the State already incorporates probabilistic matching in its direct certification procedures and a very high percentage of students certified for school meals based on categorical eligibility were certified through direct certification rather than application (Figure I.1). Therefore, we would have expected the State already to have directly certified most eligible students who were easy to match. The high independent match rate that we found, however, could largely reflect the timeliness of the enrollment data used in Nebraska's initial match. For most of the school year, the State's direct certification matching procedure uses continuously updated school enrollment data. Yet for the initial match, Nebraska requires districts to upload current enrollment data files as of the beginning of the new school year. If the districts do not do so, the initial match would be performed with the previous year's enrollment data. This would likely result in some categorically eligible students not being directly certified until after the school year began and submitting an application for school meal benefits. As noted earlier, our analysis sample of categorically eligible applicants will include any children initially certified for free school meals based on the NSLP applications, later matched through direct certification, and not reclassified as having been directly certified.

Table IV.7 compares the characteristics of categorically eligible applicants' independent match status. As in most other States, probabilistically matched applicants had much less common last names than other students (percentile 27.9 in last name commonality versus 47.5 and 41.0 for deterministically matched and unmatched students, respectively). Very few sampled applicants in Nebraska had missing data, although unmatched applicants were more likely than probabilistically matched applicants to have any missing data elements.

Table IV.7. Average Characteristics of Students Certified for NSLP Benefits Based on Categorical Eligibility by Application in Nebraska, by whether Matched to State SNAP Data (percentages unless otherwise noted)

Characteristic	Matched Through Deterministic Process	Matched Through Probabilistic Process Only	Unmatched Applications
Age[a]			
Younger than 5	3.8	9.5	6.9
5–9	49.3	41.7	37.8
10–14	32.9	31.5	32.4
15–18	14.0	16.6	19.1
19 or older	0.0	0.7	1.3
Missing	0.0	0.0	2.6
Mean (years)	10.3	10.3	10.8
First Name Commonality[b]			
Average percentile	34.5	24.2††**	38.9
Average name length (number of letters)	6.08	5.95††	5.95
Last Name Commonality[c]			
Average percentile	47.5	27.9†††	41.0
Average name length (number of letters)	6.27	6.67	6.28
Missing Data on:			
First name	0.0	0.0	0.0
Last name	0.0	0.0	0.0
Date of birth	0.0	0.0	1.3
Address	0.0	0.0	0.0
Parent name	0.0	0.0	0.0
SNAP case number	1.7	0.7†	3.8
Any element	1.7	0.7†	5.1
Multiple elements	0.0	0.0	0.0
Sample Size	**167**	**130**	**68**

Source: Nebraska Department of Health and Human Services, records of matched and unmatched SNAP participants. NSLP application data from sampled school districts.

[a] Differences between group distributions were tested using a chi-squared test.
[b] Based on Social Security Administration records from 1994 to 2007.
[c] Based on 2000 Decennial Census data.

*/**/*** Mean is significantly different from mean of unmatched records at the .10/.05/.01 levels, respectively.
†/††/††† Mean is significantly different from mean of deterministic matches at the .10/.05/.01 levels, respectively.
●/●●/●●● Distribution is significantly different from distribution of unmatched records at the .10/.05/.01 levels, respectively.
○/○○/○○○ Distribution is significantly different from distribution of deterministic matches at the .10/.05/.01 levels, respectively.

6. Independent Matching Results for Texas

The data elements available in both the SNAP program data and NSLP applications in Texas included first and last names, date of birth, and address. Because the independent match in Texas is based on only four data elements, students missing any data elements cannot be matched; about 33 percent of the 893 categorically eligible applicants in the original Texas sample were missing at least one data element.

Independent Matching Algorithms for Texas

Deterministic Match

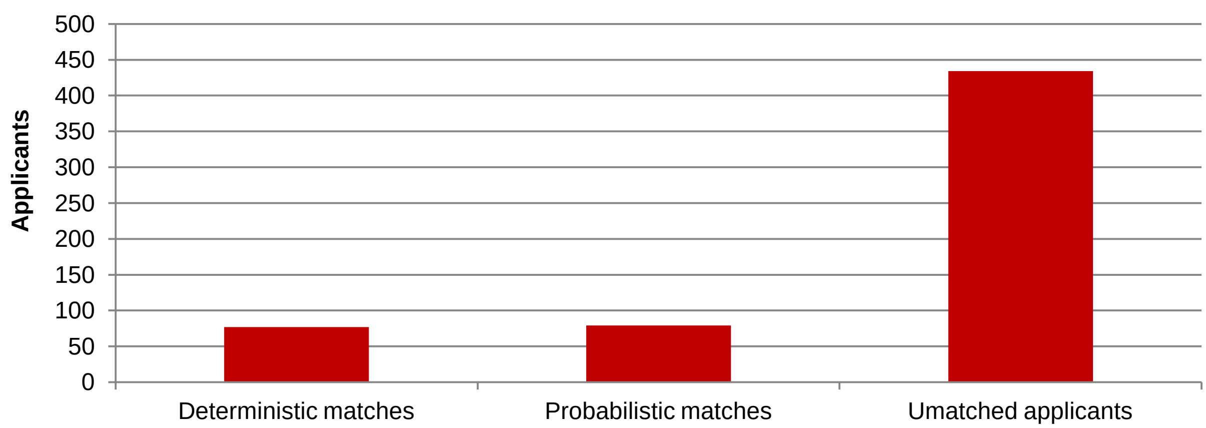
Must exactly match all of the following:
- First name
- Last name
- Date of birth
- Address

Probabilistic Match

Must closely match all of the following:
- First name
- Last name
- Address
- Date of birth

Independent matching for Texas yielded a relatively low matching rate. Among 595 applicants with no missing data elements, we matched 77 deterministically and 79 probabilistically, accounting for about one-quarter of the sample combined.

Figure IV.7. Texas Matching Results



Source: Mathematica matching analysis of Texas Health and Human Services Commission records of matched and unmatched SNAP participants and NSLP application data from sampled school districts.

Table IV.8 compares the characteristics of categorically eligible applicants by independent match status. As in most other States, probabilistically matched applicants had much less common last names than other students (percentile 28.9 in last name commonality versus 49.5 and 49.9 for deterministically matched and unmatched applicants, respectively). Probabilistically matched applicants also had less common first names than deterministically matched applicants and substantially longer last names than either other group.

Table IV.8. Average Characteristics of Students Certified for NSLP Benefits Based on Categorical Eligibility by Application in Texas, by Whether Matched to State SNAP Data (percentages unless otherwise noted)

Characteristic	Matched Through Deterministic Process	Matched Through Probabilistic Process Only	Unmatched Applications
Age[a]	●●	●	
Younger than 5	1.5	2.5	21.6
5–9	53.1	56.9	42.5
10–14	33.8	31.5	25.4
15–18	11.5	7.6	10.3
19 or older	0.0	1.4	0.3
Missing	0.0	0.0	0.0
Mean (years)	10.2**	9.7	9.1
First Name Commonality[b]			
Average percentile	47.0***	37.5††	37.1
Average name length (number of letters)	6.09	6.24	6.20
Last Name Commonality[c]			
Average percentile	49.5	28.9†††***	49.9
Average name length (number of letters)	6.10***	9.38†††***	7.62
Sample Size	**77**	**79**	**434**

Source: Texas Health and Human Services Commission, records of matched and unmatched SNAP participants. NSLP application data from sampled school districts.

[a] Differences between group distributions were tested using a chi-squared test.
[b] Based on Social Security Administration records from 1994 to 2007.
[c] Based on 2000 Decennial Census data.

*/**/*** Mean is significantly different from mean of unmatched records at the .10/.05/.01 levels, respectively.
†/††/††† Mean is significantly different from mean of deterministic matches at the .10/.05/.01 levels, respectively.
●/●●/●●● Distribution is significantly different from distribution of unmatched records at the .10/.05/.01 levels, respectively.
○/○○/○○○ Distribution is significantly different from distribution of deterministic matches at the .10/.05/.01 levels, respectively.

7. Independent Matching Results for West Virginia

The data elements available in both the SNAP program data and NSLP applications in West Virginia included first and last names, date of birth, address, parent's name, and SNAP or TANF case number.

West Virginia Matching Algorithms

Deterministic Match

Must exactly match all of the following:
- First name
- Last name
- Date of birth

And at least one of the following:
- Address
- SNAP or TANF case number
- Parent's full name

Probabilistic Match

Must closely match all of the following:
- First name
- Last name

And closely match at least two of the following:
- Address
- Date of birth
- SNAP or TANF case number
- Parent's full name

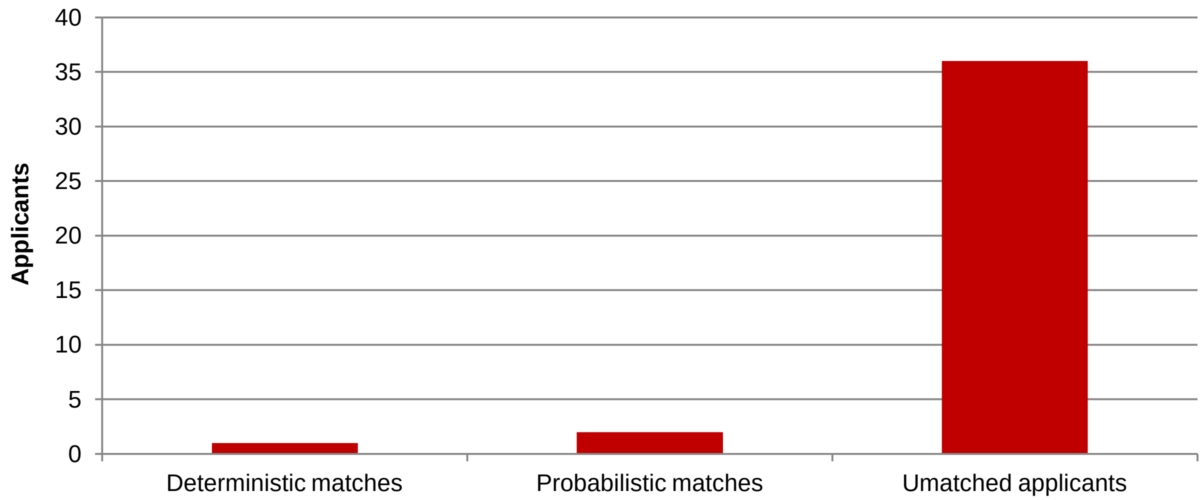
OR

Must closely match:
- First name

And closely match at least three of the following:
- Last name
- Date of birth
- Address
- SNAP or TANF case number
- Parent's full name

Almost no categorically eligible applicants in West Virginia were independently matched to State SNAP records. Only one of the 39 applicants matched deterministically and two matched probabilistically. This low match rate is at least partially related to incomplete data. Although the application files contained a fairly broad range of potential matching elements, large numbers of applications had blank addresses, case numbers, and parents' names. The small matching result might also be due to the relatively small number of students certified for NSLP benefits based on categorical eligibility by application in West Virginia or the relative success of the direct certification process in identifying categorically eligible students in State records.

Figure IV.8. West Virginia Matching Results



Source: Mathematica matching analysis of West Virginia Department of Health and Human Resources records of matched and unmatched SNAP participants. NSLP application data from sampled school districts.

Because of the small numbers of matched students, it is not possible to draw inferences about differences in the characteristics of categorically eligible applicants by independent match status. Therefore, we do not show applicant characteristics by independent match status for the West Virginia sample.

C. Cross-State Themes

Although independent matching rates varied considerably among States in this study, some common themes emerge in the results pertaining to the importance of name characteristics in matching, the importance of complete data, and the potential for probabilistic matching to overcome barriers related to complex data.

- **Longer, less common names are a barrier to deterministic matching.**

In all study States, the name characteristics of categorically eligible applicants are significantly related to independent matching status. In most cases, these relationships are consistent with longer, less common names being more difficult to match deterministically. These findings are likely related to the difficulty of recording less common and longer (particularly compound) names consistently. This pattern is clearest in Arizona, Nebraska, and Texas, where deterministically matched applicants have substantially more common names than their probabilistically matched counterparts (Tables IV.4, IV.7, and IV.8). In Arizona and Texas, differences in name length are very large as well. These findings are important because they suggest that name characteristics are a key factor in preventing matches of students who are both categorically eligible for school meal benefits and identifiable in State SNAP records.

The ability to match long or uncommon last names can have implications for matching outcomes for different racial and ethnic groups. One district in Texas provided data on student race and ethnicity in its NSLP application records. The matching results for that district show large, statistically significant differences in race and ethnicity between probabilistic and deterministic matches. More than 80 percent of deterministically matched students were white, compared with 41 percent of probabilistically matched students (Table IV.9). More than half of probabilistically matched students were Hispanic, compared with less than 20 percent of deterministic matches. Differences in name characteristics likely explain these divergent racial and ethnic outcomes: probabilistic matches had less common first and last names, and much longer last names. Probabilistically matched students had first names in percentile 33.1 and last names in percentile 31.0. Deterministic matches had names in percentile 52.3 and 48.4. Probabilistically matched last names were more than three letters longer, on average, than deterministically matched names.

Table IV.9. Average Characteristics of Students Certified for NSLP Benefits Based on Categorical Eligibility by Application in a Rural, Medium-Sized Texas District, by Whether Matched to State SNAP Data (percentages unless otherwise noted)

Characteristic	Matched Through Deterministic Process	Matched Through Probabilistic Process Only	Unmatched Applications
Race/Ethnicity[a]	●●●	○○○	
African American, not Hispanic	0.0	2.7	1.4
Asian, not Hispanic	0.0	0.0	2.1
White, not Hispanic	80.4	40.5	31.5
Hispanic	19.6	56.8	65.0
First Name Commonality[b]			
Average percentile	52.3**	33.1†††**	42.6
Average name length (number of letters)	6.10	6.29	5.99
Last Name Commonality[c]			
Average percentile	48.4	31.0†††**	43.2
Average name length (number of letters)	6.03***	9.29†††***	7.77
Sample Size	**63**	**41**	**176**

Source: Texas Health and Human Services Commission, records of matched and unmatched SNAP participants. NSLP application data from sampled school district.

[a] Differences between group distributions were tested using a chi-squared test.
[b] Based on Social Security Administration records from 1994 to 2007.
[c] Based on 2000 Decennial Census data.

*/**/*** Mean is significantly different from mean of unmatched records at the .10/.05/.01 levels, respectively.
†/††/††† Mean is significantly different from mean of deterministic matches at the .10/.05/.01 levels, respectively.
●/●●/●●● Distribution is significantly different from distribution of unmatched records at the .10/.05/.01 levels, respectively.
○/○○/○○○ Distribution is significantly different from distribution of deterministic matches at the .10/.05/.01 levels, respectively.

- **Complete data make deterministic matching easier.**

For States for which the independent matching is based on only four data elements (Connecticut, Indiana, and Texas), no student with a missing data element can be matched with this study's independent matching algorithm. This requirement did not exclude any students in Indiana; however, one-quarter of the Connecticut sample and one-third of the Texas sample were not matchable due to missing data elements. Patterns of missing data by independent match status in the other States in this study also highlight the importance of complete data. For example, in Alabama, deterministically matched applicants were dramatically less likely to have missing data elements than were probabilistically matched applicants and unmatched applicants. Similarly, the very high match rates in Nebraska might be related to the very low rates of missing data there. It is important to note again, however, that the application data used in this study's independent matching analysis are likely of lower quality than the student enrollment data available for direct certification matching. Enrollment data likely have more data fields containing identifying student information, more data maintenance and quality checks, and less missing data.

- **Probabilistic matching might offer a way to overcome barriers related to data recording difficulty and data completeness.**

Findings from the independent match analysis point to the usefulness of probabilistic matching in resolving issues related to harder-to-record data items, such as longer, less common names and address. This is supported by the findings discussed previously related to differences in name characteristics for deterministically and probabilistically matched applicants. It is further supported by the finding in some States that, in many cases, address is the data field that prevents a deterministic match. A primary challenge in matching using street addresses is the existence of multiple correct variations. A large portion of inexact matches on street address resulted from variations such as *Street* versus *St.* or different ways to represent apartment numbers. Therefore, relaxing the requirement of an exact match in every field through probabilistic matching allows for verification of the eligibility of more categorically eligible students. This conclusion also applies to missing data; by relaxing the requirement that all data elements are nonmissing, more students who legitimately match State SNAP records can be identified.

V. CONCLUSION

This report has examined the characteristics of categorically eligible students not matched by direct certification and explored implications of the findings for improving direct certification matching processes. Seven States participated in this study, submitting SNAP caseload data and NSLP application data from sampled districts. We used these data in two sets of analysis. In the first, we analyzed statewide SNAP participant data from two participating States—Arizona and West Virginia—comparing characteristics of children who were directly certified and those who were not. Because all children in this sample are categorically eligible for free school meals, these comparisons identified patterns in age, name characteristics, and local area school and economic characteristics associated with more successful or more challenging direct certification.

In the second part of the study, we analyzed data on children certified for school meal benefits by application based on categorical eligibility. These data, drawn from randomly sampled districts in all seven participating States, represented categorically eligible students who could have been matched in direct certification but were not. We sought to identify these categorically eligible applicants in State-level SNAP participation files using a two-stage matching process. In the first stage, we conducted a deterministic match, requiring exact matches on key variables such as name and date of birth. This process mirrored the deterministic processes in place in many States. In the second stage, we used a probabilistic match that incorporated more flexible algorithms and allowed inexact matches between data fields. We conducted this match using off-the-shelf probabilistic matching software available for purchase to entities conducting matching and similar to other available matching software.[1]

The results of this independent matching process indicate the extent to which students categorically eligible for NSLP benefits can be identified in SNAP participation data. By comparing the characteristics of students matched deterministically, those matched probabilistically, and those not matched in either process, we identified characteristics associated with more challenging matching, as well as the potential value of probabilistic matching in direct certification.

Here, we synthesize some of the study's key findings:

- **Systemic features of the data—such as data completeness, data richness, and integration of private school data—might be associated with the success of matching.**

Data completeness. In our independent matching analysis, results from Alabama and West Virginia demonstrated most clearly the difficulty of identifying matches with high levels of missing data. These States had the lowest overall matching rates in the study and among the highest rates of missing data. Moreover, in Alabama, students who were matched deterministically had more

---

[1] We used the probabilistic software to identify most likely matches between NSLP application and SNAP participation data. We then manually reviewed each prospective match. We did not use confidence score thresholds to identify matches in our analysis. See Chapter 1 for more details on our matching process.

complete data than students who were not.[2] Nebraska had the highest matching rate in the study and the most complete data. Our analysis reinforced an intuitive principle of data matching: without complete data, it is difficult to establish matches between enrollment and program data.

Data richness. Including a larger number of data elements common across enrollment and program data can make it easier to identify matches. This principle of matching interacts with the first: having many variables available can compensate for some data incompleteness. In our independent matching analysis, we achieved a high matching rate with Arizona, despite incomplete data, in part because the State's data included six matching elements. All applicants in the data set lacked data in at least one field, but less than 6 percent were missing data in more than one. In contrast, in West Virginia, nearly 80 percent of applicants lacked data in multiple fields. Having six matching fields available could not compensate for the scale of the missing data. Nebraska, the State with the highest matching rate in our independent matching, combined both advantages. The State's data had six matching elements available and a very low rate of missing data, enabling us to match more than 80 percent of the sample.

Integration of private school data. Our analysis of SNAP records in Arizona and West Virginia revealed that children with SNAP records residing in counties with higher levels of private school enrollment were less likely to be matched through direct certification. This is likely due to some private schools not using the same data submission and other direct certification processes as public schools.

- **Even with systemic data needs addressed, data elements that are difficult to store consistently present challenges to direct certification matching.**

Across our analysis of State SNAP records and our independent matching analysis, students with longer, less common names—particularly last names—were less likely to be matched than those with shorter, more common names. This trend held across nearly every State in the study and was particularly evident in Arizona and Texas, States with large Hispanic populations.

The second data challenge to direct certification we identified was data variations or errors. Misspellings, illegible applications, nicknames, alternate last names, and variations in street address spelling can all hinder effective matching. In some cases, these could be data errors. In others, they could simply be two variations of correct spelling (such as *St.* versus *Street*). Similarly, if program and enrollment data represent different periods of time, families could move, resulting in mismatched addresses. These variations can complicate matching even with complete, rich data sources.

- **Flexible matching algorithms and probabilistic matching might be effective strategies in mitigating some challenges to successful matching.**

Our independent matching analysis revealed several matching strategies that can alleviate the effect these challenges have on matching results. First, flexible matching algorithms can take advantage of rich data sets to identify matches, even with some data inconsistencies. The matching results in Nebraska highlight the advantages of using rich data sets to incorporate flexible matching

[2] We cannot make an analogous comparison for West Virginia because so few students were matched in that State. As noted, this low match rate is at least partially due to the fact than 80 percent of students were missing multiple data elements.

algorithms. In Nebraska, 17 percent of the probabilistic matches were possible, not because they had inexact matches in some data fields, but because they simply did not match on last name, a field required in our deterministic algorithm. Because our probabilistic algorithm allowed a greater range of variable combinations, we were able to confirm these matches based on other available data fields. Flexible deterministic algorithms are possible as well, if sufficient data elements are available.

Results presented here indicate that probabilistic matching can also help combat some data challenges discussed earlier. Probabilistic matching increases matching results by allowing inexact matches. Minor variations in spelling, street abbreviations, or other inconsistencies are less likely to preclude a match in probabilistic matching than in a deterministic system. In our independent matching analysis, a large majority of categorically eligible applicants who were matched to States SNAP records were identified through probabilistic matching. This finding is important because, as noted previously, these applicants represent students who were categorically eligible for free meal benefits but who were not directly certified. Probabilistic matching proved particularly valuable at identifying matches among students with long or uncommon last names—students direct certification systems often have difficulty matching. This advantage was most apparent in Arizona and Texas, where probabilistically matched students had much longer and less common names than deterministically matched children. In Arizona, probabilistic matching compensated for a very low deterministic match rate, leaving Arizona with the third highest overall match rate in the study.

Probabilistic matching does require additional software and more staff effort than deterministic matching.[3] Our methods differed somewhat from the processes States or districts would likely use when performing probabilistic matching. We deviated from how a typical user might operate probabilistic matching software in conducting direct certification matching. Specifically, after receiving the results, we manually reviewed each prospective matched pair, sorting the pairs in a Microsoft Excel spreadsheet to identify those that met our matching algorithms. This process was likely more labor intensive than staff in direct certification systems would probably use, and it was feasible only because the sample sizes in our study were smaller than State-level school enrollment files (we manually reviewed slightly more than 3,000 matched pairs). However, probabilistic matching software offers tools that can lower the burden of manual review, such as defining thresholds of match confidence scores to identify definite matches, potential matches requiring further review, and definite nonmatches. Establishing the thresholds would require familiarity with the State's specific data, likely gained with a more thorough manual review process similar to the one conducted for this study. However, the upfront work of setting the thresholds would reduce the staff time required in subsequent matching rounds and might make implementation of probabilistic matching more feasible.

[3] Some characteristics of probabilistic matching can be accomplished with data analysis software not specifically designed for probabilistic matching. Staff conducting matching could use data management software to align common spelling variations such as *St.* and *Street*. These programs would not identify likely matched pairs as readily as specially designed probabilistic matching software, however.

APPENDIX A

IN- DEPTH CASE STUDY NSLP DIRECT CERTIFICATION PROFILES

IN-DEPTH CASE STUDY NSLP DIRECT CERTIFICATION PROFILES INTRODUCTION

The In-Depth Case Study NSLP Direct Certification Profiles expand on the information presented in the summary profile by providing additional detail in how direct certification worked in the seven in-depth case study States in SY 2012-2013. The profiles provide narrative descriptions of each State's approach to direct certification; details on the data, systems, and algorithms used in the matching process; the history of the State's direct certification program; plans for future improvement; and strengths and challenges staff reported in the process.

A diagram illustrating each step in the direct certification process follows each narrative description. The flow chart depicts the sequence of events and indicates the agency and district functions in the process. Each flow chart contains a legend identifying the symbols used in the chart. The symbols represent the key steps and system components involved in the process to directly certify school age children for free school meals.

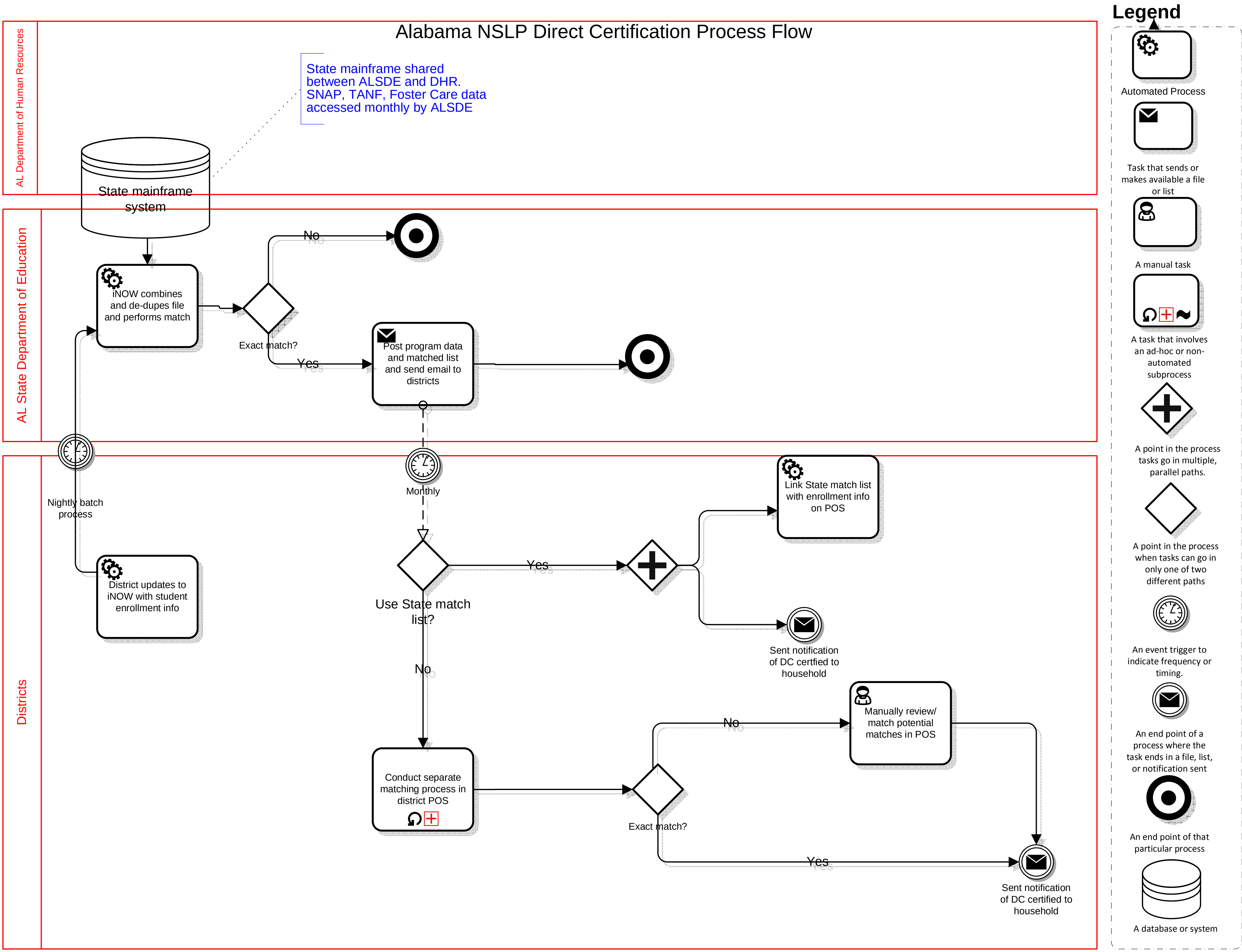
APPENDIX A.1

IN- DEPTH CASE STUDY NSLP DIRECT CERTIFICATION PROFILES

ALABAMA

Table A.1. Profile of Direct Certification Procedures for Alabama, SY 2012–2013

Approach to Matching	Alabama is a central matching State that allows districts great flexibility in how to carry out direct certification. The State Department of Education produces a list of directly certified students and provides it to district child nutrition offices. Districts can either match this list to their local enrollment data or they can match to the State program enrollment data directly.
Timing of match or data distribution	The State provides its matched list to districts monthly and encourages districts to match monthly. Districts may match more frequently during some times in the year.
Use of program participation data and integration with other agencies	The State matches using data from SNAP, TANF, and Foster Care, using data provided by the Department of Human Resources. Staff reported a productive interagency relationship.
Matching algorithms or guidelines	The state's algorithm uses an exact match of the Social Security Number and either the last name or date of birth for direct certification. Districts are permitted to use other algorithms if they choose.
Approach to identifying children from the same household	Districts are responsible for identifying other children in direct certification households.
Transmission procedures for direct certification results or matching data	The Department of Human Resources provides program data to the Department of Education by moving it to a shared location on the state mainframe. The Department of Education makes the matched file available to the districts for download via secure VPN.
History of Direct Certification Process	Alabama successfully piloted an automated process in one school in 1996-1997 that led to statewide implementation of direct certification in 2001. Gradual improvements and grants led to statewide student management system (iNOW) that allowed ALSDE to transition from annual matching to monthly matching in 2010-2011.
Plans for Improving Direct Certification Process	The district plans to update their data systems to push the matched list to the districts every month rather than requiring them to download it.
Strengths of Process	• Recent automation may have improved accuracy of matching. • Strong data security reduces risk to students • Positive interagency relationships help the process run smoothly. • Good communication between Child Nutrition office and IT staff in the Department of Education ensures that data systems meet program needs.
Challenges of Process	Private schools use a manual matching process. The wide variety of point-of-sale systems in use by the districts may lead to variation in direct certification procedures. Respondents expressed data quality concerns and DHR staff suggested that more data sources could be used for direct certification if additional assistance programs used a common definition of poverty.


Alabama NSLP Direct Certification Process Flow
AL Department of Human Resources
AL State Department of Education
Districts
State mainframe shared between ALSDE and DHR. SNAP, TANF, Foster Care data accessed monthly by ALSDE
State mainframe system
iNOW combines and de-dupes file and performs match
Exact match?
No
Yes
Post program data and matched list and send email to districts
Nightly batch process
District updates to iNOW with student enrollment info
Monthly
Use State match list?
Yes
No
Link State match list with enrollment info on POS
Sent notification of DC certfied to household
Conduct separate matching process in district POS
Exact match?
No
Yes
Manually review/ match potential matches in POS
Sent notification of DC certified to household
Legend
Automated Process
Task that sends or makes available a file or list
A manual task
A task that involves an ad-hoc or non-automated subprocess
A point in the process tasks go in multiple, parallel paths.
A point in the process when tasks can go in only one of two different paths
An event trigger to indicate frequency or timing.
An end point of a process where the task ends in a file, list, or notification sent
An end point of that particular process
A database or system

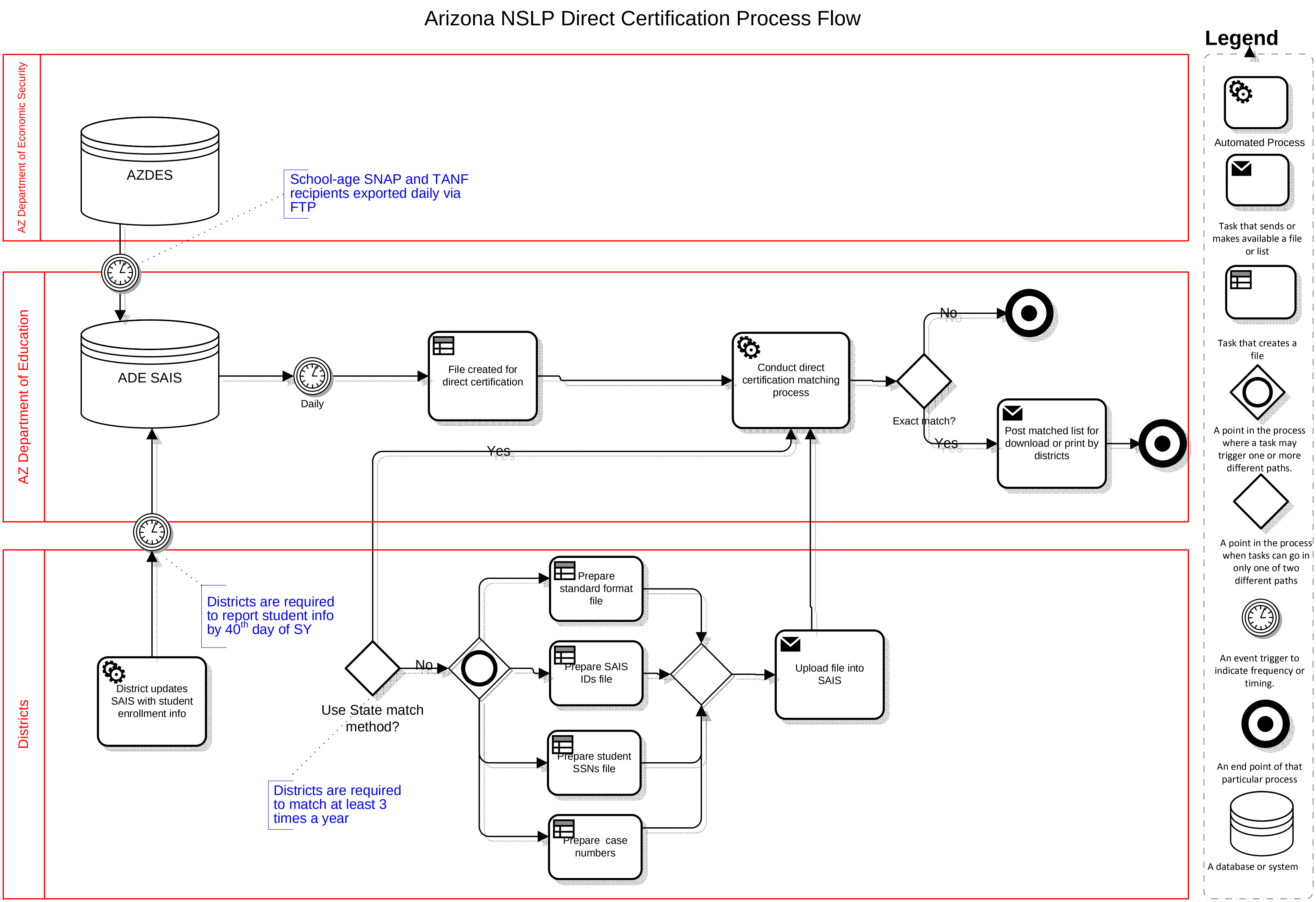
APPENDIX A.2

IN- DEPTH CASE STUDY NSLP DIRECT CERTIFICATION PROFILES

ARIZONA

Table A.2. Profile of Direct Certification Procedures for Arizona, SY 2012–2013

Approach to Matching	The Arizona Department of Economic Security (AZDES) provides SNAP and TANF program data to the Arizona Department of Education (ADE) daily. ADE stores the program data and statewide school enrollment data, and districts logon to the child nutrition web portal and initiate matches using one of five match methods. Districts can query the direct certification system any time in the year to determine the certification status for individual students.
Timing of match or data distribution	ADE requires districts to perform a match at least three times but districts often do more frequent matching. The initial match is performed in September. Districts can look up individual students' direct certification status at any time.
Use of program participation data and integration with other agencies	Arizona uses SNAP and TANF program data for direct certification. AZDES pushes the program data file to ADE daily through an FTP server.
Matching algorithms or guidelines	An exact match on all of the elements (first name, last name, date of birth; or SSN, or student ID; or SNAP/TANF case number) is required for a student to be directly certified regardless of match method used
Approach to identifying children from the same household	The districts are responsible for extending categorical eligibility to students within the same household.
Transmission procedures for direct certification results or matching data	Once a match is complete, districts can download or view match or unmatched results from the central matching system web portal. At any time, districts can pull the direct certification status for individual students by querying the State system.
History of Direct Certification Process	Direct certification began in Arizona in 2003. The State revised the matching system in 2006, creating a more user-friendly process for districts.
Plans for Improving Direct Certification Process	Arizona is considering revising the direct certification matching algorithm and introducing probabilistic matching. ADE is also planning on enhancing the report functionality in the central matching system as well as incorporating Medicaid and possibly foster care data in the near future.
Strengths of Process	The State provides multiple options and flexibility for districts to perform direct certification matching through centralized system. Districts can look up the certification status of students at any time, enabling them to directly certify new and transfer students.
Challenges of Process	Because there is an exact match required for the three elements in order for a student to be directly certified, many potential matches are lost. Additionally, the lack of review process for unmatched or partially matched students limits the direct certification accuracy. High migrant population makes matching eligible kids not registered in the NSLP program problematic. There are a good amount of subgroups of schools that participate in FDPIR, but are not part of the matching process currently. Some issues in the reporting of the FNS-742 data at the district level.



Arizona NSLP Direct Certification Process Flow
AZ Department of Economic Security
AZ Department of Education
Districts
AZDES
School-age SNAP and TANF recipients exported daily via FTP
ADE SAIS
Daily
File created for direct certification
Conduct direct certification matching process
Exact match?
No
Yes
Post matched list for download or print by districts
District updates SAIS with student enrollment info
Districts are required to report student info by 40th day of SY
Use State match method?
Yes
No
Districts are required to match at least 3 times a year
Prepare standard format file
Prepare SAIS IDs file
Prepare student SSNs file
Prepare case numbers
Upload file into SAIS
Legend
Automated Process
Task that sends or makes available a file or list
Task that creates a file
A point in the process where a task may trigger one or more different paths.
A point in the process when tasks can go in only one of two different paths
An event trigger to indicate frequency or timing.
An end point of that particular process
A database or system

APPENDIX A.3

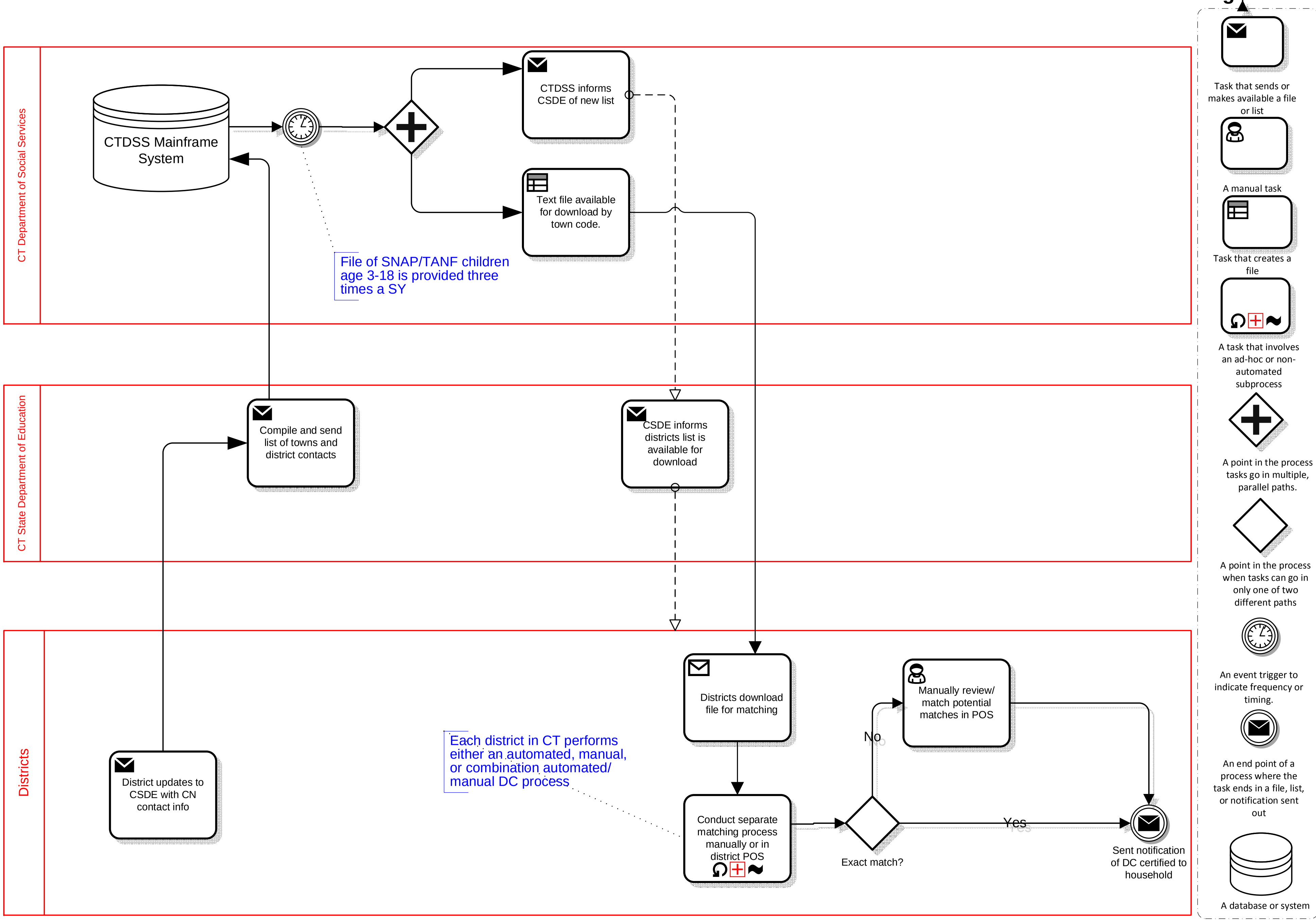
IN- DEPTH CASE STUDY NSLP DIRECT CERTIFICATION PROFILES

CONNECTICUT

Table A.3. Profile of Direct Certification Procedures for Connecticut, SY 2012–2013

Approach to Matching	Connecticut is a local matching state. The Department of Social Services (CTDSS) provides the SNAP and TANF enrollment data to the districts three times per year. Each district matches its local enrollment data against the SNAP and TANF program data to complete direct certification. District procedures vary greatly across the State. Connecticut will transition to a central matching model in fall 2015.
Timing of match or data distribution	CTDSS makes SNAP and TANF program data available to districts three times per year: in August/September, in November/December, and in March.
Use of program participation data and integration with other agencies	Connecticut uses SNAP and TANF program data, both maintained by CTDSS. It is exploring using Foster Care data in the future, which would involve working with the Department of Child and Family Services.
Matching algorithms or guidelines	Procedures vary by district.
Approach to identifying children from the same household	Procedures vary by district.
Transmission procedures for direct certification results or matching data	CTDSS makes the SNAP and TANF program data available to districts on a password-protected website as fixed-length text files.
History of Direct Certification Process	Connecticut has conducted direct certification in some districts since the early 1990s. In the beginning, State staff sent the program data to districts on tapes. More districts gradually began conducting direct certification until 2005, when all districts in the State participated. Districts matched once per year until 2006 when all districts matched three times per year.
Plans for Improving Direct Certification Process	Connecticut plans to transition to a central matching model and increase the frequency of direct certification matching from three times per year to weekly in fall 2015.
Strengths of Process	The strength of Connecticut’s local matching model is that each district is responsible for its own students. Staff reported that they therefore have a particularly strong incentive not to miss any eligible students.
Challenges of Process	The weaknesses of the current local matching model are infrequent matching and inconsistent procedures across the state.

Connecticut NSLP Direct Certification Process Flow



APPENDIX A.4

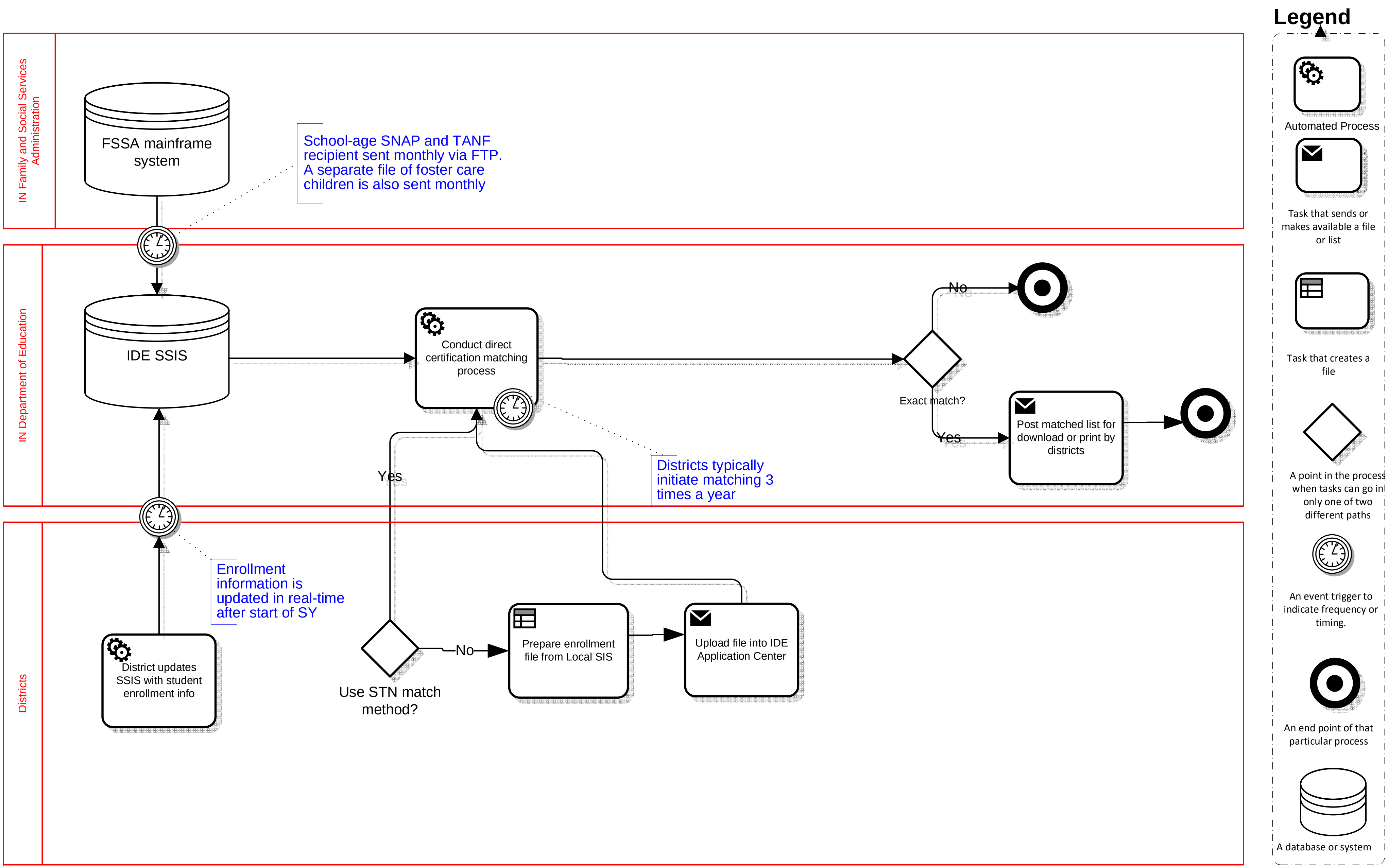
IN- DEPTH CASE STUDY NSLP DIRECT CERTIFICATION PROFILES

INDIANA

Table A.4. Profile of Direct Certification Procedures for Indiana, SY 2012–2013

Approach to Matching	Indiana uses a central matching system and conducts two types of direct certification matching: With the "traditional matching" method, districts upload their local enrollment files to the State's matching tool. The State then matches these local files with State SNAP, TANF, and Foster Care program data to produce lists of matched students. With the "student test number matching" (STN) method, the State draws student enrollment information directly from the statewide student information system, which is updated in real time during the school year. This method is easier, but can only be done during the school year. Therefore, the initial match, which is conducted prior to the start of school each year, uses the traditional matching method. Subsequent matches use the student test number matching method.
Timing of match or data distribution	The initial match is conducted annually prior to the start of school. Program data are updated monthly while student enrollment data is updated in real time during the school year. The State matches these two data sources together monthly, while districts upload the matched data into their local point-of-sale systems at least three times annually. Beginning in SY 2013-2014, monthly matching will be conducted automatically statewide.
Use of program participation data and integration with other agencies	The Indiana Family and Social Services Administration provides monthly data files containing SNAP, TANF, and Foster Care information.
Matching algorithms or guidelines	Indiana directly certifies students with exact matches on first name, last name, date of birth, and county. First and last name matches may be exact matches by spelling or by soundex.
Approach to identifying children from the same household	The State generates a list of unmatched siblings, identified as children in the program data who do not match the enrollment data but who have the same SNAP or TANF case number as a directly certified student. Districts may use this list to extend eligibility.
Transmission procedures for direct certification results or matching data	Districts download the matched list from the State direct certification system as often as monthly. For subsequent matches, districts have the option of downloading the entire district matched list or a list of newly matched students.
History of Direct Certification Process	The direct certification matching algorithm has remained unchanged since it was introduced in the late 1990s.
Plans for Improving Direct Certification Process	Indiana plans to improve the direct certification system so that monthly matches occur automatically. Districts will no longer have to initiate the process manually. The State has also considered introducing probabilistic matching.
Strengths of Process	Direct certification saves staff time. Completing the initial match early and getting notification letters to families quickly can preempt application submissions. Individual student look-up allows districts to certify newly eligible students more quickly and reduce applications.
Challenges of Process	District processes can create a bottleneck in the direct certification system. Even if students are matched efficiently at the State level, they are not certified until districts load the updated information into their point-of-sale systems.

Indiana NSLP Direct Certification Process Flow



APPENDIX A.5

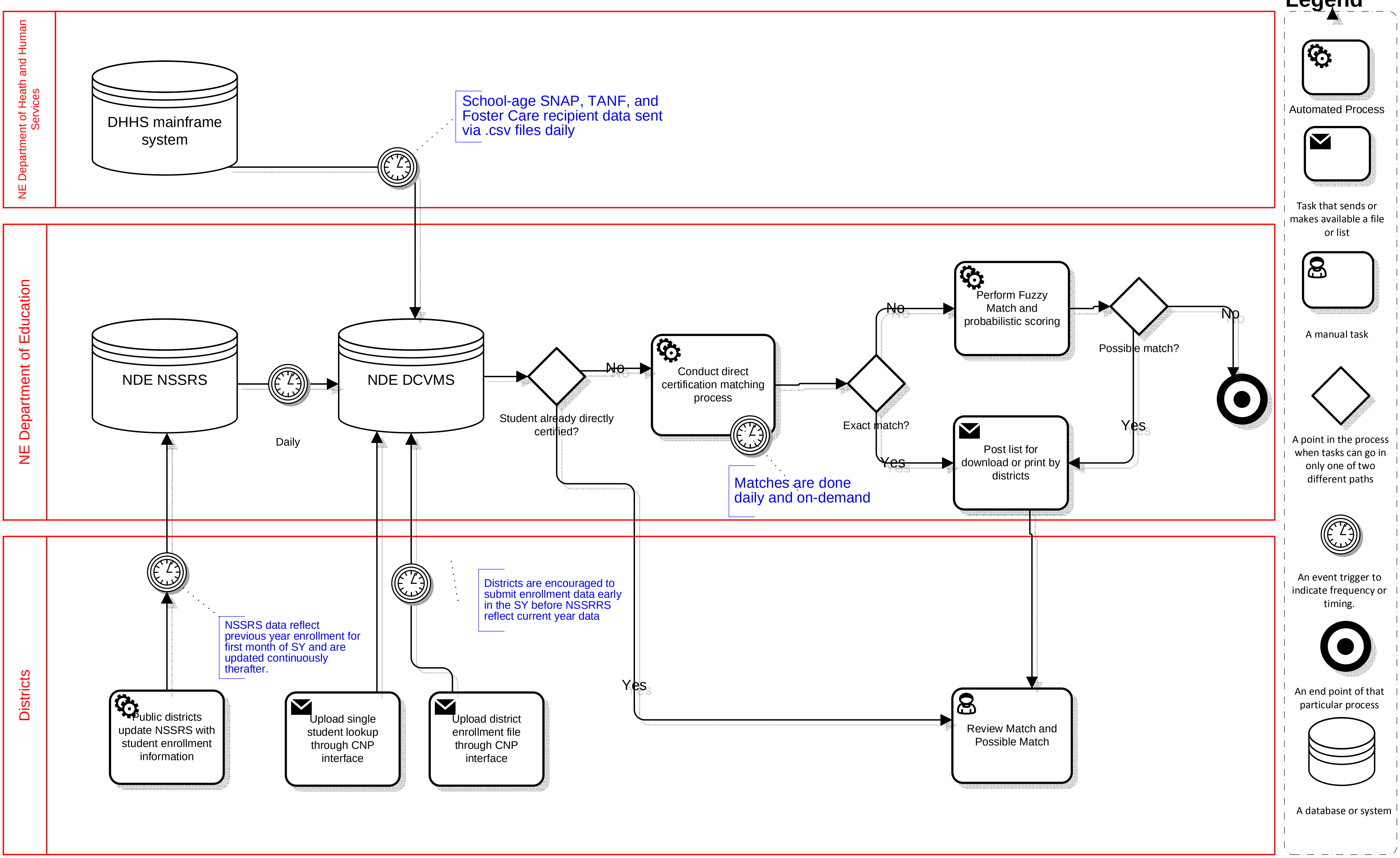
IN- DEPTH CASE STUDY NSLP DIRECT CERTIFICATION PROFILES

NEBRASKA

Table A.5. Profile of Direct Certification Procedures for Nebraska, SY 2012–2013

Approach to Matching	Nebraska uses a central matching system that is based on probabilistic matching of school enrollment data to SNAP, TANF and Foster care data. State Department of Education staff access student enrollment data through the Nebraska Student and Staff Record System (NSSRS). They return lists of definite and possible matches to districts. Districts then investigate possible matches and incorporate matched students into their local student information and POS systems. Districts also have access to an individual student lookup feature that allows for inclusion of student information not available in the State enrollment system.
Timing of match or data distribution	Initial match is conducted before the beginning of each school year with nightly matches conducted throughout the year. Initial matches are not conducted with current enrollment data until September unless districts upload their own enrollment data.
Use of program participation data and integration with other agencies	Nebraska Department of Health and Human Services set up an automated process that provides the Department of Education with daily files of SNAP, TANF, Medicaid and Foster Care participants. This process requires no staff time unless changes are requested. Although establishing an MOU between the relevant agencies was time consuming, both agencies praise the quality of their relationship.
Matching algorithms or guidelines	The main matching algorithm uses four fields: first name, last name, date of birth, and gender. Additional data fields that are not available in the State student enrollment data (but that are included in the State program data) can be used in the individual student lookup feature. The probabilistic matching algorithm was originally based on an internally developed algorithm but was recently switched to Microsoft fuzzy logic to improve accuracy and efficiency.
Approach to identifying children from the same household	Districts are responsible for extending eligibility to children in households receiving SNAP, TANF, or FDPIR. Most districts use POS systems that include electronic matching for extending eligibility.
Transmission procedures for direct certification results or matching data	Districts may download match lists as often as daily and are encouraged to process lists weekly. The State also recommends that districts use the individual student lookup feature whenever there is a new student or transfer.
History of Direct Certification Process	Nebraska received a direct certification grant from FNS in 2009 that was used to develop their web-based probabilistic matching system.
Plans for Improving Direct Certification Process	Nebraska plans to incorporate data on homeless and migrant students into the direct certification process.
Strengths of Process	System was designed to save time for districts, both in processing applications and conducting direct certification. Using a web-based system increases access and allows for user-friendly features. The State believes that daily matching and use of Foster Care data adds substantially to their match rates. Single student lookup is very effective, especially for Nebraska's many small rural schools. Smooth communication with partner agency and automated program data transfer improve efficiency.
Challenges of Process	District technical skill level is often low, which must be mitigated with multiple modes of effective training. Establishing the initial MOU with the Department of Health and Human Services was time consuming.

Nebraska NSLP Direct Certification Process Flow



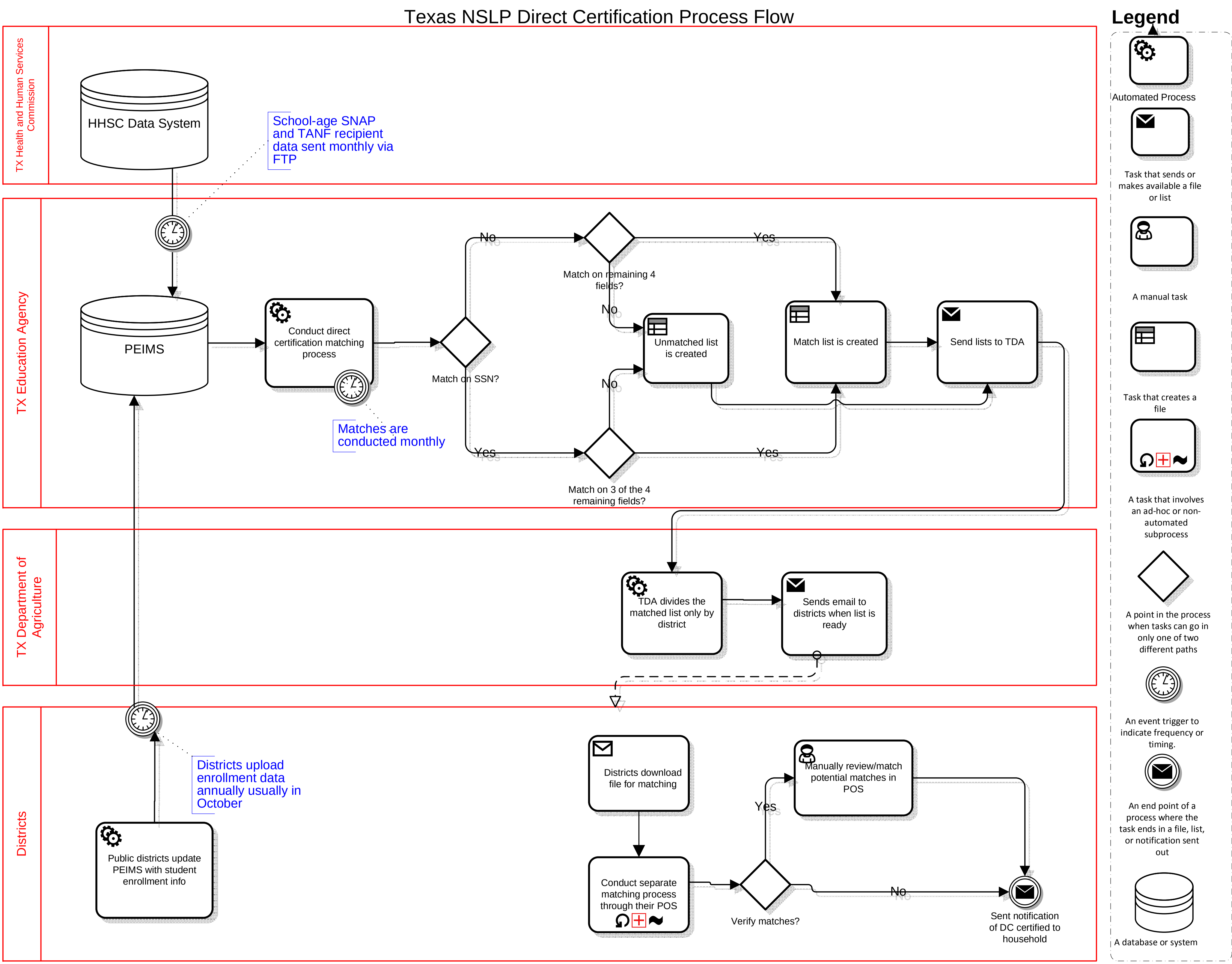
APPENDIX A.6

IN- DEPTH CASE STUDY NSLP DIRECT CERTIFICATION PROFILES

TEXAS

Table A.6. Profile of Direct Certification Procedures for Texas, SY 2012–2013

Approach to Matching	Texas is a central matching State with a fairly limited scope for district activities. State staff match the State enrollment file with SNAP and TANF program data. They then split the resulting matched list by district using the address information in the SNAP and TANF data. Each district receives a list containing only the students that appear to attend schools in that district. District staff then match the state list with their local enrollment files in their point-of-sale systems. Students assigned to the incorrect district's list are not directly certified.
Timing of match or data distribution	The State matches the enrollment data with the SNAP and TANF program data monthly. The SNAP and TANF data are updated monthly; the enrollment data is updated annually each spring and presents a snapshot of enrollment from the previous October.
Use of program participation data and integration with other agencies	The Texas Human Services Commission (HSSC) provides the SNAP and TANF program data for direct certification. The Texas Education Agency (TEA) conducts the matching using statewide enrollment data. The Texas Department of Agriculture (TDA) splits the State list into district-specific lists and makes them available to the districts.
Matching algorithms or guidelines	The Texas Education Agency conducts the matching in two phases. In the first phase, they directly certify students who exactly match on Social Security Number and three of the four other elements: date of birth, first name, last name, or gender. In the second pass, they directly certify students who do not match on Social Security Number but match on all four of the other elements.
Approach to identifying children from the same household	Districts are responsible for identifying children from the same household. They either do this through the statewide student information system (PEIMS) or through their local point-of-sale system.
Transmission procedures for direct certification results or matching data	Districts download the matched lists each month from the TDA secure web portal.
History of Direct Certification Process	Texas has conducted direct certification since the early 1990s. Though the algorithm has remained constant for most of that time, the organizational structure, the matching frequency, and the matching systems have changed. In the beginning, TEA conducted matching annually with assistance from private contractors. Contractors initially used SAS programs in the matching process. In 2004, legislative changes required that TDA assume responsibility for matching. Over time, the matching frequency increased to quarterly and then monthly, and the State transitioned from a SAS-based system to an automated matching system.
Plans for Improving Direct Certification Process	Beginning in SY 2013-2014, TDA will make the entire unmatched list available to districts.
Strengths of Process	• A strong partnership between the State agencies facilitates effective data sharing and problem solving. • High quality IT support keeps systems operating effectively. • Automation improves efficiency of matching process.
Challenges of Process	• Some students end up on the wrong district's list and therefore do not get directly certified. • The statewide student enrollment data is updated only annually and made available on a six-month delay. Therefore, the data are 6 to 17 months out-of-date when used for matching.



Texas NSLP Direct Certification Process Flow
TX Health and Human Services Commission
HHSC Data System
School-age SNAP and TANF recipient data sent monthly via FTP
TX Education Agency
PEIMS
Conduct direct certification matching process
Matches are conducted monthly
Match on SSN?
No
Yes
Match on remaining 4 fields?
Match on 3 of the 4 remaining fields?
Unmatched list is created
Match list is created
Send lists to TDA
TX Department of Agriculture
TDA divides the matched list only by district
Sends email to districts when list is ready
Districts
Public districts update PEIMS with student enrollment info
Districts upload enrollment data annually usually in October
Districts download file for matching
Conduct separate matching process through their POS
Verify matches?
Manually review/match potential matches in POS
Sent notification of DC certified to household
Legend
Automated Process
Task that sends or makes available a file or list
A manual task
Task that creates a file
A task that involves an ad-hoc or non-automated subprocess
A point in the process when tasks can go in only one of two different paths
An event trigger to indicate frequency or timing.
An end point of a process where the task ends in a file, list, or notification sent out
A database or system

APPENDIX A.7

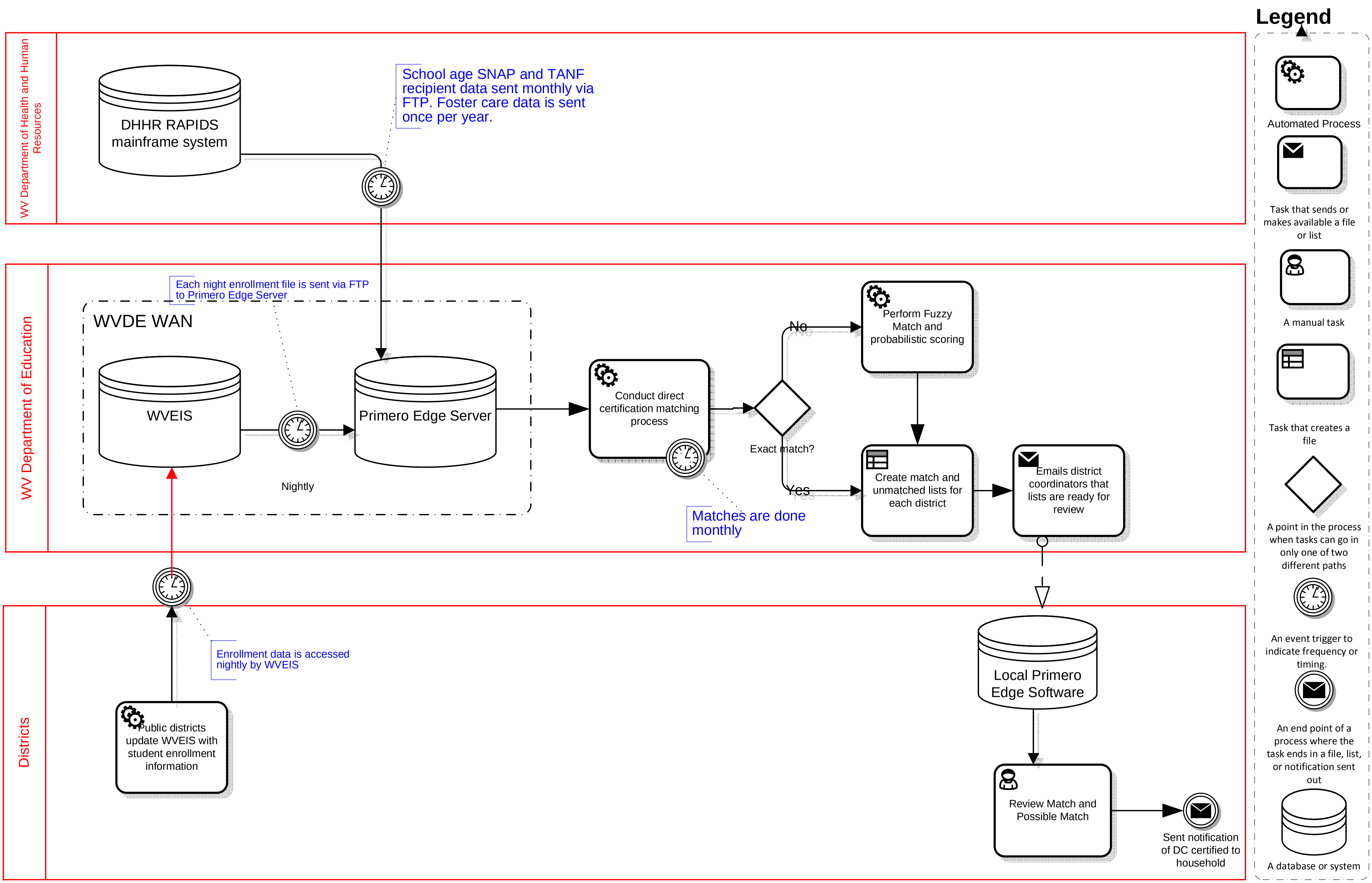
IN- DEPTH CASE STUDY NSLP DIRECT CERTIFICATION PROFILES

WEST VIRGINIA

Table A.7. Profile of Direct Certification Procedures for West Virginia, SY 2012–2013

Approach to Matching	West Virginia is a central matching state in which the State Department of Education (WVDE) matches program data against the statewide school enrollment data and makes matched, unmatched, and partially matched lists available to each district.
Timing of match or data distribution	Matching occurs daily. School enrollment data are updated in real time. SNAP and TANF data are updated monthly, following the second Saturday in each month. Foster Care data are updated annually.
Use of program participation data and integration with other agencies	The Department of Health and Human Resources provides SNAP and TANF data monthly and Foster Care data annually to the WVDE for direct certification matching.
Matching algorithms or guidelines	WVDE directly certifies students who exactly match on Social Security Number or an exact match on first name, last name, and date of birth. Name matches can be by spelling or phonetically through soundex algorithms.
Approach to identifying children from the same household	Districts identify other members of direct certification households by matching on home address. Districts can also identify these individuals by referencing applications from previous years.
Transmission procedures for direct certification results or matching data	Districts can view matched and partially/unmatched listing of students through the Primero Edge system.
History of Direct Certification Process	West Virginia began using SNAP and TANF data for direct certification in 2004. Each district initially operated different point-of-sale systems. However, around 2007, the State hired a private vendor to operate a central point-of-sale system (Primero Edge) for the entire state. Now all public schools—and most private schools—use the same system statewide.
Plans for Improving Direct Certification Process	West Virginia plans to transition to semi-monthly or even weekly matching. The State also plans to introduce a continuous direct certification training program and to incorporate private schools into the system more fully. The State also plans to invest additional resources to improve its system infrastructure to make the system more reliable and faster and to expand its bandwidth.
Strengths of Process	The primary advantage of West Virginia's central model is that State staff have access to data from all districts. System automation allows accurate and timely matching. Strong interdepartmental relationships help the system run smoothly.
Challenges of Process	Bandwidth limitations impede system performance during peak times.

West Virginia NSLP Direct Certification Process Flow



MATHEMATICA
Policy Research

www.mathematica-mpr.com

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

Princeton, NJ ■ Ann Arbor, MI ■ Cambridge, MA ■ Chicago, IL ■ Oakland, CA ■ Washington, DC