U.S. DEPARTMENT OF AGRICULTURE

Understanding Risk Assessment in Supplemental Nutrition Assistance Program Payment Accuracy

December 2025

Nondiscrimination Statement

In accordance with federal civil rights law and USDA civil rights regulations and policies, the USDA, its agencies, offices, employees, and institutions participating in or administering USDA programs are prohibited from discriminating based on race, color, national origin, religion, sex, disability, age, marital status, family/parental status, income derived from a public assistance program, political beliefs, or reprisal or retaliation for prior civil rights activity, in any program or activity conducted or funded by USDA (not all bases apply to all programs). Remedies and complaint filing deadlines vary by program or incident.

Persons with disabilities who require alternative means of communication for program information (e.g., Braille, large print, audiotape, American Sign Language, etc.) should contact the state or local agency that administers the program or contact USDA through the Telecommunications Relay Service at 711 (voice and TTY). Additionally, program information may be made available in languages other than English.

To file a program discrimination complaint, complete the USDA Program Discrimination Complaint Form, AD-3027, found online at How to File a Program Discrimination Complaint and at any USDA office or write a letter addressed to USDA and provide in the letter all of the information requested in the form. To request a copy of the complaint form, call (866) 632-9992. Submit your completed form or letter to USDA by:

1. **Mail:** USDA Food and Nutrition Service, 1320 Braddock Place, Room 334, Alexandria, VA 22314; or
2. **Email:** FNSCIVILRIGHTSCOMPLAINTS@usda.gov.

USDA is an equal opportunity provider, employer, and lender.

U.S. DEPARTMENT OF AGRICULTURE

Understanding Risk Assessment in Supplemental Nutrition Assistance Program Payment Accuracy

Final Report

December 2025

Contract: GS10F0136X

Order: 140D04221007

Authors

Betsy Thorn, Kevin Baier, Jacob Beckerman-Hsu, Lindsay Giesen, Kelley Calvin, James McCall, Julia Esposito, Nile Campbell, and Sarah Chance

Submitted to

USDA Food and Nutrition Service
Office of Evaluation, Analysis, and Regulatory Affairs
1320 Braddock Place
Alexandria, VA 22314

Project Officer: Kameron Burt

Submitted by

Westat Insight
1310 North Courthouse Road
Suite 880
Arlington, VA 22201

Project Director: Betsy Thorn

This study was conducted by Westat Insight under Contract No. GS10F0136X/140D04221007 with the U.S. Department of Agriculture's (USDA) Food and Nutrition Service. The findings and conclusions in this report are those of the authors and should not be construed to represent any official USDA or U.S. Government determination or policy.

Suggested Citation

Thorn, B., Baier, K., Beckerman-Hsu, J., Giesen, L., Calvin, K., McCall, J., Esposito, J., Campbell, N., & Chance, S. (2025). *Understanding risk assessment in Supplemental Nutrition Assistance Program payment accuracy*. Westat Insight. U.S. Department of Agriculture, Food and Nutrition Service.

Acknowledgments

This report was prepared by Westat Insight under Contract Number GS-10-F0136X/ 140D04221007 from the U.S. Department of Agriculture's Food and Nutrition Service (FNS). It represents a team effort, with many individuals making important contributions, and we gratefully acknowledge their assistance. We recognize the sustained leadership of Kameron Burt, social science research analyst and FNS Project Officer. Many other people in the FNS National Office and Regional Offices also contributed. We thank the following individuals for their time and assistance: Shawnee Anderson, John McCleskey, Anita Singh, and Amanda Wyant. We also thank the staff of the 44 State agencies that participated in the study's survey, especially the staff in Connecticut, Kansas, Rhode Island, Utah, Virginia, and Wisconsin SNAP State agencies, for their time and contributions to this study. This work could not have been done without their input.

The authors thank a panel of former State agency leaders who reviewed study materials, including Arlene Dura, Sherry Haar, Tom Hedderman, Taci Lawson, Darnell Mascarenas, and Patricia McGinn.

Finally, many Westat Insight and Westat colleagues made important contributions to this project. We thank Maeve Gearing for her leadership as qualitative lead. We are grateful to Kelsey Gray, Brian Jones, Mustafa Karakas, John Knaus, and Michael Ribar for their thoughtful feedback and quality assurance review. We thank Holly Alexander, Katya Scanlan, Chantell Atere, and Kim Kerson for their editorial and production support.

Contents

Executive Summary 1

A. Overview of SNAP RA Tool Implementation and Use 1

B. Assessment of RA Tool Effectiveness 3

C. Identification of Promising Practices 6

Chapter 1. Introduction 7

A. Examples of RA Tools in Human Services Programs 8

B. SNAP Administration and SNAP Quality Control 9

C. Study Objectives and Overview 11

D. Limitations and Considerations 14

Chapter 2. Development and Implementation of SNAP RA Tools 15

A. Variation in Tool Use Across State Agencies 15

B. RA Tool Development 17

C. Variables Used in RA Tools 20

D. How State Agencies Use RA Tool Results 21

E. Reasons Some State Agencies Do Not Currently Use RA Tools 25

Chapter 3. Case Studies of Six State Agencies' RA Tools 27

A. Connecticut's Case Review List 28

B. Kansas' Pre-Authorization Screening Checklist 32

C. Rhode Island's Pre-Authorization Review Tool 35

D. Utah's Performance Review Case-Profiling Tool 39

E. Virginia's SNAP Cross Check and Discrepancy Tool 42

F. Wisconsin's Pre-Certification Review Tool 46

Chapter 4. RA Tool Effectiveness 53

A. Evaluating Tool Performance in Selecting Cases for Review 54

B. Evaluating RA Tool's Effects on the Payment Error Rate 58

Contents

Chapter 5. RA Tool Promising Practices 63

A. Robust Data Analysis 63

B. Collaborative Design Process With Technical and Programmatic Perspectives 64

C. Transparent Programming and Design 65

D. Performance Testing Before Implementation 66

E. Continuous Monitoring of Tool Performance 66

F. In-Depth and Ongoing Training for End Users 67

References 68

Appendix A. Detailed Methods A-1

A. Literature Review Methodology A-1

B. Survey Development, Fielding, and Results A-2

C. Case Studies A-8

D. RA Tool Effectiveness Methodologies A-9

Appendix B. Data Collection Instruments B-1

Appendix C. Supplemental Tables C-1

Appendix D. Research Objectives and Questions Crosswalk D-1

Appendix E. Considerations for Developing, Using, and Monitoring the Performance of RA Tools E-1

A. Key Questions to Ask When Developing an RA Tool E-1

B. Key Considerations for Monitoring and Evaluating an RA Tool E-10

Contents

Tables

Table 3.1. Sample of potential errors flagged by the SNAP Cross Check and Discrepancy Tool 43

Table 4.1. Confusion matrix with definitions for SNAP RA tools 54

Table 4.2. Confusion matrix for Kansas' Pre-Authorization Screening Checklist 55

Table 4.3. Confusion matrix for Missouri's FNS Error Prone Trends 56

Table 4.4. Confusion matrix for Rhode Island's Pre-Auth Review 58

Table 4.5. Median PERs October 2014 to December 2019 59

Table 4.6. Run chart analysis in Minnesota 59

Table 4.7. Impact of Minnesota's RA tool on its PERs 61

Table A.1. Pretest respondents A-2

Table A.2. Recruitment schedule A-3

Table A.3. Survey response rate by FNS Region A-4

Table A.4. Survey response rate by SNAP caseload size A-4

Table A.5. Average survey question completion rate by survey screener group A-5

Table A.6. Survey question completion rate for 44 responding State agencies A-6

Table A.7. Survey nonresponse bias analysis results A-7

Table A.8. Interviews conducted by case study State agency A-9

Table A.9. Impact of Minnesota's RA tool on their PERs, sensitivity analysis models A-15

Table C.1. State agency use of RA tools C-1

Table C.2. Adequate funding for RA tool C-3

Table C.3. Dates of use for SNAP RA tools C-4

Table C.4. Motivations for developing RA tools currently in use C-5

Table C.5. Data analyzed to determine high risk of payment error C-6

Contents

Table C.6. Data analysis used to develop SNAP RA tools C-7

Table C.7. Developers of RA tools currently in use C-8

Table C.8. State agency involvement in tools currently in use developed by vendors/contractors C-9

Table C.9. Testers of implemented RA tools currently in use C-10

Table C.10. Number of State agencies offering each type of training C-11

Table C.11. Factors considered and included in currently active RA tools C-12

Table C.12. Variables used in statewide RA tools C-14

Table C.13. Formats of RA tools currently in use C-15

Table C.14. State agency allows local agencies to customize criteria used by RA tool C-16

Table C.15. Time period when current tools are used to identify cases at risk for payment error C-17

Table C.16. SNAP case file data used by active tools C-18

Table C.17. Staff who use RA tool to flag cases at risk for payment error C-19

Table C.18. Ways State agencies take action on cases flagged as at risk for payment error C-20

Table C.19. Staff who follow up on cases at risk for payment error based on RA tool results C-21

Table C.20. State agency use of RA tool data for other purposes C-22

Table C.21. State agency assessment of impact on error rates C-23

Table C.22. Frequency of effectiveness evaluations of tools currently in use C-24

Table C.23. Reason developed RA tools were not implemented C-24

Table C.24. State agency consideration of RA tool development C-25

Table C.25. Reasons RA tools are no longer in use C-25

Table D.1. Understanding risk assessment in SNAP payment accuracy crosswalk of research questions and data collection instruments D-2

Contents

Table E.1. Effects of an example RA tool used after the State agency finishes collecting information from households for SNAP applications E-6

Table E.2. Example confusion matrix for a proposed RA tool E-8

Table E.3. Example confusion matrix for a proposed RA tool stratified by the presence of children in the household E-9

Figures

Figure ES.1. RA tool status by State agency 2

Figure ES.2. Sensitivity and specificity for RA tools in three State agencies 4

Figure ES.3. Minnesota's quarterly payment error rates, before and after tool implementation 5

Figure 2.1. RA survey results by State agency 16

Figure 2.2. SNAP RA tool implementation timeline 17

Figure 2.3. Motivations for developing RA tools currently in use 18

Figure 2.4. Data analyzed to determine high risk of payment error 18

Figure 2.5. Developers and testers of SNAP RA tools 19

Figure 2.6. State agency involvement in RA tools currently in use developed by vendors or contractors 20

Figure 2.7. Variable categories considered and included in RA tools by State agencies 21

Figure 2.8. When State agencies apply RA tools to SNAP cases 23

Figure 2.9. Steps State agencies take after their RA tool flags a case at risk for payment error 24

Figure 2.10. Frequency of RA tool effectiveness evaluations 25

Figure 3.1. Use of Connecticut's Case Review List 29

Figure 3.2. Screenshot of sample fields in the Case Review List 30

Figure 3.3. Use of Kansas' Pre-Authorization Screening Checklist 33

Contents

Figure 3.4. Rhode Island's pilot of the Pre-Auth Tool 36

Figure 3.5. Use of Utah's Performance Review Case-Profiling Tool..... 40

Figure 3.6. Use of Virginia's SNAP Cross Check and Discrepancy Tool 43

Figure 3.7. Use of Wisconsin's Pre-Certification Review Tool..... 47

Figure 3.8. Sample dashboard on Pre-Certification Review Tool effectiveness 50

Figure 3.9. Sample dashboard on errors found with Pre-Certification Review Tool 51

Figure 4.1. Minnesota's quarterly payment error rates, October 2014 (FY 2015, quarter 1) to December 2019 (FY 2020, quarter 1)..... 60

Figure A.1. Example of run chart analysis A-10

Figure A.2. Quarterly payment error rates, October 2014 (FY 2015, quarter 1) to December 2019 (FY 2020, quarter 1)..... A-17

Executive Summary

Introduction

Risk assessment (RA) tools offer human services agencies the ability to allocate resources more efficiently by ensuring cases with the highest levels of risk receive a greater degree of staff time and attention. As advanced analytic platforms and methods become more accessible, more agencies are developing these tools for program administration. This report provides an overview of the RA tools currently used by the State agencies that administer the Supplemental Nutrition Assistance Program (SNAP). It assesses the effectiveness of a subset of those tools and identifies best practices in RA tool development, implementation, and evaluation. To support this effort, the study team reviewed the literature, surveyed all 53 SNAP State agencies, interviewed State and local staff in six State agencies, and analyzed SNAP Quality Control (QC) data for 10 State agencies.

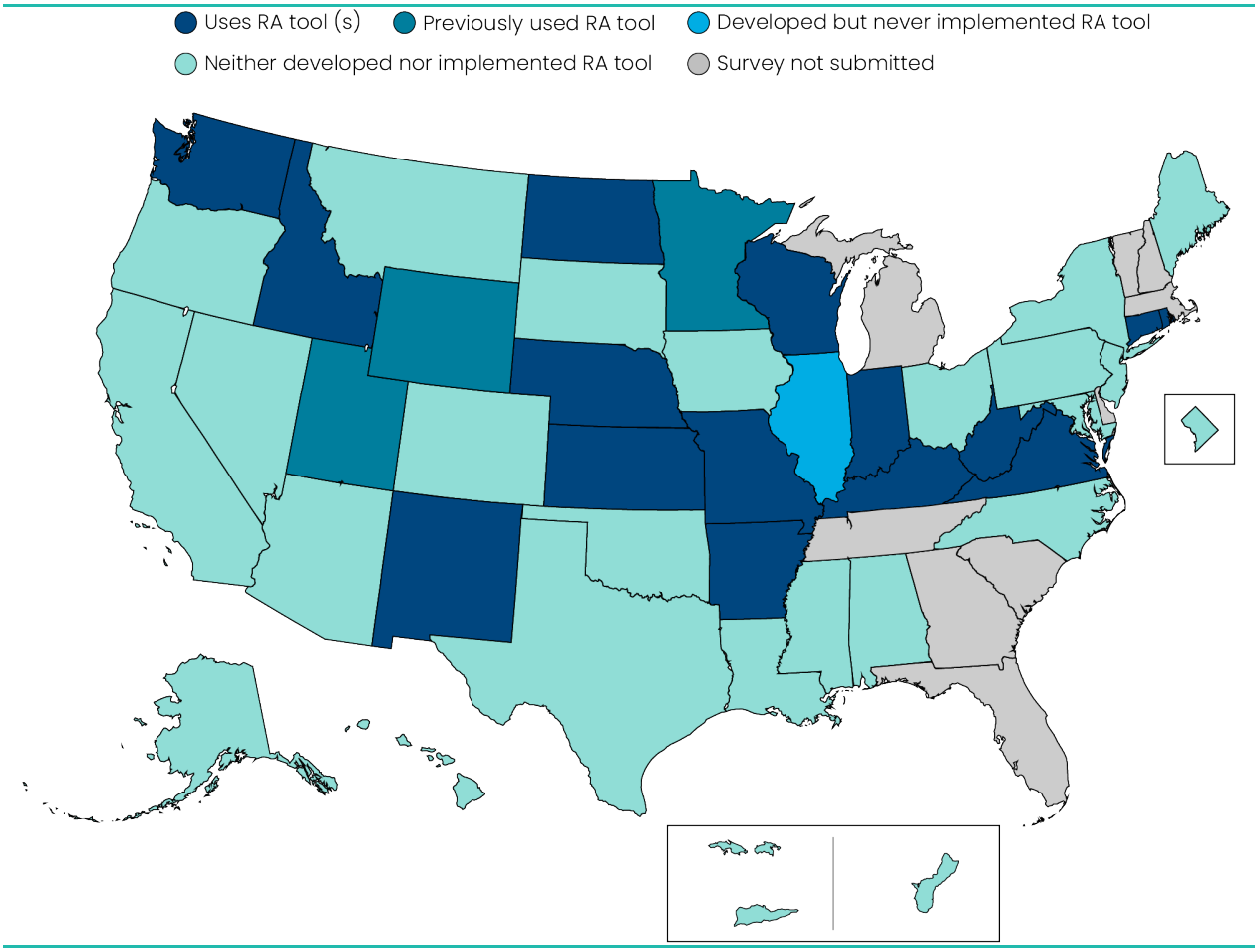
A. Overview of SNAP RA Tool Implementation and Use

Of the 43 State agencies and one local agency[1] that responded to the SNAP Risk Assessment Study Online Survey—

- Fifteen State agencies currently use an RA tool (34 percent; figure ES.1).
- Three State agencies previously used an RA tool but have since discontinued its use (7 percent).
- One State agency has developed but not yet implemented an RA tool (2 percent).
- Twenty-four State agencies and one local agency have neither developed nor implemented an RA tool (57 percent).

[1] New York City responded instead of New York State.

Figure ES.1. RA tool status by State agency



Note: Data for Figure ES.1 is available in Table C.1 in Appendix C.
Note: New York City responded instead of New York State.
Data source: SNAP Risk Assessment Study Online Survey

Among the 15 State agencies that use RA tools, most developed the tools to help address high payment error rates (PERs) (10 State agencies) and/or to help them concentrate their available resources on suspected high-risk cases (7 State agencies). One State agency used machine learning, and seven used descriptive statistics as part of the RA tool development process. Other State agencies were uncertain about whether their agency conducted substantial statistical analysis specifically for RA tool development, and some relied on previous efforts to identify error-prone case characteristics as part of the SNAP QC or quality assurance review process. State agencies typically drew upon SNAP QC data (12 State agencies) or other case file data (10 State agencies) for their findings. Seven State agencies engaged a contractor to develop their RA tool. Interview respondents described mixed experiences working with contractors: Some State agency staff wanted greater insight into tool specifications to ensure alignment with SNAP policy.

The RA tools that SNAP State agencies use come in several formats. Eight State agencies described their tool as written instructions or a checklist. Kansas, for example, uses a checklist in Microsoft Word to instruct frontline eligibility workers to request a secondary review from an

eligibility worker supervisor if the case meets all of several criteria related to household size, income levels, and benefit amount. Six State agencies have an RA tool programmed into their eligibility system or other database, and three describe their systems as taking some other format. In 11 State agencies, local agencies or counties have the flexibility to customize the parameters of the RA tool. Wisconsin, for example, allows consortia of local offices to select which, if any, criteria to use to identify high-risk cases; the State agency also allows consortia to place upper bounds on the number of high-risk cases to help manage workloads. The most common criteria used by RA tools are presence of earned income (10 State agencies), presence of self-employment income (10 State agencies), household size (9 State agencies), and presence of unearned income (8 State agencies).

RA tools are incorporated into many steps in SNAP administration, but they are most commonly used after new information on a case has been received but before benefits have been issued. Some State agencies use their tools after benefit issuance. For example, Connecticut uses an RA tool to select cases for quality assurance review at the end of the month, after benefits have been issued. Interview respondents noted a tradeoff between the preference to use RA tools before benefit issuance to improve benefit calculation accuracy and the need to limit the steps needed before issuance to maintain application processing timeliness standards. Cases flagged by RA tools most frequently undergo a secondary review or a quality assurance review.

B. Assessment of RA Tool Effectiveness

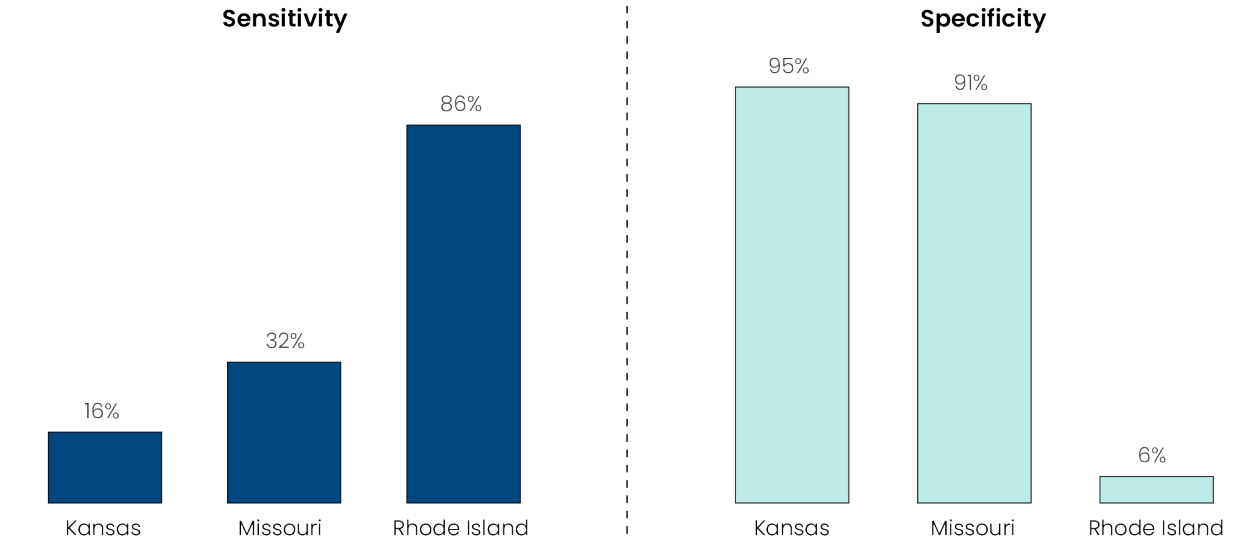
The study team took two approaches to evaluating RA tools. The first was to assess whether the tools can accurately and efficiently identify cases with payment errors. The second was to examine whether RA tool implementation is associated with a lower PER.

For the first approach, the study team assessed RA tools for three State agencies. The team determined which cases the tools would have flagged as high risk or low risk by applying the tool's parameters to all cases in the SNAP QC Data File for the fiscal year prior to implementation of the tool. The team then compared these flags with whether the cases truly had payment errors as determined through the SNAP QC process. Using these results, the study team calculated two measures of performance for each tool:

- Sensitivity—the proportion of cases with a true payment error that are flagged as high risk
- Specificity—the proportion of cases without a payment error that are flagged as low risk

Figure ES.2 presents sensitivity and specificity results for each of the three State agencies.

Figure ES.2. Sensitivity and specificity for RA tools in three State agencies



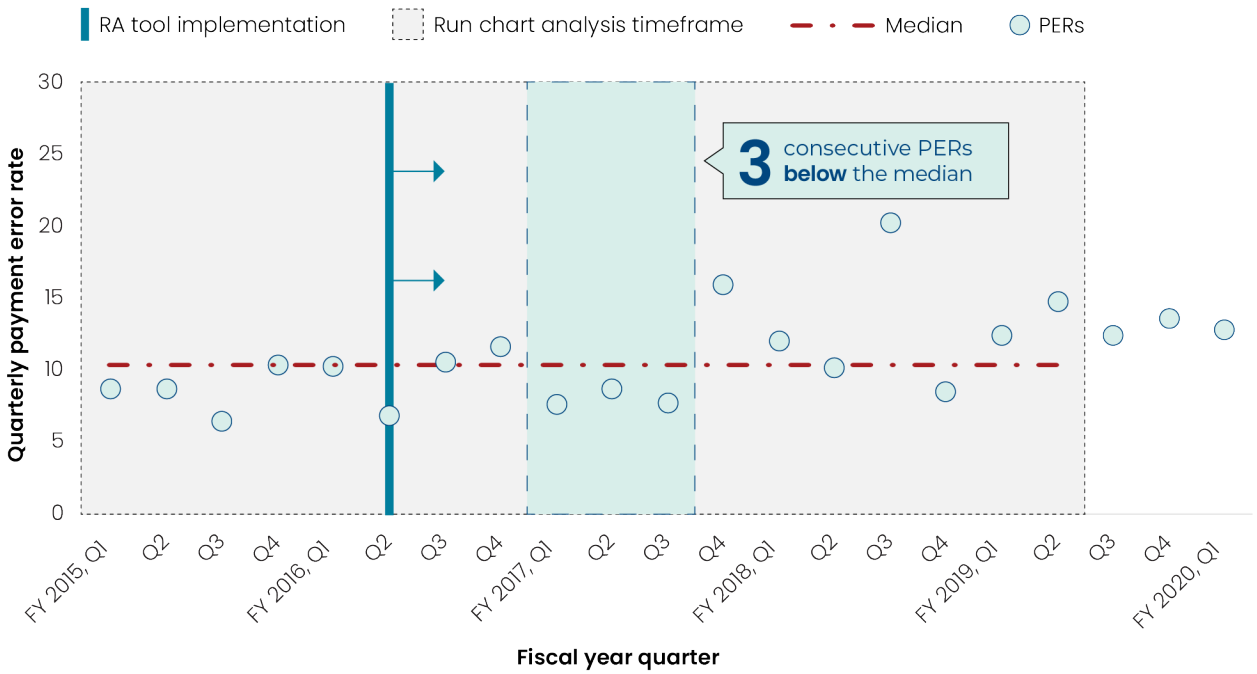
Data source: SNAP Quality Control data

Kansas and Missouri both have tools with high specificity (95 and 91 percent), meaning that the tools rarely identify cases without a payment error as high risk. However, both State agencies' tools have low sensitivity, meaning many cases with payment errors are not identified as high risk. Rhode Island's tool shows the opposite pattern; with high sensitivity, it does not miss many cases with payment errors, but its low specificity reflects that many cases without payment errors are flagged as high risk.

Ideally, implementing an RA tool leads to fewer cases with payment errors. The second evaluation approach the study team used was to test the association between RA tool implementation and State agency PERs. Due to certain SNAP QC data limitations, only one State agency's tool (Minnesota's) could be evaluated. Minnesota implemented its tool in January 2016. The team examined PERs for Minnesota and six comparison State agencies[2] that did not have a tool from October 2014 to December 2019. A run chart analysis (figure ES.3) found a streak of three consecutive quarters with PERs below the median (across the timeframe of the analysis); this was below the study team's preestablished metric of success of eight consecutive quarters.

[2] The comparison group comprises Arizona, Guam, North Dakota, New York, South Dakota, and the U.S. Virgin Islands.

Figure ES.3. Minnesota's quarterly payment error rates, before and after tool implementation



Note: Given the SNAP Quality Control data challenges prior to FY 2015, the run chart analysis timeframe includes the quarter of RA tool implementation and five quarters before and 12 quarters after the quarter of RA tool implementation. A shift occurs when the quarterly PER is below the median quarterly payment error of the run chart analysis timeframe for eight or more consecutive quarters after the quarter of RA tool implementation. Minnesota did not experience a shift.
FY = fiscal year; Q = quarter; PER = payment error rate
Sources: SNAP Quality Control data; SNAP Risk Assessment Study Online Survey

The study team built on the run chart findings by estimating a difference-in-differences model. This type of model assumes that trends in the outcome measures (in this case, quarterly PERs) would have continued on the same trajectory if a treatment (implementation of an RA tool) did not occur. The study team used data from Minnesota and the six comparison State agencies to estimate the effect of the RA tool implementation on PERs in Minnesota. The difference-in-differences model found a statistically significant decrease of 1.4 percentage points in the PER in Minnesota following the implementation of the RA tool.

C. Identification of Promising Practices

Drawing on the literature review, response to the SNAP Risk Assessment Study Online Survey, key informant interviews, and SNAP QC Data analysis, the study team identified six promising practices for SNAP to consider when developing and implementing RA tools.

1. RA tool development should begin with **robust data analysis**. Agencies should use the data available to them to identify the case characteristics most associated with negative outcomes and continue periodic data analysis after implementation because the root causes of negative outcomes may change.
2. RA tool development should include a **range of technical and programmatic perspectives**, such as policy experts to ensure RA tool parameters are consistent with program policy, data analysts or statisticians to ensure rigorous data analysis, IT staff for tool programming and integration with other systems, and future end users to ensure the tool is appropriately integrated into existing procedures.
3. RA tools should use **transparent programming and design**. Broad understanding of the tools' underlying parameters and algorithms facilitates effective use and appropriate corrections or updates.
4. RA tool developers should **test the tool's performance** before implementation. Testing should ensure that the tool appropriately identifies high-risk cases and can be effectively incorporated into end users' workflows.
5. After implementation, RA tools should undergo **continuous performance monitoring**. RA tool developers should plan to track outcomes and metrics to support performance monitoring. Agencies that use RA tools should routinely review trends in outcomes and revise tool parameters or procedures as necessary.
6. RA tool users should receive **in-depth and ongoing training**. Agency staff should be clearly trained in how and why to use the tool. Many agencies may also use tool outcomes (e.g., error trends) to identify additional training needs for staff.

Chapter 1

Introduction

Ensuring Federal funds are disbursed appropriately is a key function of the U.S. Department of Agriculture's (USDA) Food and Nutrition Service (FNS) and the State agencies that administer the Supplemental Nutrition Assistance Program (SNAP). With the oversight of FNS, State agencies are responsible for ensuring that only eligible recipients receive SNAP and in the correct amount. Despite preventive efforts, some ineligible households receive benefits, and some eligible households receive benefits in incorrect amounts. To address this challenge, some State agencies use risk assessment (RA) tools. These tools are designed to identify which cases are likely to have a payment error to help State agencies direct their efforts to improve accuracy where those efforts are most needed. This report describes how SNAP State agencies use RA tools, assesses the effectiveness of the tools in selecting cases for review and increasing payment accuracy, and identifies promising practices in RA tool development and implementation.

For the purposes of this research, the study team used three criteria to define a SNAP RA tool:

1. The tool is used during the application or recertification process or on active cases.[3]
2. The tool identifies the risk of a case having a payment error (e.g., high/low risk, high/medium/low risk).
3. Cases may be treated differently according to their risk category (e.g., high-risk applications are subject to additional scrutiny before benefit determination, high-risk active cases are prioritized for a quality assurance review).

[3] Active cases are cases that were issued SNAP benefits during the relevant timeframe.

A. Examples of RA Tools in Human Services Programs

Many human services agencies at the Federal, State, and local levels employ RA tools, which draw on large volumes of data and advanced analytic approaches to help organizations make decisions about resource allocation. RA tool examples include the following:

- The Oregon Audits Division uses predictive models to detect SNAP fraud. This effort relies in part on data matching but also uses analysis of SNAP transaction patterns to identify retailers and participants that might be violating program regulations. The findings of this analysis enable Oregon to prioritize retailers and participants who had been flagged as high risk for investigation (Oregon Audits Division, 2018).
- Allegheny County, Pennsylvania, developed a tool to predict the likelihood that a child referred to the child welfare system would be placed in foster care. The Allegheny County Family Screening Tool helps workers decide which cases referred through the county hotline should be subject to investigation. An evaluation determined the tool led to changes in the types of children whose cases were screened for investigation and reduced disparities in case-opening rates between Black and White children, without changing the overall workload for workers (National Institute of Justice, 2021).
- The Prisoner Assessment Tool Targeting Estimated Risk and Needs predicts the risk of recidivism for every person in the custody of the Federal Bureau of Prisons (National Institute of Justice, 2021). The Department of Justice developed this tool to help determine who is eligible for early release.

RA tools can take many forms. At their core, they use information about a case to estimate the chance a case has an error, which in turn guides how the agency handles the case. An example of a simple RA tool is a checklist of case characteristics associated with payment errors (e.g., large household size, multiple sources of income). The more of these characteristics a given case has, the higher the likelihood of a payment error. The agency could then take additional action on cases with a certain number of characteristics associated with risk of payment error. A more complex RA tool could employ machine learning algorithms to automatically assess case data entered into a State agency's eligibility system. The tool could then instruct eligibility workers on the correct course of action—immediate approval for benefit issuance or secondary review—according to the case's estimated risk.

B. SNAP Administration and SNAP Quality Control

1. SNAP Administration

To receive SNAP benefits, households[4] must apply to the State agency[5] in the State where they currently live and meet certain eligibility requirements.[6] Applications can typically be found on State agencies' websites or at local SNAP offices and are used to determine a household's eligibility. Whereas eligibility for SNAP benefits can vary by State agency because of various administrative options and waivers of statutory and regulatory requirements, households that meet Federal income[7, 8] guidelines are often eligible. Under Federal guidelines, households must meet the following criteria:

- Gross monthly income must be at or below 130 percent of the Federal poverty guidelines.
- Net income must be at or below the poverty guidelines.
- Assets[9] must fall below certain limits.

Households that do not meet these criteria are often not eligible to receive SNAP benefits. Households with individuals without documented immigration status, some students attending higher education institutions, and select formerly incarcerated individuals are not eligible to receive SNAP benefits.[10]

After a household submits a SNAP application, the State agency or local SNAP office processes the application and makes a benefit determination within 30 days.[11] During the 30-day processing period, households are required to complete an eligibility interview to provide verification of the information stated in their application. During the interview, SNAP eligibility workers accomplish the following tasks:[12]

- Discuss aspects of the application that the household did not complete.
- Screen household members for exemptions from certain SNAP rules.
- Identify and resolve questionable or inconsistent information.
- Explain program rules to the household and answer questions.
- Explain household rights and responsibilities.

[4] In SNAP, households are considered a group of people who live together and purchase and prepare meals together.
[5] State agency refers to the SNAP programs in the 50 States, the District of Columbia, the U.S. Virgin Islands, and Guam.
[6] For more information, see FNS's SNAP Eligibility website.
[7] SNAP gross and net incomes limits are higher in Alaska and Hawaii.
[8] SNAP considers income from all sources, including earned income, unearned income such as Social Security benefits, and child support.
[9] SNAP considers assets to be resources that could be available to households to purchase food (e.g., bank accounts).
[10] For more information, see A Quick Guide to SNAP Eligibility and Benefits from the Center on Budget and Policy Priorities.
[11] SNAP households that qualify for "expedited service" should have benefit determinations made within seven days of their application date compared with the standard 30-day processing period.
[12] For more information, see FNS's State SNAP Interview Toolkit.

- Identify all verification requirements and assist the household in obtaining needed verification.

Once the interview is complete, the eligibility worker determines whether a household is eligible for SNAP benefits. If a household is found to be eligible for SNAP benefits, the eligibility worker will determine the household's benefit allotment based on the household size. Then, the eligibility worker will notify the household of its allotment amount, stipulations, and length of the certification period before the household must reapply.

Benefits are issued to households via an Electronic Benefits Transfer (EBT) card, which most State agencies mail to the household. EBT cards work like debit cards and can be used at authorized retailers to pay for food using SNAP benefits. Households can use SNAP benefits for food and for plants and seeds to grow food for households to eat.

When a household receives SNAP benefits, it is responsible for updating its information monthly, quarterly, or semiannually if certain changes in its circumstances occur. Households must recertify for SNAP often, commonly on a 6-, 12-, or 24-month basis.

2. SNAP Quality Control

The Food Stamp Act of 1977 established the current SNAP quality control (QC) system, which was designed by FNS to monitor and measure errors in SNAP eligibility and benefit determination. Through SNAP QC, FNS complies with the Payment Integrity Information Act of 2019, which requires Federal agencies to track and mitigate improper payments, defined as payments that should not have been made or were made in an incorrect amount. State agencies must conduct a QC review of a random sample of currently certified cases each month (referred to as active cases), identifying underpayments, overpayments, and total payment error. At the end of the review period for each month's cases, the State agencies share the case files and results with Federal SNAP staff, who review a subsample of the cases for accuracy. The results of this Federal QC review and the State agencies' QC review results are used to calculate each State agency's official payment error rate (PER).[13, 14]

FNS most recently reported that the national PER for fiscal year (FY) 2024 was 10.93 percent (FNS, 2025). Error rates at the State agency level ranged from 3.28 percent in South Dakota to 24.66 percent in Alaska.

Each year, State agencies with a PER greater than six percent must develop corrective action plans to reduce their payment errors; those with PERs greater than 105 percent of the national performance measure for a second consecutive year are placed in financial liability and may be required to pay financial penalties. FNS also uses the data produced through the SNAP QC process to develop an annual SNAP QC Data File. These files contain data on the demographic, economic, and case processing characteristics of more than 50,000 SNAP-participating

[13] The PER is calculated by taking the difference between the value of benefits provided to participants and the value of benefits participants would have received had their cases been processed correctly and dividing the difference by total benefits issued.
[14] State agencies also conduct QC reviews of negative cases that were denied or closed to produce a case and procedural error rate.

households as well as data on the results of the QC review. FNS releases a public-use version of the SNAP QC Data File (FNS, 2020). The restricted-use version contains several additional geographic variables.

Several factors have influenced SNAP QC processes and results in recent years. Because of challenges associated with State-level QC processes, FNS reported PERs for only 11 State agencies for FY 2015. FNS issued revised guidance on the QC process to the State agencies in FY 2016, addressing factors such as the appropriate use of error review committees. The Department of Justice also settled lawsuits with at least eight State agencies and a consultant over violations of the False Claims Act related to SNAP QC results. Finally, SNAP QC was paused during parts of FY 2020 and FY 2021 due to flexibilities in SNAP administration provided by Congress because of the COVID-19 public health emergency.[15] FNS renewed reporting of SNAP QC results starting from FY 2022.

C. Study Objectives and Overview

RA tools offer human services organizations, such as SNAP State agencies, opportunities to efficiently improve payment accuracy. This study provides detailed information on the State agencies that use RA tools in support of SNAP administration, the development of these tools, their effectiveness (when it is possible to measure), and promising practices for FNS and State agencies to consider in the development and use of RA tools.

1. Study Objectives

The specific objectives of the study follow (see appendix D for a full list of research objectives and questions).

Overview of RA Tool Implementation and Use

- Determine which State agencies are using RA tools to reduce error rates and who developed these tools.
- Determine what factors and variables State agencies are using in their RA tools.
- Identify how State agencies act on the results of their RA tools.

Assessment of RA Tool Effectiveness and Accuracy

- Using QC data provided by FNS, determine whether State agencies' RA tools are successful in reducing error rates.

[15] Section 4603(a)(2) of the Continuing Appropriations Act, 2021 and Other Extensions Act (Public Law 116-159), as amended by the Consolidated Appropriations Act, 2021 (Public Law 116-260), allowed State agencies to suspend QC sampling and reviews from June 1, 2020, through June 30, 2021.

Identification of Best Practices

- Determine best practices in development and use of RA tools.

2. Approach Overview

To address the study objectives, the study team employed a four-part data collection approach (for more information on methodology, see appendix A):

- Literature review to identify and evaluate RA tool use in human services program administration
- Survey of all SNAP State agencies to capture information on their development and use of RA tools
- Key informant interviews to provide in-depth information on select State agencies' use of RA tools
- SNAP QC data analysis to evaluate RA tool performance and effectiveness

Literature Review

The study team conducted a systematic literature review by searching academic databases and publicly available online sources to identify relevant peer-reviewed literature, Federal and State reports, publications from business and trade organizations, and vendor-produced materials related to RA tools.

The study team searched for RA tool literature in three broad categories:

- RA tools in social welfare programs
- RA tools in the legal system
- RA tools in data science, statistics, and computer science

These searches yielded 107 documents for review. After deduplicating and screening each item for relevance to the study, the study team downloaded 23 relevant publications. All documents were uploaded into NVivo 12 for coding, following the designated codebook. After the coding was complete, each coder was assigned a subset of the data to review and analyze. For more information about the literature review and search terms, see appendix A.

SNAP Risk Assessment Study Online Survey

The study team developed a web-based survey for SNAP State agencies to complete (see appendix B). The study team programmed and administered the survey using Qualtrics, a web-based survey software. The survey included five modules: (1) State agencies with current case-profiling[16] tools, (2) State agencies that previously used case-profiling tools, (3) State agencies

[16] Following conversations with subject matter experts in Westat's Nutrition Assistance Research Advisory Panel, including former SNAP directors and QC team leads, the study team used the term "case profiling" instead of "risk assessment" in the survey instrument because State staff would be more familiar with it.

that developed but never implemented case-profiling tools, (4) State agencies that neither developed nor implemented case-profiling tools, and (5) State agency context.

To decrease survey respondent burden, respondents did not answer questions in each module; rather, their response to the initial screening question directed them to the appropriate module(s). For example, State agencies that indicated on the screener question that they currently use a case-profiling tool(s) were shown only module 1 (State agencies with current case-profiling tools) and module 5 (State agency context). However, State agencies that indicated they developed but never implemented case-profiling tools were shown only module 3 (State agencies that developed but never implemented case-profiling tools) and module 5 (State agency context).

Key Informant Interviews

The study team conducted interviews with staff in six State agencies: five State agencies with current RA tools (Connecticut, Kansas, Rhode Island, Virginia, Wisconsin) and one State agency that formerly had an RA tool but stopped using it (Utah). The study team considered a variety of factors in selecting State agencies for case studies, including FNS Region, county- or State-level administration of SNAP, monthly caseload, benefits issued, and FY 2023 PERs.

The study team collaborated with each State agency to identify the appropriate staff members or contractors to describe RA tool development and implementation, use, and effectiveness. All interviews were conducted virtually and were recorded and transcribed.

SNAP QC Data Analysis

The study team used SNAP QC data to assess two approaches to measuring RA tool effectiveness. The first measure of effectiveness is whether the tools can accurately distinguish between cases with and without payment errors and efficiently select cases for review. For this analysis, the study team applied State agencies' RA tools to their SNAP QC sample data to identify which cases the tool flagged as high risk. The study team then compared these flags with the SNAP QC findings, which indicated which cases had a payment error, to quantify how well the State agency's RA tools performed.

A second measure of RA tool effectiveness is whether the State agencies that implemented the tools saw a subsequent decrease in their PER. Official PERs are published only once per year; this frequency makes it difficult to detect changes in the PER that might be attributable to an RA tool. Thus, the study team used SNAP QC data and the statistical code developed by FNS to calculate quarterly PERs and then tested for an association between RA tool implementation and changes in the quarterly PERs.

D. Limitations and Considerations

This study has several limitations that warrant consideration.

- Survey completion rate
  - The study team achieved an overall survey response rate of 83.0 percent.[17] Forty-four of 53 eligible State agencies[18] completed the SNAP Risk Assessment Study Online Survey. The average survey question completion rate was 97.6 percent. The lack of a 100-percent survey response rate may limit the generalizability of the findings.
- Historic QC data and PERs
  - SNAP QC methods have changed in recent years, resulting in changes in PERs. Before FY 2016, many State agencies conducted QC reviews using approaches that were inconsistent with the *SNAP QC Review Handbook* (FNS Handbook 310). This practice resulted in FNS calculating PERs for only 11 State agencies in FY 2015; in FY 2016, no PERs were reported. In the same fiscal year, FNS released revisions to the guidance it provides State agencies on the QC process. PERs were again published beginning in FY 2017 and were significantly higher than they were before 2015. Compounding the challenges with QC data, at Congress' direction, FNS placed a moratorium on SNAP QC during the COVID-19 public health emergency, and no PERs were reported for FY 2020 or FY 2021 (see section B2 for additional detail). Therefore, the study team had usable PER data only from FY 2015 to FY 2019 for State agencies with calculated FY 2015 PERs to estimate the effect of RA tool implementation.
  - Other factors at State agencies may affect PERs. For example, substantial changes in how the program is administered may lead to changes in PERs. Examples of such changes include a transition from in-person to telephonic interviews or from county-administered to State-administered interviews, the implementation of a new eligibility system, or the adoption of new policy options. Staffing challenges may also affect PERs, especially when they result in undertrained workers being responsible for processing too many cases. State agencies experienced many challenges in program administration and staffing as they adjusted operations during COVID-19. To address these challenges, the study team included questions in the survey and interviews on other potential causes of changes in PERs.

[17] Though Arizona did not click the submit button to complete its survey, the agency responded to all substantive survey questions, so the Westat and FNS study team decided to include Arizona's response data and categorize them as a complete.

[18] This count includes all 50 States; Washington, DC; Guam; and the Virgin Islands. New York City responded for New York State.

Chapter 2

Development and Implementation of SNAP RA Tools

This chapter provides a national picture of SNAP State agencies' use of RA tools, the factors and variables used in RA tools, and how State agencies apply the results of their RA tools. All results in this chapter are based on the SNAP Risk Assessment Study Online Survey.

Key Takeaways

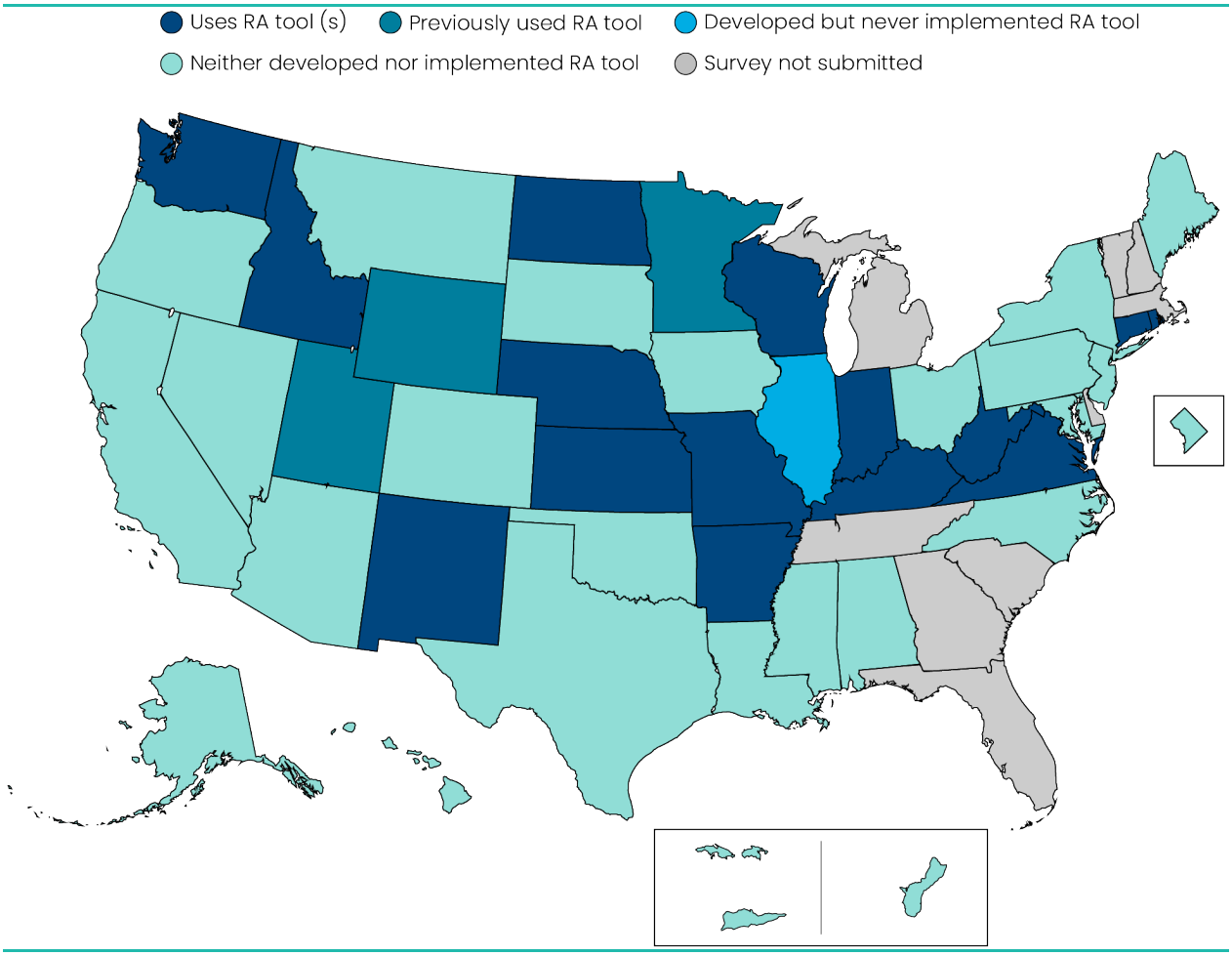
- Among State agencies that responded to the survey, more than a third reported using an RA tool.
- Nearly half the State agencies that use RA tools received help from vendors or contractors when designing the tool.
- State agencies that use RA tools most commonly reported using economic characteristics such as the presence of earned income and self-employment income in their tools.
- More than half of the State agencies that use an RA tool indicated they used their tools to detect at-risk cases after the eligibility determination for new cases but before benefits have been issued.
- Nearly half of the State agencies that have never developed or implemented an RA tool are unfamiliar with them.

A. Variation in Tool Use Across State Agencies

This section identifies which State agencies currently use or previously used RA tools, developed but never implemented RA tools, and never developed or implemented RA tools. Figure 2.1 shows the variation in State agencies' experiences with RA tools in the United States. Of the 53 State agencies eligible for the SNAP Risk Assessment Study Online Survey, 15 (28.3 percent) reported they currently use an RA tool, 3 (5.7 percent) reported they formerly used an RA tool, 1

(1.9 percent) reported having developed but never implemented a tool, and 25 (45.3 percent) reported having never developed or implemented a tool. The remaining nine State agencies (17 percent) did not respond to the survey. See appendix table C.1 for more details.[19]

Figure 2.1. RA survey results by State agency

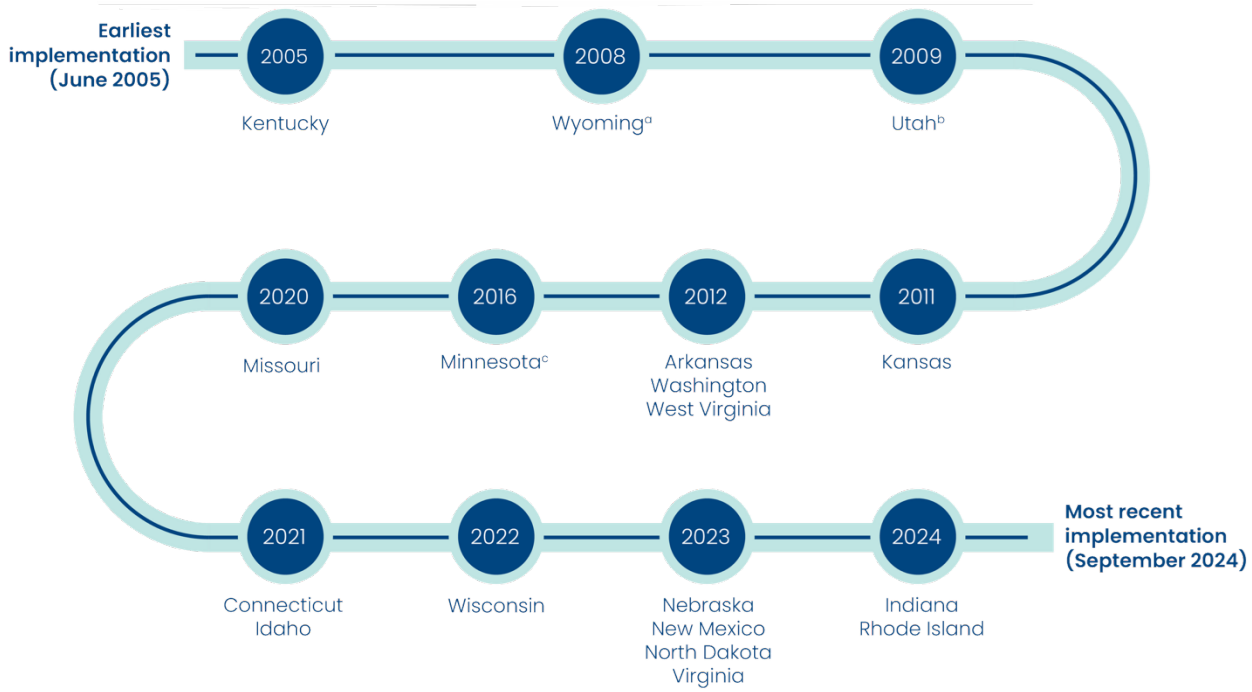


Note: Data for Figure ES.1 is available in Table C.1 in Appendix C.
Note: New York City responded instead of New York State.
Data source: SNAP Risk Assessment Study Online Survey

State agencies have implemented RA tools at various points over the past two decades (figure 2.2). Among reporting State agencies, the earliest State to implement its tool was Kentucky in 2005, followed by Wyoming in 2008. Eight State agencies implemented their tools between 2020 and 2023, and Indiana and Rhode Island implemented their tools most recently in 2024. Minnesota, Wyoming, and Utah discontinued their tools in 2019, 2023, and 2024, respectively.

[19] New York City responded for New York State. New York is a county-administered SNAP State agency, and local social services districts (not New York's Office of Temporary and Disability Assistance) are responsible for administering SNAP. The study team contacted New York City's social services district at the direction of the New York State SNAP director.

Figure 2.2. SNAP RA tool implementation timeline

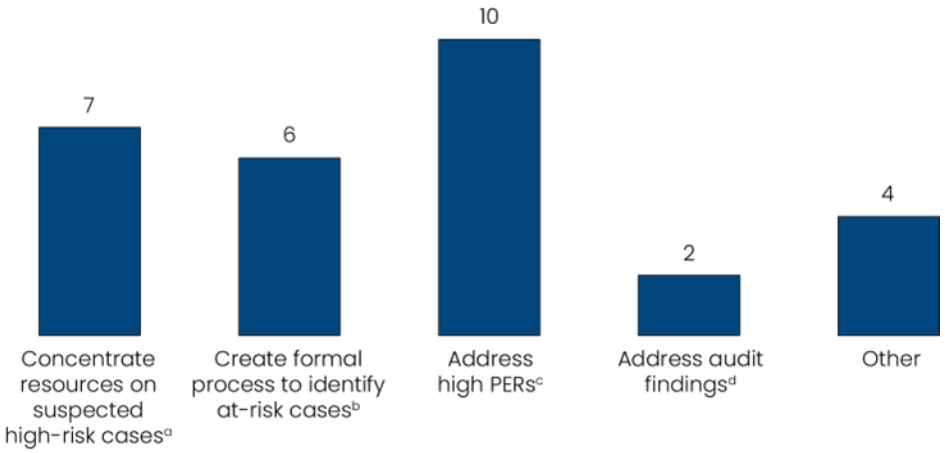


Note: See appendix table C.3 for more details.
[a] Wyoming discontinued its RA tool in 2023.
[b] Utah discontinued its RA tool in 2024.
[c] Minnesota discontinued its RA tool in 2019.
Data source: SNAP Risk Assessment Study Online Survey, questions A3 and B4

B. RA Tool Development

Among the 15 State agencies with RA tools, 10 (66.7 percent) reported they developed their tools to help address higher PERs (figure 2.3). Seven State agencies (46.7 percent) reported they developed their tools to help them concentrate their available resources on suspected high-risk cases. Six State agencies (40 percent) reported they developed their tools to create a formal process for identifying at-risk cases, and two (13.3 percent) reported they developed their tools to address audit findings. Lastly, four State agencies (26.7 percent) reported other motivations for their development of RA tools. These motivations can be categorized as (1) being proactive in maintaining accuracy rates, (2) building in quality checks at the beginning of the process, and (3) helping staff better identify at-risk cases to improve the case review process.

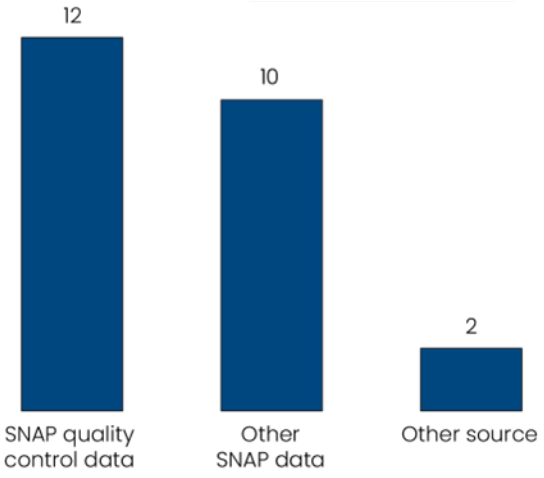
Figure 2.3. Motivations for developing RA tools currently in use



Note: The motivations are not mutually exclusive. See appendix table C.4 for more details.
[a] To concentrate resources (e.g., staff, funds, time) on only those SNAP cases suspected of being at high risk of payment error
[b] To create a formal process for identifying SNAP cases at risk for payment error
[c] PER = payment error rate
[d] To address the findings or recommendations of an audit or management evaluation
Data source: SNAP Risk Assessment Study Online Survey, question A4

Among the 15 State agencies with RA tools currently in use, 12 (80.0 percent) reported using SNAP QC data, and 10 (66.7 percent) reported using other SNAP data when developing their tools (figure 2.4). No State agency reported analyzing proprietary vendor or contractor data to inform their tools. Two State agencies (13.3 percent) reported using another data source.

Figure 2.4. Data analyzed to determine high risk of payment error



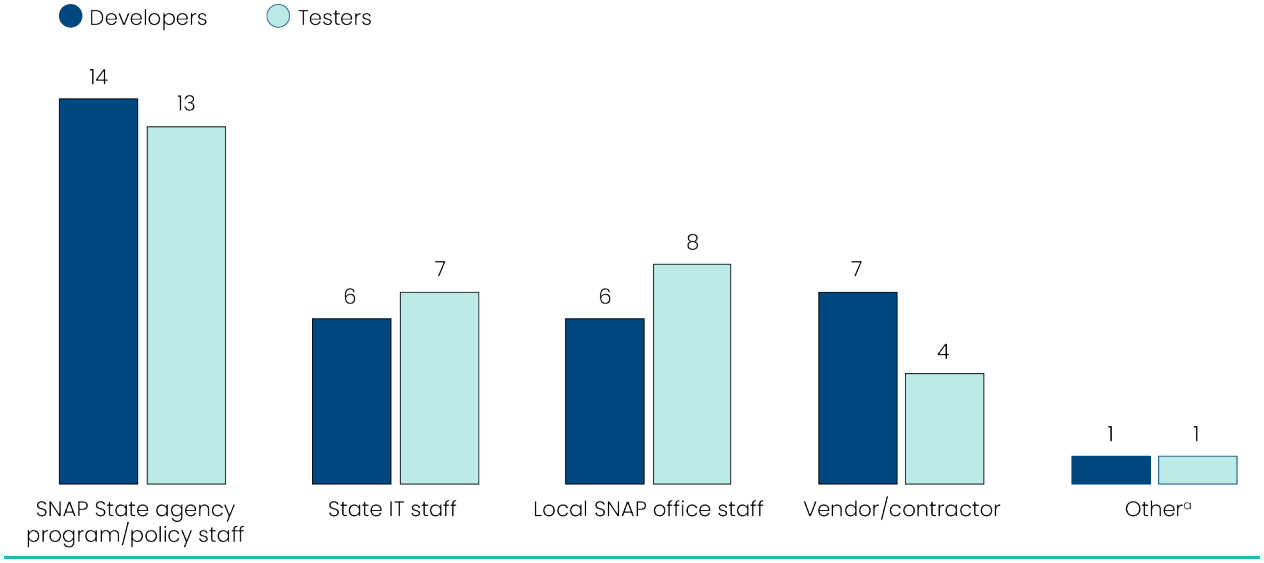
Note: No State agencies reported analyzing proprietary vendor or contractor data to inform their tools. See appendix table C.5 for more details.
Data source: SNAP Risk Assessment Study Online Survey, question A8

Among the 15 State agencies with RA tools currently in use, State agencies most often reported using descriptive statistics to inform RA tool development (7 State agencies; 46.7 percent). One State agency also reported using machine learning (6.7 percent), three reported that they did not know whether their agency conducted any analysis (20.0 percent), and two reported that their State agency did not conduct data analysis (13.3 percent) to inform their tool. Four of the agencies with current RA tools (26.7 percent) reported some other analysis approach to inform the development of their tools. These responses included review of additional data (i.e., agency internal review data), a prior manual process, or prior FNS technical assistance. See appendix table C.6 for more details on the data analyses conducted to develop SNAP RA tools.

Among the two State agencies reporting they did not conduct any analysis, when asked what data source informed the tool, one selected "Other source" and described the eligibility determination process. The other State agency provided information about the data source that informed the tool. When asked about the decision to include selected indicators in its tool, this State agency described its understanding of characteristics that are most strongly associated with payment errors rather than highlighting key indicators. In other words, rather than conducting an analysis specifically to help develop its RA tool, this State agency may have relied on institutional knowledge to make decisions about the indicators that matter for identifying cases at risk for a payment error.

Among the 15 State agencies with RA tools currently in use, most enlisted help from SNAP State agency program/policy staff for both development (93.3 percent) and testing (86.7 percent) of their RA tools (figure 2.5). More State IT staff and local SNAP office staff helped with testing

Figure 2.5. Developers and testers of SNAP RA tools

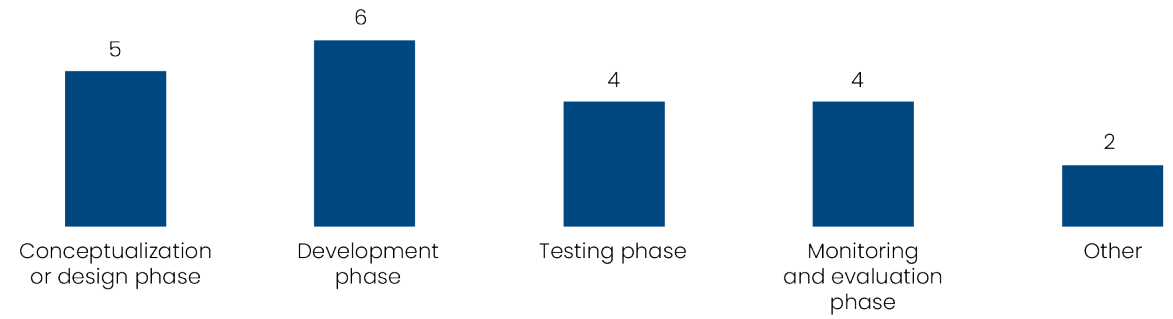


Note: See appendix tables C.7 and C.9 for more details.
[a] The study team recoded FNS Regional or National Office to "Other" for tool developers to facilitate comparability with tool testers.
Data source: SNAP Risk Assessment Study Online Survey, questions A7 and A8

compared with the development of RA tools. Agencies also reported that vendors/contractors more often helped develop RA tools and were less likely to help test the tools.

Among the seven State agencies that received help from vendors or contractors in designing their RA tools, the degree of State agency involvement varied across four phases of tool development (conceptualization or design, development, testing, and monitoring and evaluation). State agency staff in five agencies (71.4 percent) helped during the conceptualization or design phase, staff in six agencies (85.7 percent) were involved during the development phase, and staff in four agencies (57.1 percent) were involved during both the testing phase and the monitoring and evaluation phase (figure 2.6).

Figure 2.6. State agency involvement in RA tools currently in use developed by vendors or contractors

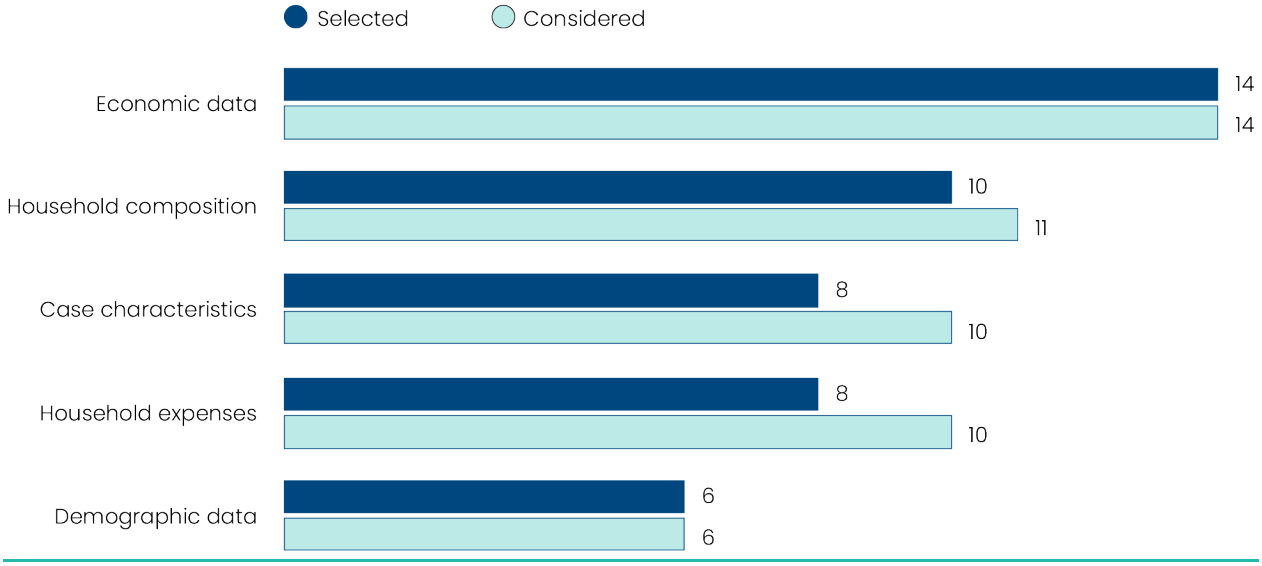


Note: See appendix table C.8 for more details.
Data source: SNAP Risk Assessment Study Online Survey, question A6

C. Variables Used in RA Tools

SNAP State agencies that implemented RA tools considered and included a variety of variables in their tools. Variable categories included (1) economic characteristics, (2) household composition, (3) demographic characteristics, (4) household expenses, and (5) case characteristics. Variables indicating a case's economic situation were among those that State agencies more commonly reported incorporating in their tools to detect cases at risk for a payment error (figure 2.7). For example, among the 15 State agencies with RA tools currently in use, 10 State agencies (66.7 percent) reported including self-employment income and presence of earned income in their tools (see appendix table C.11). Agencies also reported they incorporated presence of unearned income (53.3 percent) and zero income (46.7 percent) at a higher frequency than the other remaining economic characteristics (see appendix table C.11). Related to a case's household composition and demographic information, most State agencies (60.0 percent) also reported including total number of household members and employment status to help their tools detect cases at risk for a payment error (see appendix table C.11).

Figure 2.7. Variable categories considered and included in RA tools by State agencies



Note: See appendix table C.11 for more details. Percentages may sum to more than 100 percent because responses are not mutually exclusive
n = 15
Data source: SNAP Risk Assessment Study Online Survey, questions A9–A13

Another category of variables State agencies often included in their RA tools was household expenses. The most commonly reported household expenses agencies included in their tools were utility expenses (including SUAs [Standard Utility Allowances]) and non-utility shelter expenses (both reported by 46.7 percent of agencies; see appendix table C.11). In addition to these expenses, 33.3 percent of agencies also reported including legally obligated child support, excess shelter deduction, dependent care expenses, and medical expenses (see appendix table C.11).

Few State agencies reported including case characteristics into their RA tools; the most common was benefit amount (reported by 46.7 percent of agencies). No agency reported considering or including race, ethnicity, sex, level of education, or citizenship in their tools.

D. How State Agencies Use RA Tool Results

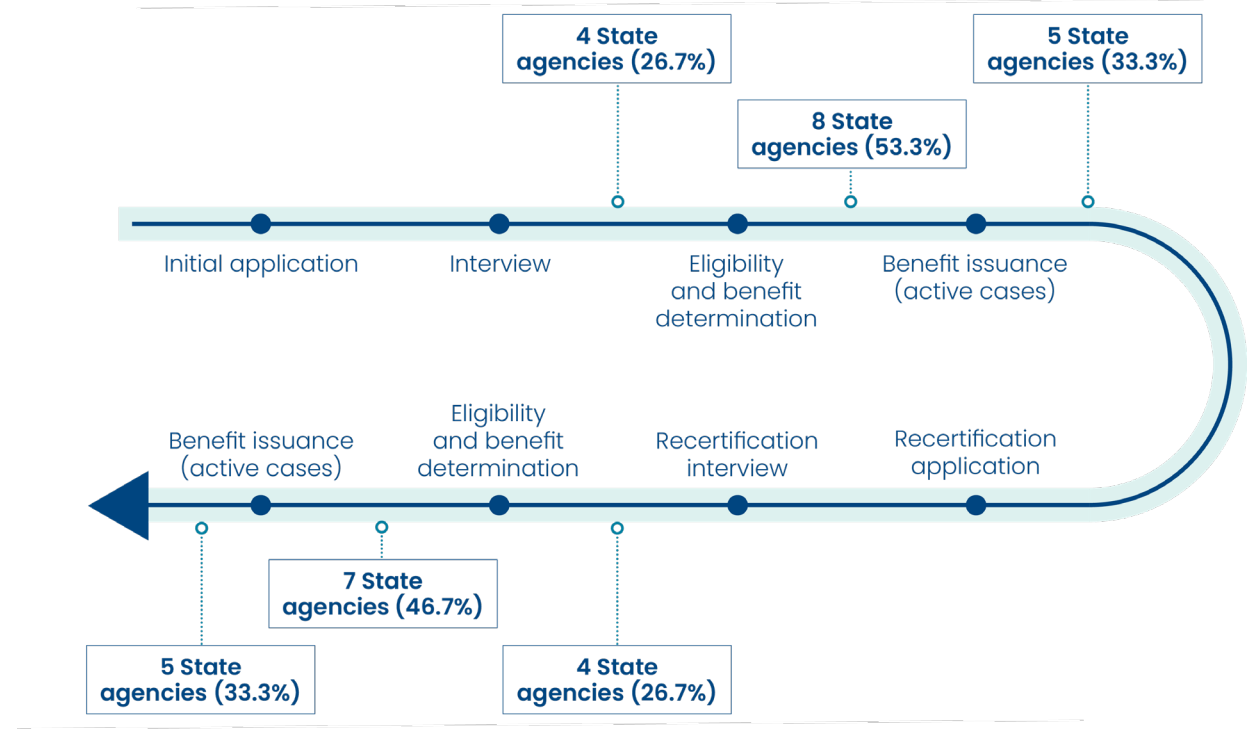
State agencies implement RA tools in various formats. Among the 15 State agencies with RA tools currently in use, 9 State agencies (60.0 percent) reported local agencies can customize the tools (appendix table C.14). The most reported formats for tools include written instructions, paper checklists, and electronic checklists. The second most reported format is an algorithm programmed into the agency's eligibility system or other systems or databases. A small proportion of agencies indicated other formats for their tools, including an *"Excel file with full listing broken down by defined criteria"* and a *"stand alone data collection system."* One agency described the process for implementing its tool by stating that a *"data team identifies random cases based on criteria"*; see appendix table C.13 for more details. RA tools most often use household interview data (reported by 14 State agencies, 93.3 percent) and data from the

household application (reported by 13 State agencies, 86.7 percent). Nine agencies (60.0 percent) reported their tools use data from data matches. See appendix table C.16 for more details.

Local, county, and State agency staff receive training on how to use RA tools through several modes. State agencies most commonly conduct trainings online without a live presenter or facilitator. This mode of training was reported by six agencies (40.0 percent) for local staff, four agencies (26.7 percent) for county staff, and six agencies (40.0 percent) for State staff. Other types of training indicated by State agencies include written tutorials and in-person trainings; see appendix table C.10 for more details.

State agencies reported wide variation in when they use RA tools to detect cases at risk for a payment error (figure 2.8). Among the 15 State agencies with RA tools currently in use, 8 State agencies (53.3 percent) indicated they use their tools to detect at-risk cases *"after the eligibility determination for new cases but before benefits have been issued."* The second most frequently reported time in which agencies use their tools to identify at-risk cases was *"after the recertification determination but before benefits have been issued"* (reported by seven agencies, or 46.7 percent). Five agencies (33.3 percent) reported using their tools after initial or recertification benefits are issued. A small proportion of State agencies offered additional nuance in their open-ended responses describing when they use RA tools. These included *"any time benefits are re-calculated—change reported [sic] or periodic reporting"*; *"before an eligibility determination is completed for applications, changes, and recertifications"*; *"after the periodic report determination but before benefits have been issued"*; and *"whenever supervisors and the case review team have a case review total to meet and want to find a case sample to review."*

Figure 2.8. When State agencies apply RA tools to SNAP cases



Note: Four State agencies responded "other time period" to this question and are not integrated into the numbers shown. See appendix table C.15 for more details.
Data source: SNAP Risk Assessment Study Online Survey, question A6

Figure 2.9 presents a chart highlighting the process State agencies follow when they use RA tools to identify cases at risk for a payment error, including what happens to identified cases and the staff most often responsible for each step. State agencies most often reported local eligibility worker supervisors, local eligibility workers, and State quality assurance staff as users of RA tools (appendix table C.17).[20] After a State agency's RA tool flags a case at risk for a payment error, the case typically undergoes a second review by an eligibility worker or eligibility worker supervisor, or it undergoes a quality assurance review (appendix table C.18). After determining a case's risk of a payment error is high, local eligibility worker supervisors or eligibility workers follow up. Other commonly reported staff who follow up on at-risk cases include State quality assurance staff, county eligibility worker supervisors, and county eligibility workers; see appendix table C.19.

[20] County-level QA staff, eligibility workers, and other staff were among those least reported to use an agency's RA tool.

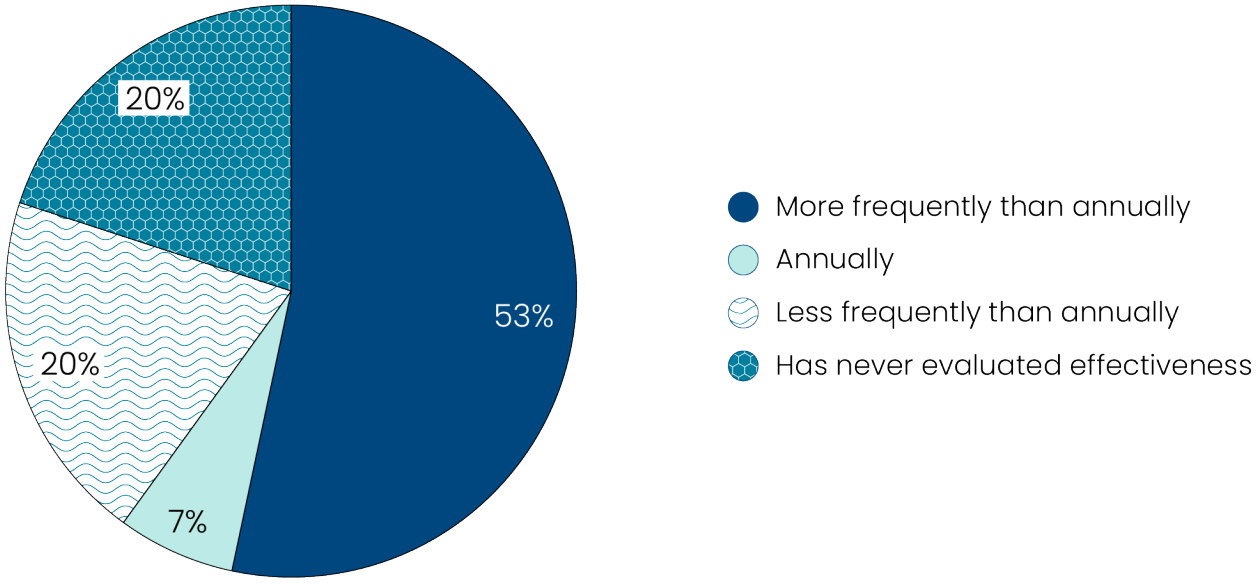
Figure 2.9. Steps State agencies take after their RA tool flags a case at risk for payment error

Note: See appendix tables C.17, C.18, and C.19 for more details.
Data source: SNAP Risk Assessment Study Online Survey, questions A23, A24, A26, A27, A29, A30, and A32

Respondents from three State agencies (20.0 percent) use their RA tools for other purposes in addition to identifying cases likely to have payment errors. One respondent wrote, *"We use this data to make updates in our process manually and possibly training."* Another respondent said their agency *"has an integrated eligibility system for SNAP, CCAP [Child Care Assistance Program], LIHEAP [Low Income Home Energy Assistance Program], Medicaid and TANF [Temporary Assistance for Needy Families]. Error proofing is completed at the time of application and review for all programs in which a determination was made by a new eligibility worker or a worker new to a program."* The third respondent said, *"The data can be analyzed to determine training needs for individuals or training needed overall."* See appendix table C.20 for more details.

Among the 15 State agencies with RA tools currently in use, 12 agencies (80 percent) evaluate the effectiveness of their RA tools to accurately detect cases at risk for a payment error (figure 2.10). Eight agencies (53.3 percent) conduct such evaluations more frequently than annually, one agency (6.7 percent) conducts them annually, and three agencies (20.0 percent) conduct them less frequently than annually. Three agencies (20.0 percent) have never evaluated the effectiveness of their tools.

Figure 2.10. Frequency of RA tool effectiveness evaluations



Note: See appendix table C.22 for more details on the frequency of effectiveness evaluations.
Data source: SNAP Risk Assessment Study Online Survey, question A35

E. Reasons Some State Agencies Do Not Currently Use RA Tools

This section highlights key findings from the 25 State agencies that never developed RA tools and summarizes why some State agencies stopped using RA tools. One State agency developed but never implemented an RA tool.[21]

Of the 25 State agencies that have never created an RA tool, 12 agencies (48.0 percent) reported they are not familiar with RA tools. Of the 13 State agencies that are familiar with these tools, eight State agencies (61.5 percent) reported they have considered developing one to identify cases at risk for a payment error. Among the five State agencies that have not considered developing a tool, two provided substantive reasons for not developing a tool. These reasons included not wanting to affect the *"QC sample by imposing bias or inequities toward certain household types"* and that the current policies and procedures for eligibility already help maintain a low PER. See appendix table C.24 for more details.

Three State agencies implemented but discontinued their RA tools. Minnesota implemented its tool in 2016 and discontinued it in 2019. Utah implemented its tool in 2009 and discontinued it in 2024, and Wyoming implemented in 2008 and discontinued in 2023. When asked about the reason their tools were no longer in use, Minnesota responded, *"It's been on pause since the*

[21] Illinois' SNAP State agency program/policy staff and an outside vendor/contractor are developing a computerized RA tool to reduce SNAP error rates post-eligibility. The computerized tool will include a summary screen, historical records, indicator comments, and Tableau dashboards (see appendix table C.23).

pandemic started as we are currently focusing our SNAP-ME [management evaluation] review efforts in other areas. We hope to bring it back into use again in the future." Utah said it *"discovered that the error-prone cases were less error-prone than others. As a result, the Performance Review Team (PRT) recently made changes to minimize using case-profiling tools."* Wyoming reported that it *"switched to a universal caseload and at this time staff are focused on learning the steps for this. We do plan to implement tools in place in the future."* See appendix table C.25 for more details.

Chapter 3

Case Studies of Six State Agencies' RA Tools

This chapter provides detailed information on the development, implementation, and impacts of RA tools used by six State agencies: Connecticut, Kansas, Rhode Island, Utah, Virginia, and Wisconsin. Five of the six State agencies were using their RA tool at the time of data collection; Utah discontinued its tool in summer 2024. Originally, 14 State agencies were identified as possible case study States. The study team considered all State agencies that completed the survey nine weeks after the survey launch and indicated they currently use or previously used an RA tool. Through conversations with FNS, the study team narrowed the six case study State agencies to include a mixture of FNS Regions, SNAP administration types, monthly caseloads, benefits issued, PERs, and tool use timelines. The study team interviewed State agency staff from each of the six States between November 2024 and January 2025; local agency staff were also invited to participate in interviews if they used an RA tool or collaborated on its development.

Key Takeaways

- State agencies reported challenges balancing the need for accuracy and timeliness of case processing. Some indicated that implementing an RA tool to improve accuracy placed additional burden on the eligibility workers processing SNAP cases. If an eligibility worker's case was flagged for additional review by the tool, it extended the amount of time needed to process it. Some respondents also said the extra time they spent resolving flagged cases meant they had less time to process the rest of their caseload.
- Five of the six State agencies designed their RA tools to flag cases after an eligibility worker determined eligibility but before they issued SNAP benefits. The final State agency designed its tool to flag cases after SNAP benefits were issued, and eligibility workers corrected errors for the following month.
- State agencies described mixed experiences working with contractors to build, test, and refine RA tools. Some agencies wanted greater insight into the programming

specifications their contractor wrote to ensure the coding matched their goals for the tool and aligned with SNAP policy.

- Eligibility workers and eligibility worker supervisors are often the end users of RA tools, but only one State agency involved those staff in testing its RA tool before launch; for some State agencies, this resulted in tools that were not incorporated into the eligibility determination process very well.
- Five of the six State agencies' RA tools included the SNAP benefit issuance amount when selecting cases for review. Some staff indicated this variable was included because of the larger potential impact on the State's PER, not because cases with larger issuance values are at higher risk for a payment error.
- Four of the six State agencies do not track the outcomes of cases flagged by RA tools (e.g., among the cases flagged, agencies did not track how many contained a case error or payment error).
- When working to reduce PERs, States implemented multiple concurrent strategies, impeding the study team's ability to determine the true impact of RA tools on State PERs.

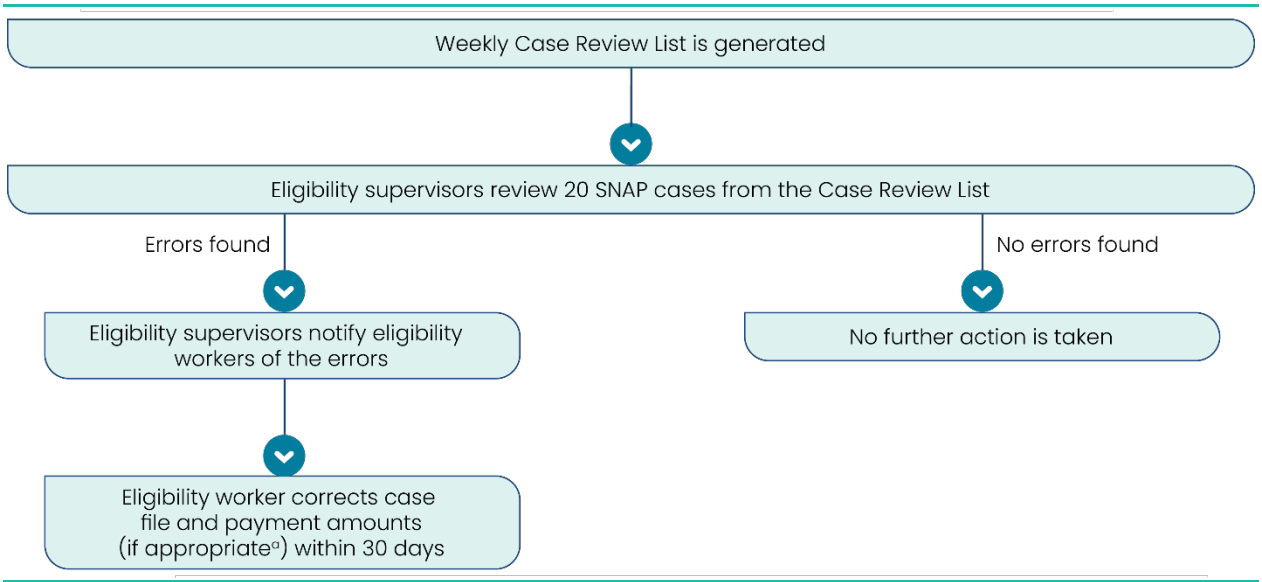
A. Connecticut's Case Review List

Connecticut's State agency uses an RA tool called the Case Review List. The State agency developed this Microsoft Excel-based tool in 2021 to address high PERs; in 2019, Connecticut's PER was 10.5 percent.[22] Historically, the State agency required eligibility worker supervisors to review at least 20 SNAP cases per month that their eligibility workers processed, but the agency had determined that some eligibility worker supervisors were "cherry picking" easier cases for review. One respondent said this happened when eligibility worker supervisors were short on time or *"apprehensive of reviewing individuals in their unit because they didn't want to cause tension."* The State agency created the Case Review List to help supervisors more objectively select cases for review based on characteristics known to put a case at risk for a payment error. Ultimately, the hope was that this approach would improve accuracy and reduce payment errors.

Today, eligibility worker supervisors use the Case Review List at the end of each month to select 20 cases at risk for a payment error and flag them for a second review (see figure 3.1). The second review occurs after benefits are issued for SNAP certifications and recertifications, so the Case Review List functions as a corrective quality assurance tool rather than a preventive one. Any errors found are corrected for the following month's benefit issuance. State QC staff also use the Case Review List for broader quality assessments.

[22] Data for 2020 and 2021 are not available.

Figure 3.1. Use of Connecticut's Case Review List



[a] Case errors may or may not lead to a payment error. For example, a household's address may have been entered incorrectly (case error) but would not result in overpayment or underpayment of SNAP benefits (payment error).

The State agency's eligibility system generates the weekly Case Review List in Excel (see screenshot in figure 3.2) and includes all SNAP certifications and recertifications issued in the prior week. The list includes the following data points for each case, as applicable:

- SNAP benefit amount
- Earned income value
- Unearned income value
- Self-employment value
- Shelter expense value
- Utility expense (yes/no)
- Number of household members

Eligibility worker supervisors manually review and filter the list to select 20 cases for review each month. The State agency suggests that eligibility worker supervisors start by identifying cases that meet the category A risk criterion, and, if fewer than 20 fit this category, they expand their selection to include cases under the category B and category C criteria:

- **Category A:** SNAP benefit amount is $400 or more per month.
- **Category B:** SNAP benefit amount is between $200 and $399 per month.
- **Category C:** SNAP benefit amount is $199 and below per month.

However, eligibility worker supervisors are not required to adhere strictly to this "ABC" approach and have discretion to select other cases for review using the other data points in the list. One respondent said she looks for cases with unusual income patterns, such as one with $0.16 listed for earned income, and flags those for review to ensure that the case data are accurate and properly documented. After the supervisor reviews each of their 20 monthly cases, they return them to the original eligibility worker who processed the case to address any errors found; they also provide positive feedback when cases are processed correctly. The eligibility worker then has 30 days to make any corrections, which sometimes requires contacting the household for more information. The goal is to correct errors before benefits are issued the following month.

> *"The reason we went this [manual] route was because we wanted to be able to target the areas that we know are consistently showing up on our reports as being our highest errors. We didn't necessarily want to leave it to chance of the system to pull cases. We wanted to be a little bit more defined and exact in it."*
>
> **—Connecticut respondent**

Figure 3.2. Screenshot of sample fields in the Case Review List

A	B	C	D	E	F	G	H	I	J	K	L
First Name	Last Name	Case Number	Application Date	Benefit Amount $16 & >	Earned Income	Unearned Income	Self-Employment Income	Shelter Expense	Utility Expense	Client Office #	Worker ID
Jane	Doe	[Case #]	6/1/2024	1491	0	0	1431.25	0	N	[Office #]	[ID#]
John	Smith	[Case #]	6/1/2024	1386	0	0	0	0	Y	[Office #]	[ID#]
Ebony	Jones	[Case #]	6/1/2024	1386	0	936	0	900	Y	[Office #]	[ID#]
Daniel	Garcia	[Case #]	6/1/2024	1302	0	1042	0	0	N	[Office #]	[ID#]
David	Black	[Case #]	6/1/2024	1155	0	0	237.32	1350	Y	[Office #]	[ID#]

Note: This screenshot does not include data from real participants.

1. Tool Development and Implementation

State agency program staff collaborated with SNAP QC and State IT staff to develop the Case Review List. The eligibility system contains an array of preprogrammed reports, and the team adapted one of those to produce the monthly Case Review List. One State agency respondent said that the data points selected for inclusion (see previous section) reflect the characteristics that put a SNAP case at risk for a payment error, per QC data.

The State agency program staff and IT staff tested the Case Review List using sample case files. They verified that the data from the sample case files matched the output in the Case Review List (e.g., that the earned income values from each sample file were accurately reflected in the Case Review List). During testing, the team decided to include the date each application was certified or recertified and program the Case Review List to pull the applications that could be reviewed as quickly as possible after certification or recertification. Prioritizing the most recently processed cases made it less likely that an error would be found by the QC team and counted toward the

State's PER. As one respondent described, *"Obviously, it's great to have the cases fixed, and that is one of the top priorities. But if you do it after [that case is] already part of the [Federal QC] sample, you're still going to be hit with the payment error, so it kind of defeats the second half of the purpose."* No other changes were made to the Case Review List after testing or implementation.

Eligibility worker supervisors attended formal training sessions on how to use the Case Review List when it was first introduced. The training consisted of a 1-day instructional session in which supervisors learned to navigate the tool and select cases for review using the "ABC" criteria. Eligibility workers do not directly interact with the Case Review List and therefore did not receive training on it.

2. Impact of the RA Tool

On Staff Workloads

Respondents did not comment on the impact of the Case Review List on staff workloads. However, they remarked that a high number of SNAP cases combined with hundreds of staff vacancies across the State (both State- and local-level positions involved with case processing and review) make it challenging to complete the required number of case reviews each month.

On the Payment Error Rate

Most respondents view the Case Review List as a valuable tool for reducing SNAP payment errors, although one person said they could not comment on its effectiveness. However, the tool's true impact is difficult to determine because the Case Review List is just one of several strategies the State agency has implemented to address payment errors. Other strategies mentioned by staff include the following:

1. Developing more targeted training materials for eligibility workers and supervisors on common sources of error
2. Linking the eligibility systems for SNAP and LIHEAP to ensure those benefits are accurately captured for relevant SNAP cases

Overall, the State agency remains committed to exploring a combination of staff training and system changes to minimize payment errors.

B. Kansas' Pre-Authorization Screening Checklist

In the early 2000s, Kansas' PER surpassed the 6-percent threshold for possible financial liabilities and corrective action, and in 2011, FNS required the State agency to implement a corrective action plan[23] to reduce its PER. The State agency created the Microsoft Word-based Pre-Authorization Screening Checklist as part of that corrective action plan. The agency designed the checklist to flag all SNAP cases at risk for a payment error for a second review before benefit issuance. Eligibility workers and supervisors still use the tool today.

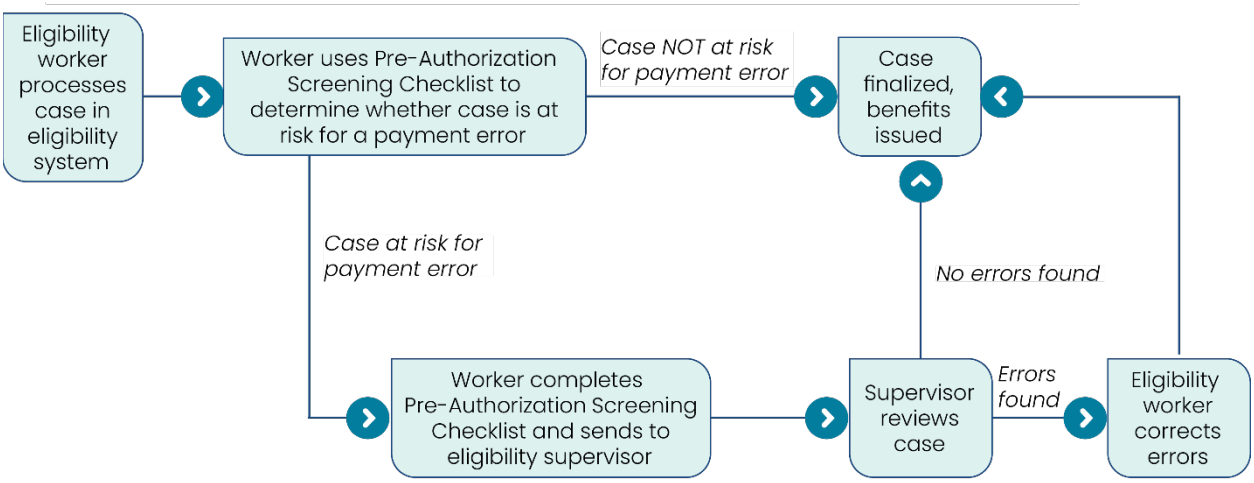
Eligibility workers use the Pre-Authorization Screening Checklist when reviewing SNAP cases undergoing certification, recertification, or a periodic report (see figure 3.3). After processing a SNAP case in Kansas' eligibility system, an eligibility worker reviews a final authorization screen with the proposed eligibility determination and estimated SNAP benefit amount. On that authorization screen, they can open the Pre-Authorization Screening Checklist, stored as a Microsoft Word file, to help them identify cases at risk for a payment error and flag them for a second review. The checklist instructs eligibility workers to flag cases with all of the following characteristics:

- Households of three or four people
- Estimated benefit issuance amount greater than $600 per month
- Earned income greater than $400 per month **and/or** child support income greater than $400 per month

If a case has none of those characteristics, the eligibility worker confirms the benefit issuance amount and authorizes payment. If the case has one or more of those characteristics, the eligibility worker completes the Pre-Authorization Screening Checklist to indicate that the case needs a second review and sends it to their eligibility worker supervisor via email. The eligibility worker supervisor then conducts a full case review. If the supervisor finds errors, they email the eligibility worker to make the needed corrections before finalizing the case and authorizing payment. If no errors are found, the supervisor emails the eligibility worker to direct them to confirm the benefit issuance amount and authorize payment. The eligibility worker documents in the case file if the case underwent a pre-authorization second review.

[23] FNS requires State agencies to develop and execute a corrective action plan if their PER is six percent or greater. During the corrective action planning process, State agencies determine appropriate actions to substantially reduce or eliminate deficiencies leading to payment errors.

Figure 3.3. Use of Kansas' Pre-Authorization Screening Checklist



1. Tool Development and Implementation

None of the State agency respondents were involved with developing the Pre-Authorization Screening Checklist and therefore could not comment on the processes to develop or test the tool at its inception. However, respondents indicated that the case characteristics used to flag cases for review have changed slightly over time. Originally, the Pre-Authorization Screening Checklist instructed eligibility workers to flag cases with all of the following characteristics:

- Households of three or four people
- Earned income greater than $400 per month **and/or** child support income greater than $400 per month
- Estimated benefit issuance amount of $400 or more per month

In 2022, the State agency updated the Pre-Authorization Screening Checklist after eligibility workers and their supervisors reported that the tool flagged too many cases for review. *"They were getting so many [Pre-Authorization Screening Checklists to review] that it was becoming apparent that the checklist wasn't narrowing [cases] down enough,"* said one respondent. The State agency asked the SNAP QC team to identify current payment error trends and suggest updates to the characteristics used to flag cases. Drawing on this updated analysis of SNAP QC data, the State agency revised the tool to increase the threshold for estimated benefit issuance from $400 to $600 per month.

Respondents had mixed feedback about the utility of these changes. On the one hand, increasing the threshold for benefit issuance helped align the tool with policy changes that increased the SNAP benefit allotments since the tool's inception in 2011. However, some respondents said the errors they currently see in SNAP QC reports in 2025 no longer align with the Pre-Authorization Screening Checklist updates made in 2022, which suggests the tool may need to be updated again.

In 2022, the State agency made changes to staff trainings on the tool. First, the agency began training all new eligibility workers on the Pre-Authorization Screening Checklist during orientation. Second, the agency revised the eligibility workers' training manual to include information about the tool and how it works. Despite the training and materials provided, all respondents indicated that eligibility workers use the Pre-Authorization Screening Checklist inconsistently. Most notably, respondents said that eligibility workers do not reliably use the tool to identify which cases qualify for a second review. They attributed this inconsistency to conflicting information taught in trainings and mixed communications about tool use (see textbox). Some State agency respondents perceived that the more experienced eligibility workers view the tool as a punishment or an indication that agency leadership does not trust them to process cases correctly. This view may also affect the willingness of staff to use the tool.

> *"I feel that [eligibility workers] are interpreting the checklist differently [...] If staff don't understand what the checklist is for and they're not submitting the cases to the supervisors to review them, then those are cases that are not being reviewed."*
>
> **—Kansas respondent**

2. Impact of the Tool

On Staff Workloads

The Pre-Authorization Screening Checklist is integrated into the daily case processing tasks of eligibility workers and their supervisors, but staff described mixed experiences using it. Some eligibility workers reported the Pre-Authorization Screening Checklist is not burdensome to complete. However, State agency respondents had received feedback from other eligibility workers and eligibility worker supervisors (they indicated these are typically more experienced eligibility workers) that using the Pre-Authorization Screening Checklist slows case processing. Eligibility worker supervisors said the review process initiated by the Pre-Authorization Screening Checklist is both onerous and disruptive. They noted that completing a full review of a single case can take up to an hour, which impedes their ability to complete other required tasks each day.

> *"For [eligibility] supervisors, it's adding more work, definitely, because we want to have a very quick turnaround so we're stopping what we're doing to get those cases read as soon as possible."*
>
> **—Kansas respondent**

State agency respondents indicated they are devising enhancements to the tool to make it less burdensome. They also plan to standardize trainings and materials, which they hope will promote more consistent use of the tool by eligibility workers and supervisors statewide.

On the Payment Error Rate

At the time it was first implemented, the tool was expected to flag 14 percent of cases with payment errors based on its performance (see table 4.2. Confusion matrix for Kansas' Pre-Authorization Screening Checklist in chapter 4). If Kansas corrected the issues with all flagged cases, the effect on the payment error would likely be modest. However, the true impact of the Pre-Authorization Screening Checklist on the State's PER cannot be determined with existing data. Since the tool's inception in 2011, the State agency has implemented several strategies to reduce the PER, making it difficult to determine the influence of the Pre-Authorization Screening Checklist on its own.

C. Rhode Island's Pre-Authorization Review Tool

> *"The only way to beat an error is to catch it before you authorize the case."*
>
> **— Rhode Island respondent**

For years, the Rhode Island State agency used the Quality Improvement Review System to pull for review a random sample of SNAP cases that eligibility workers had processed and approved. However, reviewing cases after processing and approval did not prevent incorrect benefits payments from being issued, and the State agency's PER reached 12.4 percent in 2023. The State agency programmed the Pre-Authorization Review Tool (Pre-Auth Tool) within its electronic eligibility system to address that issue—catching errors before benefits were issued—and allocate resources more effectively to focus on cases thought to be at risk for a payment error.

The State agency began piloting its Pre-Auth Tool in 2024 with a select group of eligibility workers and supervisors across the State. Launching the tool as a pilot enabled the State agency to refine the procedures and programming in advance of statewide implementation (expected in 2025). After implementation, the State agency expects eligibility workers and their supervisors to use the Pre-Auth Tool when processing both certifications and recertifications.

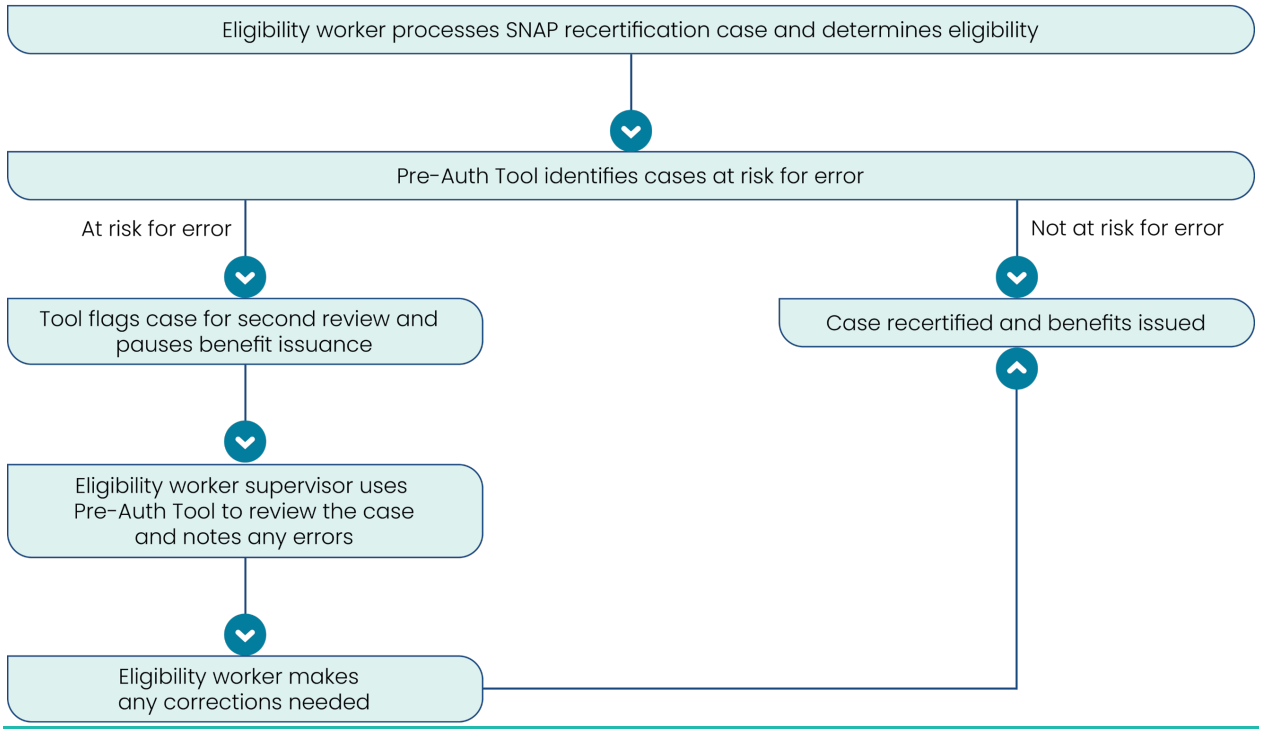
For the pilot, the Pre-Auth Tool in the State eligibility system selects from among the SNAP cases undergoing recertification. Local agencies have flexibility to limit the number of cases selected for a second review each day according to the number of available eligibility worker supervisors. After an eligibility worker makes a determination about a SNAP recertification case and before benefits are issued, the Pre-Auth Tool flags the case for a second review if it has one or more of the following characteristics:

- Five or more people in the household
- Earned income of $1,400 or higher per month
- Unearned income of $1,500 or higher per month
- Any self-employment income
- Any rental income

- SNAP benefit amount of $200 or higher per month

If a case has one or more of these criteria, the Pre-Auth Tool flags it, puts benefit issuance on hold, and sends it to an eligibility worker supervisor for a second review (figure 3.4). Eligibility worker supervisors receive email alerts when a case needs their review and use the Pre-Auth Tool to assess the case file and indicate where corrections, if any, are needed. Eligibility workers receive a similar email alert when the supervisor completes the review and can read the notes and make any necessary corrections before approving the case and issuing benefits.[24]

Figure 3.4. Rhode Island's pilot of the Pre-Auth Tool



1. Tool Development and Implementation

State agency staff reviewed SNAP QC data to understand which characteristics were associated with risk for payment error and identified the characteristics listed in the previous section. They then collaborated with a contractor to develop the Pre-Auth Tool and incorporate it into the State's electronic eligibility system. Although the elements used to flag cases cannot be changed, State agency respondents say the contractor programmed sufficient flexibility into the tool to allow them to adjust the value thresholds (e.g., earned income of $1,400 or higher per month) as needed.

[24] The procedural steps described may have gaps because the study team could not meet with all State agency and local agency staff with knowledge of the tool during the data collection period.

After programming was complete, the contractor tested the tool's performance and passed it to the State agency for additional testing. State agency staff encountered numerous challenges during testing and when training staff to use the Pre-Auth Tool. During testing, State agency respondents said they struggled to understand how to use the Pre-Auth Tool efficiently and encountered various functionality issues. Respondents expressed that many of the challenges they encountered during testing stemmed from the limited involvement of State agency staff in the programming of the Pre-Auth Tool. This lack of visibility into the programming specifications hindered their understanding of the tool's logic and their ability to pinpoint the reasons why the tool incorrectly flagged certain cases. It also made it difficult for those State agency staff to articulate the mechanics of how the tool works when training eligibility workers and supervisors participating in the pilot. Reflecting on the challenges encountered during testing, one State agency respondent expressed skepticism that the contractor had truly tested the tool.

> *"It's just challenging to see how it was designed. It's counterintuitive to me. … It doesn't honestly make sense a lot of times. So those have been I think the biggest challenges that I have come across because you're trying to translate this [process for the eligibility] supervisors and have the supervisors all complete the reviews in the same way."*
>
> **— Rhode Island respondent**

Before launching the Pre-Auth Tool pilot, the State agency developed a PowerPoint presentation to demonstrate how to navigate and use it to review cases and indicate where case corrections were needed. Staff posted the Microsoft PowerPoint in the State agency's learning management system for eligibility workers and supervisors participating in the pilot to review. One respondent said they also developed tipsheets and held more hands-on trainings because the Pre-Auth Tool *"isn't really straightforward."*

Despite the trainings provided, respondents said that eligibility staff have found the Pre-Auth Tool difficult to use during the pilot. They described it as not user-friendly and lacking intuitive functionality. For instance, one respondent said the Pre-Auth Tool requires a precisely sequenced series of steps to work correctly and equates it to learning a "secret code" of buttons to press in a video game. Another respondent agreed that using the tool is not intuitive: *"The workers, I feel, are still struggling. They're getting better. I think that the process when it first rolled out probably really was not communicated well enough."* That complexity has led to frustration among eligibility staff and requests for more training and guidance.

The State agency respondents empathize with the eligibility workers and eligibility worker supervisors and reference their own challenges using the Pre-Auth Tool after it was rolled out for the pilot. Their comments reveal that some of the challenges stem from misunderstandings with the contractor about what constitutes an error. For example, one respondent said that several months of monitoring led them to realize that the contractor had incorrectly programmed the tool to flag errors for all cases with the status "no further actions required."

Although the Pre-Auth Tool has been difficult to implement, State agency respondents believe that the functionality has improved and that the tool's effectiveness will depend on teaching eligibility workers and supervisors to navigate its complexities. They added that the overall transition from the Quality Improvement Review System to the Pre-Auth Tool was a positive change because (1) it enabled them to catch errors before benefit issuance, and (2) using the tool helped eligibility workers learn to better spot potential errors while processing cases. They remain optimistic that all staff will become more adept at using the Pre-Auth Tool with time.

2. Impact of the Tool

At the time of data collection, the Pre-Auth Tool was still in the pilot stage; therefore, the impact on Rhode Island's PER and staff workloads could not be assessed. Respondents expressed optimism that, once the contractor resolves the functionality challenges and eligibility staff use it consistently, the tool will help improve accuracy during case processing and reduce the State's PER.

All information on staff use of the Pre-Auth Tool and the outcomes of the cases flagged feed into a dashboard. The dashboard is accessible to both eligibility worker supervisors and State agency staff who use it to view key metrics, including the following:

- Number of flagged cases
- Number of flagged cases pending review
- Number of completed reviews
- Error trends (e.g., common mistakes made related to calculating household income)

State agency staff review the data on error trends to determine what additional trainings eligibility workers might need.

D. Utah's Performance Review Case-Profiling Tool

Utah introduced its Performance Review Case-Profiling Tool (Case-Profiling Tool) in 2009, when the State agency's PER surpassed the 6-percent threshold for possible financial liabilities and corrective action planning. At the same time, Utah transitioned from its legacy system to a modernized, integrated eligibility system for multiple health and human services programs, including SNAP. The State agency developed the Case-Profiling Tool in Utah's electronic eligibility system to increase accuracy and reduce its PER. The machine learning model–based tool later evolved to serve the secondary purpose of reviewing the accuracy of individual staff who processed SNAP cases.

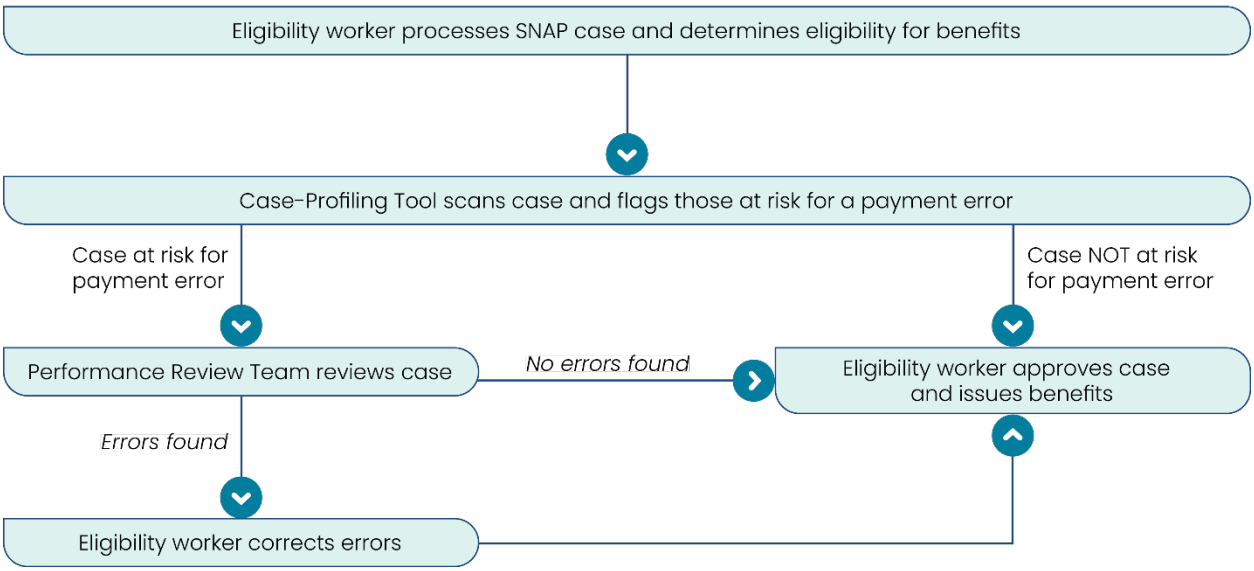
Utah's Case-Profiling Tool flagged all SNAP cases with one or more of the following characteristics for a second review:

- Households with eight or more people
- Any earned, unearned, or self-employment income
- Assets valued over $300
- Medical expenses
- Dependent care expenses
- Child support expenses

The State agency had determined that cases with any of those characteristics were more at risk for a payment error. The agency also requested that the Case-Profiling Tool flag cases approved for $600 or more in SNAP benefits because a mistake made on a case with a relatively high benefit value would have a greater impact on the State's overall PER.

The Case-Profiling Tool flagged SNAP cases for review after an eligibility worker processed the case and before benefits were issued (figure 3.5). The Case-Profiling Tool automatically added SNAP cases to the queue of Utah's Performance Review Team, a centralized unit that reviewed cases across the State and for all programs (e.g., SNAP, TANF, Medicaid), which assessed the cases for accuracy. Both the eligibility worker and their supervisor received an automated email notification when the Performance Review Team found an error, and they could access a report to review the errors and the Performance Review Team's suggested corrections.

Figure 3.5. Use of Utah's Performance Review Case-Profiling Tool



The State agency experienced years of high accuracy and low PERs. However, in 2024, Utah discontinued the Case-Profiling Tool in its original form and revised it to pull a random sample of SNAP cases for the Performance Review Team. Eligibility workers requested that the tool no longer be used to gauge their performance because it pulled the most complex cases rather than a random sample.

1. Tool Development and Implementation

The State agency collaborated with the Utah Department of Technology Services to design and program the Case-Profiling Tool. To inform the design, the State agency reviewed SNAP QC data and determined which characteristics put cases at higher risk for a payment error. The results of the analysis revealed the most at-risk cases were those that included one or more of the characteristics listed in the previous section. The State agency requested the Department of Technology Services program the tool to flag those cases in Utah's electronic eligibility system.

After programming the Case-Profiling Tool, the Department of Technology Services collaborated with the Performance Review Team and State agency staff to test the tool's ability to correctly identify and flag at-risk cases. Respondents recalled that testing revealed the Case-Profiling Tool both (1) flagged cases that did not fit the criteria and (2) missed cases that met the criteria. Adding a layer of complexity, Utah was transitioning from a legacy system to an integrated eligibility system for SNAP, Medicaid, TANF, and other health and human services programs. Respondents recalled some challenges transferring SNAP cases from the old system to the new one and ensuring all household information was retained. The Department of Technology Services continued refining the tool's programming to address the issues found until the Case-Profiling Tool performed as expected in the new eligibility system.

The State agency adjusted its Case-Profiling Tool and associated procedures over time. Although respondents could not recall every alteration, one change involved the specific thresholds for each characteristic used to flag cases. For instance, one respondent said the tool was originally programmed to flag cases with more than $800 in earned income, then more than $100 in earned income, and eventually any amount of earned income. These adjustments occurred as the State agency reviewed annual QC data and the characteristics of cases with errors evolved. The same respondent said that the State agency's review of the QC data also led it to change the threshold used to flag cases by the number of household members, but the respondent could not recall the specifics.

> *"I think we did that right as we did not build these variables into the system in a way that they are not changeable. You will need to evolve through this process. What your needs might be today might not be what your needs are tomorrow, and you need to build it in such a way that allows you flexibility."*
>
> **—Utah respondent**

Over time, the State agency also adjusted some of the procedural steps for using the Case-Profiling Tool. Originally, cases flagged by the Case-Profiling Tool were held for 24 hours to give the Performance Review Team time to review and correct any errors in advance of benefit issuance. However, the high volume of cases the tool flagged meant the Performance Review Team could not always review a case before the 24-hour hold expired and benefits were issued.[25] If the Performance Review Team reviewed the case and caught an error after that 24-hour period, State agency staff corrected payments the following month. In 2024, Utah discontinued the 24-hour hold process after staff voiced concerns about the impracticality of reviewing all flagged cases within such a short timeframe. *"Very few times … would [the Performance Review Team] get to the case within that hold period,"* said one respondent.

2. Impact of the Tool

When the Case-Profiling Tool was in use between 2009 and 2024, the State agency used a reporting tool in the eligibility system that tracked (1) the number of cases flagged that were found to contain errors and (2) the number of cases for which incorrect payments were stopped before issuance (i.e., cases caught during the 24-hour hold period). The latter report was eventually phased out, but the State agency still uses a reporting tool to track overall case errors.

On Staff Workloads

Before implementing the Case-Profiling Tool, Utah's Performance Review Team reviewed 10 cases per month per eligibility worker to confirm the accuracy of eligibility determinations and benefits. The cases that the Performance Review Team reviewed spanned all programs in the

[25] In addition to reviewing the cases the Case-Profiling Tool flagged, the Performance Review Team was responsible for reviewing 10 cases per month per eligibility worker to ensure all staff were accurately processing SNAP cases.

integrated eligibility system. The Case-Profiling Tool added more SNAP cases to the Performance Review Team's queue for review, and staff struggled to keep up with the higher volume.

Eligibility workers typically did not interact directly with the Case-Profiling Tool because it operated in the background of the eligibility system. However, the tool affected their employee performance reviews. The Performance Review Team pulled a sample of 10 cases per month per worker to review for accuracy, prioritizing the cases the Case-Profiling Tool flagged. Because the cases flagged as at risk for a payment error were typically more complex (i.e., larger households with multiple types of income), workers were being graded on the most difficult cases rather than a representative sample.

Following years of low PERs and requests from eligibility workers not to use the Case-Profiling Tool for performance reviews, the State agency discontinued the tool in its original form in 2024. As of the time of data collection, a new version of the tool pulls a random sample of SNAP cases for the Performance Review Team rather than using specific criteria to flag cases at risk for a payment error. Respondents indicated the State agency will continue using this version of the tool to (1) ensure SNAP cases are processed accurately and (2) assess the performance of eligibility workers based on a more accurate representation of their caseloads.

On the Payment Error Rate

Respondents noted that Utah's PER decreased during the years the State agency used the Case-Profiling Tool (i.e., 2009–2024). *"I don't remember the last year that we were at risk for a sanction since the implementation of this tool,"* remarked one respondent. However, the study team could not determine the extent to which the Case-Profiling Tool was responsible for that decrease because Utah implemented concurrent changes that may also have had an impact. One change was the transition to a new integrated eligibility system. In addition, during the years it used the Case-Profiling Tool, the State agency created an employee recognition program for eligibility workers who demonstrated 100 percent accuracy in the cases reviewed by the Performance Review Team. These and other initiatives make it impossible to draw a direct correlation between the use of the Case-Profiling Tool and the State agency's PER.

E. Virginia's SNAP Cross Check and Discrepancy Tool

Virginia's SNAP Cross Check and Discrepancy Tool launched in 2023 after State agency consultants noticed the most common mistakes made by eligibility workers also appeared as errors in the SNAP QC data and quality assurance reviews. The State agency developed the checklist-based tool and embedded it within the eligibility system to reduce errors and ensure that eligibility workers correctly processed initial certifications, recertifications, and case changes, such as adding a household member.

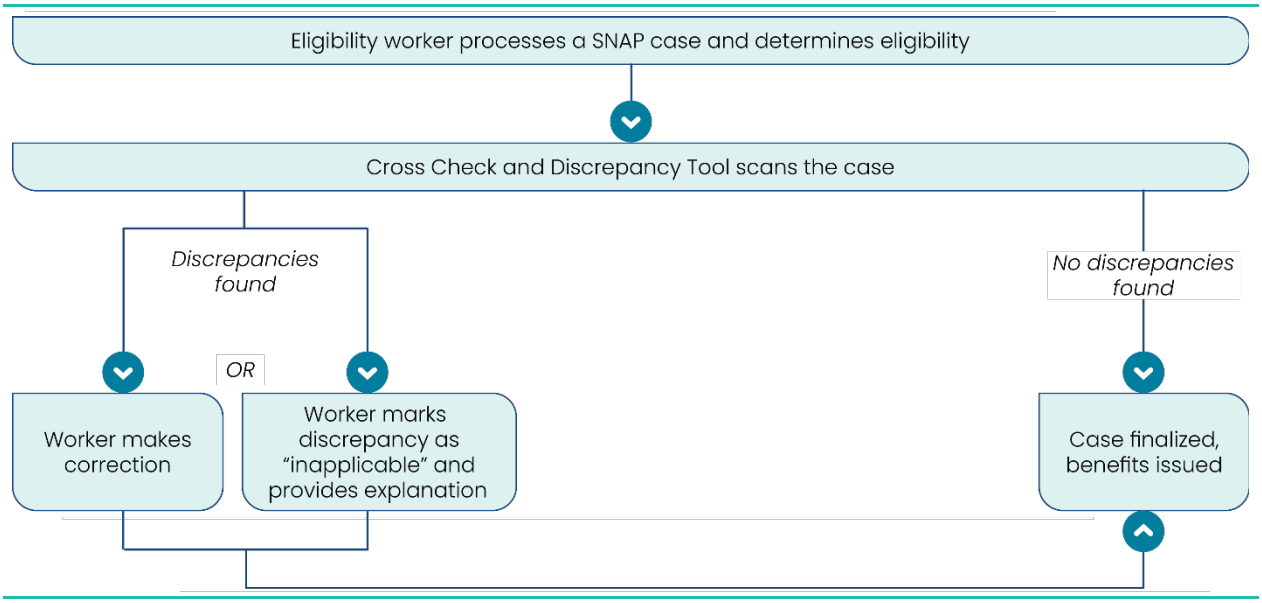
After an eligibility worker enters all relevant case data into the eligibility system, the Cross Check and Discrepancy Tool scans the case and alerts the worker to potential errors. Table 3.1 lists a sample of the potential errors the tool looks for when scanning a case.

Table 3.1. Sample of potential errors flagged by the SNAP Cross Check and Discrepancy Tool

Potential error
■ Expenses listed but no income reported ■ Employment status changed but no change to income ■ New employment reported but no changes made to dependent care expenses when a child younger than age 6 lives in the home ■ New adult household member listed but no income reported for them ■ New disability reported but no disability-related unearned income (e.g., Social Security Disability Insurance) reported

When the tool flags a potential error, the eligibility worker must either (1) correct mistakes or (2) mark the alert as "inapplicable" and leave a comment with an explanation (figure 3.6). Eligibility workers cannot finish certifying a case or issue benefits until they address the items that the tool flagged.

Figure 3.6. Use of Virginia's SNAP Cross Check and Discrepancy Tool



1. Tool Development and Implementation

The State agency submitted a formal work request to its contractor to develop the Cross Check and Discrepancy Tool in August 2021, and the tool launched statewide in August 2023. The State agency based its determination of the potential sources of error that the Cross Check and Discrepancy Tool would look for on analyses that SNAP QC and quality assurance staff

conducted. None of the respondents had been involved in those analyses and therefore could not clearly describe the decision-making process.

After identifying the potential errors the tool should address, the State agency collaborated with a contractor to program the Cross Check and Discrepancy Tool and embed it in the eligibility system. The State agency selected this contractor because it had prepared a similar tool for another State. Throughout development, the State agency communicated with the contractor to discuss the specifications of its request and answer any questions. Virginia's IT department also collaborated with the contractor to create test scripts and scenarios to ensure the programming logic and overall functionality operated as intended. Once the test scripts and scenarios were finalized and the programming was complete, the State agency conducted user testing. As one respondent described, *"They test it in our test environments that are a mockup of our actual production system to see how it will work."* SNAP eligibility workers did not test the tool.

The State agency has a monthly videoconference with eligibility workers statewide to provide feedback and general updates. State agency staff introduced the Cross Check and Discrepancy Tool during the August 2023 call, delivering a Microsoft PowerPoint presentation on the purpose of the tool and how to use it. This was the only training provided to eligibility workers. Some respondents noted that the training was not as effective or informative as it should have been and that new eligibility workers are likely unaware of the existence of the training materials. Regarding the effectiveness of the training, one respondent described it as "a high-level training" that was not tailored to eligibility workers.

> *"With eligibility, we need detailed training. And what we find [is] most of our training is high-level training. It's geared more towards directors or people who are not actually going to touch the case."*
>
> **—Virginia respondent**

Despite testing the Cross Check and Discrepancy Tool before launch, respondents noted two ongoing issues. First, eligibility workers can bypass the cross-checks. When the tool alerts an eligibility worker to a potential error and the eligibility worker chooses to mark the alert as inapplicable, the tool prompts them to leave a comment with an explanation. However, the tool does not assess the quality of the comment, and some eligibility workers simply enter "N/A" or "X" into the comment field. One respondent said eligibility workers *"put 'N/A' so [they] can get off the screen"* because they need to meet with other clients waiting to be seen. Second, two respondents noted the Cross Check and Discrepancy Tool is not user-friendly for eligibility workers and does not always accurately read the information entered. One respondent explained that the tool unnecessarily checks for errors at times. For

> *"I think really having local workers, frontline workers involved in creating tools and understanding what effects it has on the outcome and the impact that it has on our workload is so important. I know we have to have these tools and they're great, but they also need to look at the impact and how it should flow throughout the process."*
>
> **—Virginia respondent**

example, the system requires eligibility workers to go through the Cross Check and Discrepancy Tool any time an eligibility worker wants to create a verification checklist for an applicant with an incomplete case. The State agency plans to rectify these issues, but respondents could not specify a timeline for making changes. Some respondents said these issues might have been addressed earlier had eligibility workers been involved with the development or testing of the Cross Check and Discrepancy Tool.

2. Impact of the Tool

On Staff Workloads

Two respondents said the Cross Check and Discrepancy Tool adds burden and delays processing. Part of the burden is that the tool indiscriminately cross-checks cases whenever a worker touches them rather than focusing on the ones being certified or recertified for SNAP benefits. One respondent said the tool cross-checked a case an eligibility worker was closing for a deceased client. Another source of frustration is how the tool handles dual-certified SNAP and Medicaid cases in Virginia's integrated eligibility system. The tool requires eligibility workers processing changes to the Medicaid portion of a dual-certified case to resolve cross-checks and discrepancies for SNAP. Issues such as these place additional burden on staff and delay case processing at a time when Virginia's local agencies are already understaffed.

On the Payment Error Rate

The primary goal of the Cross Check and Discrepancy Tool is to reduce case errors, some of which may lead to payment errors. However, no monitoring or reporting tool exists to enable the State agency to estimate the impact of the tool on its overall PER. Virginia's PER cannot be directly correlated with the introduction of the Cross Check and Discrepancy Tool because the State agency has implemented concurrent strategies to reduce errors. For instance, the State agency regularly shares information on the State's PER and QC data findings with eligibility workers and supervisors to help those staff understand the types of errors found.

Eligibility worker supervisors can use a reporting tool in the eligibility system to monitor two aspects of tool use by eligibility workers:

1. Types of potential errors flagged by the RA tool
2. An eligibility worker's selection of "inapplicable" in response to a potential error and the explanation they provide in the comment field

However, respondents say eligibility worker supervisors seldom use this reporting tool.

Putting Virginia's PER in context, respondents suggested the State agency's chronic shortage of eligibility workers and eligibility worker supervisors contributes to higher case error rates and PERs. They reported that a lack of eligibility workers overall leads to higher caseloads for the ones who remain. *"They're missing elements of cases because they're trying to process and get things done … because they have so many cases to process at this point,"* said one respondent. Respondents also said the State agency lacks knowledgeable staff to adequately train new workers, and the local agencies lost a high volume of experienced staff during COVID-19.

> *"We have agencies right now that are full of workers that maybe have less than three years of experience. We have supervisors in the agencies that have less than three years of experience. … And there's just so many agencies that are so new, and this is overwhelming. So, I think that feeds a lot into our error rate."*
>
> **—Virginia respondent**

F. Wisconsin's Pre-Certification Review Tool

In FY 2017 and FY 2018, Wisconsin's PER surpassed the 6-percent threshold for possible financial liabilities and corrective action planning. To improve accuracy and reduce the PER, Wisconsin created the Pre-Certification Review Tool in 2019 to identify SNAP cases at risk for a payment error. Originally created as a checklist in Microsoft Word, Wisconsin later embedded the Pre-Certification Review Tool in its electronic eligibility system to identify at-risk cases and flag them for a second review. The Pre-Certification Review Tool considers all SNAP cases undergoing certification, a 6-month review, and recertification after an eligibility worker makes a preliminary eligibility determination but before benefits are issued.

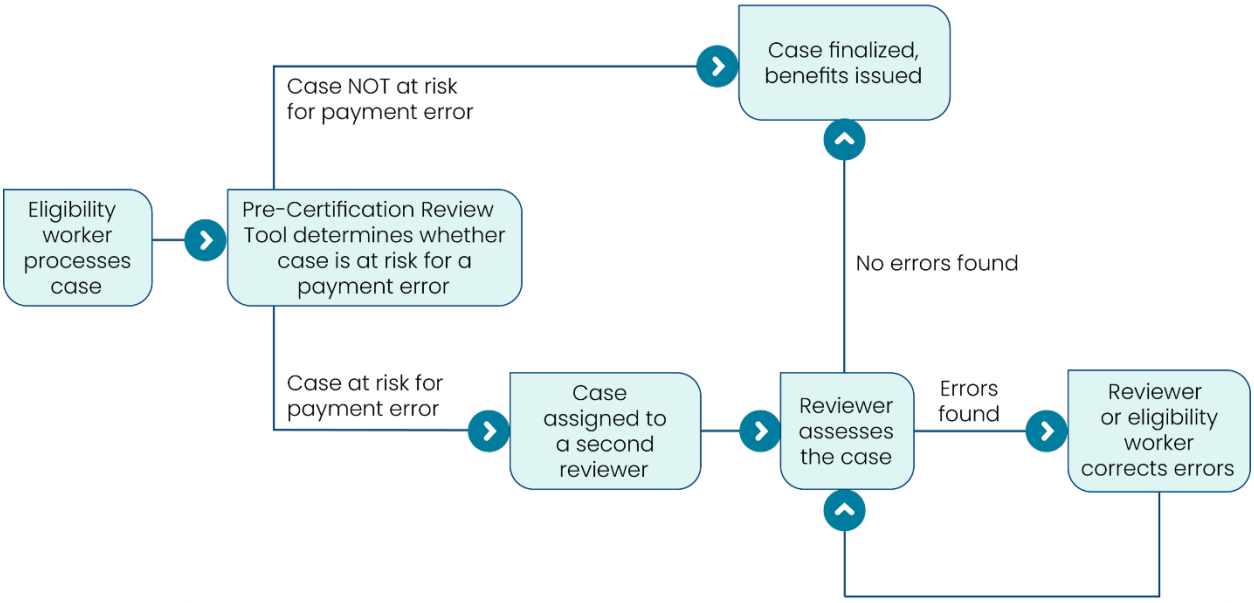
The criteria used to identify and flag cases at risk for a payment error have changed over time and vary by each consortium (see textbox), and that evolution is discussed in greater detail in the next section. Regardless of the criteria used, the process for using the electronic Pre-Certification Review Tool to assess a case and correct any errors follows the same steps (figure 3.7). After an eligibility worker processes a SNAP case, the Pre-Certification Review Tool determines whether it meets the criteria for cases at risk for a payment error, as specified by each consortium. When an application meets the criteria, a warning message in the eligibility system alerts the eligibility worker that a Pre-Certification Review is needed, and the system disables the worker's ability to confirm the benefit issuance amount until that review is complete. The case then moves to a reviewer's

What Are "Consortia?"

In Wisconsin, groups of county-level agencies join to form consortia to streamline and unify administrative efforts and share resources. Eleven consortia administer economic support programs, including SNAP, across the State.

dashboard, and the system alerts the reviewer that a case needs their attention. The reviewer assesses the entire case and searches for errors. If they find no errors, the case status changes to "Review Completed," and the eligibility worker approves SNAP benefit issuance. If the reviewer finds errors, the case status updates to "Corrective Actions Needed," and either the eligibility worker or the reviewer makes the corrections. Some cases, based on the reviewer's discretion, may go through an additional review to ensure all identified errors were corrected. After all errors are corrected, the case status changes to "Review Completed," and benefits are issued.

Figure 3.7. Use of Wisconsin's Pre-Certification Review Tool



Source: Wisconsin's Eligibility Worker Training Manual, 52.2.1. Introduction to Pre-certification Reviews for FoodShare

Each consortium, and sometimes individual eligibility worker supervisors, has the authority to customize the electronic Pre-Certification Review Tool for its caseloads. For instance, one consortium might set the tool to flag cases approved for more than $100 per month in SNAP benefits, whereas another sets the threshold at $200 per month in benefits. Supervisors can program the Pre-Certification Review Tool to flag for review all cases entered by specific staff, such as new eligibility workers.

At the time of data collection in 2024, Wisconsin's PER was below the 6-percent threshold again, and the State agency did not require consortia to use the Pre-Certification Review Tool. As one respondent reported, *"[the State] would love for all agencies to utilize it. But right now, because our error rates are doing fairly well, it's not mandatory."* Some consortia still voluntarily use the tool.

1. Tool Development and Implementation

Initially, the State agency created the Pre-Certification Review Tool as a checklist in Microsoft Word. The checklist itemized three case characteristics that staff determined were associated with higher rates of payment error after reviewing SNAP QC data:

- Households with three or more people
- Any earned income
- SNAP benefit amount of $100 or more

As eligibility workers processed SNAP certifications and recertifications, they used the Pre-Certification Review Tool to review each for the three criteria. If a case met all three criteria, the worker emailed a secondary reviewer (typically an eligibility worker supervisor) to notify them that a case needed their review. The reviewer then conducted a full case assessment and checked for errors. Upon completing their assessment, the reviewer returned the application to the eligibility worker to fix any errors found, finalize the case, and approve benefit issuance.

Respondents reported that when the Pre-Certification Review Tool existed as a Microsoft Word checklist, eligibility workers and supervisors inconsistently used it. Respondents also noted that the review process was cumbersome and prone to delays because reviewers struggled to conduct timely case assessments due to competing priorities and a high volume of cases flagged. Ultimately, some cases at risk for a payment error were processed without a second review.

Transitioning to an Electronic Tool

To embed the process more consistently into the workflows of eligibility staff and reduce processing times, State agency staff began conceptualizing a transition from the Microsoft Word checklist to an electronic tool integrated within the eligibility system in 2021. They worked with the contractor that built and maintained Wisconsin's eligibility system to develop two resources:

1. An electronic version of the Pre-Certification Review Tool
2. A data dashboard to monitor the outcomes and impacts of the electronic Pre-Certification Review Tool

Multiple respondents remarked on the positive working relationship between the State agency and its contractor.

To identify which data points to include in the electronic Pre-Certification Review Tool, SNAP QC staff analyzed findings from the State agency's error rate reports (provided by FNS). QC staff also pulled a random sample of SNAP cases and identified the characteristics of those with the highest payment errors. Following these analyses, State agency staff recommended including the following elements in the tool:

- Number of household members

- Income values: employment-earned income, gross employment earnings, self-employment earnings, countable gross income, child support income, unearned income
- Medical expenses
- Shelter expenses
- Utility expenses
- Dependent care expenses
- Child support payments
- Preliminary approved SNAP benefit amount
- Percentage of total SNAP benefit allotment based on household size
- Case type: certification, 6-month review, recertification, or ongoing

State agency staff then met with the contractor and outlined their goals and specifications. Through multiple rounds of feedback, the State agency and contractor finalized the specifications, and the contractor programmed the tool in the eligibility system.

Both the contractor and State agency staff tested the electronic Pre-Certification Review Tool during the development phase. The contractor used a testing environment with SNAP case data to test the accuracy of the tool; it looked at whether the tool correctly flagged cases that met the error-prone criteria and omitted cases that did not meet the criteria. To test the tool, the State agency formed an internal team composed of agency staff, State QC staff, and eligibility workers and supervisors. Respondents said they included a variety of staff in testing to ensure that all individuals who would use the tool could understand it.

Before launching the electronic Pre-Certification Review Tool statewide, eligibility workers and supervisors were trained to use it. Wisconsin's contractor and State agency staff developed training materials and a presentation, which explained the tool's purpose and included information and demonstrations about how it worked. State agency staff also created short (i.e., five pages or less) tipsheets on how the tool worked with screenshots illustrating the steps to use it. Finally, the State agency added information about the electronic Pre-Certification Review Tool to staff training manuals to ensure all new eligibility workers would be trained to use it.

Following staff trainings, the electronic Pre-Certification Review Tool launched statewide in April 2022. After the first year of implementation, State agency staff met with some of the consortia to solicit their input on possible changes to the tool. The consortia staff requested the tool be amended to include school enrollment status and immigration status, which they said reflected the cases under their purview more at risk for a payment error. That enhancement was made in 2024.

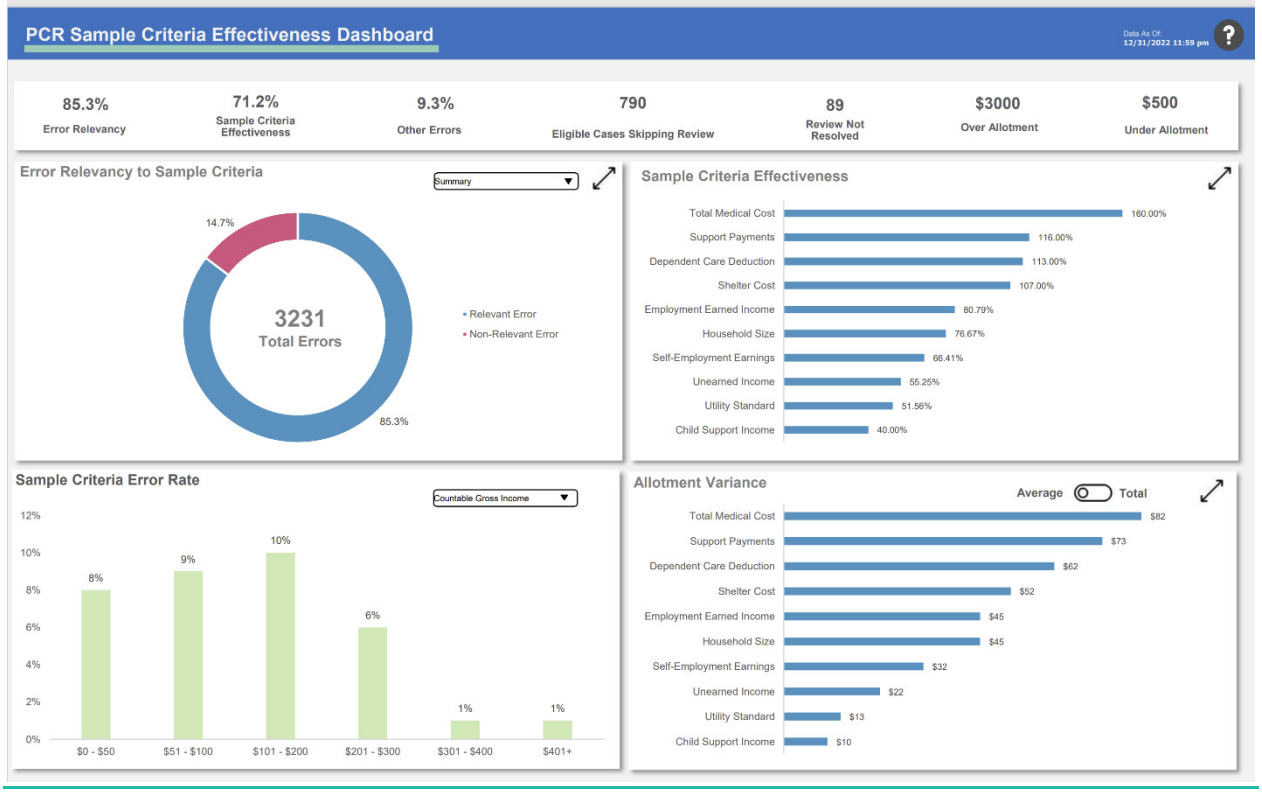
Consortia staff who had served as case reviewers requested the ability to assign themselves cases rather than waiting to be assigned cases by a supervisor. The State agency worked with the contractor to incorporate those changes.

2. Impact of the Tool

State agency staff and the contractor use Tableau dashboards, which pull data from Wisconsin's eligibility system, to track the following effectiveness ratings, outcomes, and errors identified by the Pre-Certification Review Tool:

- Criteria
  - Percentage of errors that correspond to criteria in the Pre-Certification Review Tool (see figure 3.8)
  - Dollar amount of overpayment errors caught
  - Dollar amount of underpayment errors caught

Figure 3.8. Sample dashboard on Pre-Certification Review Tool effectiveness



Note: This screenshot does not include real data.

- Outcomes
  - Number of cases flagged
  - Among cases flagged, the number and percentage of cases found to contain errors
  - Average number of errors among cases flagged

- Percent variance in benefits allotted before and after the Pre-Certification Review Tool review
- Percentage of eligibility workers with a case found to contain errors

› **Errors flagged by reviewer** (see figure 3.9)

- Cause of error (e.g., administrative error versus household reporting error)
- Location of error in a SNAP application
- Top 10 data elements in error
- Length of time flagged cases spend awaiting review

Figure 3.9. Sample dashboard on errors found with Pre-Certification Review Tool

Note: This screenshot does not include real data.

State agency staff can view the dashboard for each consortium and for the State overall, but consortia staff can view only their individual caseloads.

On Staff Workloads

Although the electronic Pre-Certification Review Tool was primarily developed to improve accuracy and reduce payment errors, a secondary goal was to reduce processing times for applications. Anecdotally, respondents reported that the electronic Pre-Certification Review Tool facilitates a timelier review process than the original Microsoft Word version. Under the former

method, flagged cases were sometimes stuck in reviewers' email inboxes either because staff were out of the office or they were struggling to work through the high volume of cases. The electronic Pre-Certification Review Tool tracks how long a case remains in a reviewer's queue and alerts the reviewer when a case has been sitting unread for too long (each consortium can set its requirements for timely reviews as long as they meet the Federal requirements). This facilitates a timelier review and approval process.

The use of the electronic Pre-Certification Review Tool affects the workloads of new eligibility staff the most. Statewide, many agencies flag the cases processed by new staff for a precertification review until at least 85 percent of their cases are processed without error and for the correct payment amounts. Respondents indicated that supervisors can adjust the programming of the Pre-Certification Review Tool to flag all cases processed by new staff, *"so that they all get pulled before [payment is] confirmed."* Overall, staff reported the tool is a useful training mechanism to help new eligibility workers understand how to review SNAP cases. They added that it is intuitive to use and rarely experiences technical issues.

Although the State agency no longer requires consortia to use the electronic Pre-Certification Review Tool, it continues to be available to all and used by some consortia.

On the Payment Error Rate

Wisconsin's PER has remained below six percent since FY 2019, several years before the State agency launched the electronic Pre-Certification Review Tool. Between April 2022 and December 2024, the tool flagged 18,109 cases for review. Among those cases flagged, 4,541 (25 percent) contained at least one case error that staff addressed before SNAP benefit issuance. However, case errors do not always lead to payment errors; a typo in an address would be considered a case error but not one that would change the amount of SNAP benefit payment.

Since the electronic Pre-Certification Review Tool's inception, the State agency estimates the tool has saved Wisconsin approximately $400,000 in overpayments (i.e., when a household is initially approved for a larger SNAP benefit amount than it is eligible to receive). The State agency also estimates the tool has corrected approximately $300,000 in underpayments to Wisconsin households (i.e., when a household is initially approved for a lower SNAP benefit amount than it is eligible to receive).

Chapter 4

RA Tool Effectiveness

This chapter presents two approaches to evaluating RA tool effectiveness. The first approach assesses whether the tools can accurately and efficiently identify cases with payment errors using data from three State agencies. The second approach examines whether RA tool implementation is associated with a lower PER using two different analytic techniques—run chart analysis and difference-in-difference regression modeling. Because of data limitations, both tests of effects on PERs use a single State agency's data.

Given the relatively small number of State agencies with suitable data for performing these analyses, the specific results should not be interpreted as a comprehensive assessment of the effectiveness of existing RA tools. However, the methods described in this chapter could serve as templates for future analyses of RA tool effectiveness. Additional details about the methods are presented in appendix A.

Key Takeaways

- States can use simple analyses to evaluate the performance of their tools in selecting cases for review, although decisions about metrics and interpretations should be informed by both policy and statistical principles.
- Of the three tools analyzed in this chapter, two fail to identify most cases with payment errors as needing review. The third tool selects many cases without payment errors for review.
- Limitations to the availability of SNAP QC data, associated with changes in QC policy and practices, and the timing of RA tool implementation prevented the study team from broadly examining the effectiveness of RA tools in lowering PERs. However, the difference-in-differences model provides some evidence to support the hypothesis that implementing an RA tool lowered one State agency's PER.

A. Evaluating Tool Performance in Selecting Cases for Review

The first effectiveness measure examines the tool's performance in accurately and efficiently identifying cases with payment errors. If an RA tool fails this test, it is unlikely to ultimately accomplish the goal of improving a State agency's PER. To examine effectiveness, the study team applied the parameters of each State agency's RA tool to the public SNAP QC data[26] for the year before the tool was implemented. The study team then compared the tool's predictions with SNAP QC error findings. The analyses are limited to Kansas, Missouri, and Rhode Island, which use tools that apply the same sets of parameters to all cases.[27]

This section focuses on two measures:

- **Sensitivity** is the proportion of cases with a true payment error that are flagged as high risk. A high sensitivity score indicates the tool selects cases that are likely to have an error for review, leading to more accurate payments for participants and lower PERs for State agencies.
- **Specificity** is the proportion of cases without a payment error that are flagged as low risk. A high specificity score indicates that the tool does not create undue additional work for State agency staff by selecting too many cases for review.

Both measures are calculated from a confusion matrix, which compares the SNAP QC result (i.e., whether each case actually had a payment error) with the RA tool result (i.e., whether each case was predicted to have a payment error) (table 4.1).

Table 4.1. Confusion matrix with definitions for SNAP RA tools

RA result	SNAP QC result	
	Payment error **Actual positives:** cases with payment errors	**No payment error** **Actual negatives:** cases without payment errors
High risk **Predicted positives:** cases predicted to have payment errors	**True positives:** cases with payment errors identified as high risk	**False positives:** cases without payment errors identified as high risk
Low risk **Predicted negatives:** cases predicted not to have payment errors	**False negatives:** cases with payment errors identified as low risk	**True negatives:** cases without payment errors identified as low risk

Note: QC = quality control

[26] Public QC data are not identical to the data State agencies maintain. For example, FNS's contractor cleans the data. As such, the way an RA tool is observed to function in the public QC data may differ from how it functions in practice for a State agency.

[27] All other State agencies with RA tools either allowed local customization, making it difficult to know precise details of the RA tool as applied to any given case, or did not provide sufficient detail about how their RA tool worked for the study team to replicate the calculations.

1. Kansas

Kansas began implementing the Pre-Authorization Screening Checklist in November 2011. At that time, the tool flagged cases with a household size of three or four, benefit amount of $400 or more (nonprorated), and earned income of more than $400 and/or child support income of more than $400.[28] Kansas uses the Pre-Authorization Screening Checklist after determining eligibility as part of an initial application, a recertification, or a periodic report but before benefits are issued. Cases flagged as high risk undergo a second review by an eligibility worker supervisor. The tool's performance is displayed in table 4.2. Kansas' tool has an estimated sensitivity of 0.16, calculated by dividing the number of predicted high-risk cases with a payment error (13) by the total number of cases with an actual payment error (81). Kansas' tool has a specificity of 0.95, calculated by dividing the number of predicted low-risk cases without a payment error (834) by the total number of cases without a payment error (878). Kansas' tool has high specificity, meaning the tool rarely selects cases that have no payment error for review, making efficient use of reviewers' time. Overall, however, it does not flag many cases, so it misses a relatively high proportion of cases with errors (low sensitivity)—that is, the tool missed 68 of the 81 cases with an actual payment error.

Table 4.2. Confusion matrix for Kansas' Pre-Authorization Screening Checklist

RA tool result	SNAP QC result	
	Payment error (N)[a] **Actual positives:** cases with payment errors	**No payment error (N) Actual negatives:** cases without payment errors
High risk **Predicted positives:** cases predicted to have payment errors	13	44
Low risk **Predicted negatives:** cases predicted not to have payment errors	68	834
Sensitivity: 0.16 (95 percent confidence interval: 0.08–0.24)		
Specificity: 0.95 (95 percent confidence interval: 0.94–0.96)		

Note: QC = quality control

[a] This column includes all cases with a payment error of any amount, not only those with a payment error amount above the threshold to be counted in the official PER.

2. Missouri

Missouri began implementing the FNS Error Prone Trends tool in August 2020. Missouri uses the tool after conducting the interview for certifications and recertifications and on active cases. Missouri conducts a quality assurance review on cases the tool flags. It flags households with earned income greater than $200 and an issuance amount of $200 or greater.

[28] The issuance threshold has been raised since the tool was first implemented.

A key limitation of the analysis on Missouri's tool is that there were many differences between FY 2019, the year of data used for this analysis, and August 2020, when the tool was implemented. As in many parts of the country, the SNAP caseload increased between FY 2019 and FY 2020.[29] Important SNAP policy changes also took place. Under the Families First Coronavirus Response Act (FFCRA) of 2020, State agencies began issuing emergency allotments, and the time limits were suspended for able-bodied adults without dependents (ABAWDs). As a result, there may have been differences in the population participating in SNAP between FY 2019 and FY 2020 or differences in payment accuracy between those years, meaning that results from analyses of 2019 data may not directly apply to what happened in 2020 when Missouri's RA tool was implemented.

This tool has an estimated sensitivity of 0.32, flagging 107 of the 332 cases with payment errors (table 4.3). It had a specificity of 0.91; of the 537 cases without a payment error, the RA tool correctly predicted 487 cases would not have a payment error. Similar to Kansas' tool, Missouri's tool does not select many cases for review overall, meaning it is inaccurate for cases with a payment error (which it often fails to flag), but it is accurate for cases without a payment error (which it seldom flags).

Table 4.3. Confusion matrix for Missouri's FNS Error Prone Trends

RA tool result	SNAP QC result	
	Payment error (N)[a] **Positives:** cases with payment errors	**No payment error (N)** **Negatives:** cases without payment errors
High risk **Predicted positives:** cases predicted to have payment errors	107	50
Low risk **Predicted negatives:** cases predicted not to have payment errors	225	487
Sensitivity: 0.32 (95 percent confidence interval: 0.27–0.37)		
Specificity: 0.91 (95 percent confidence interval: 0.88–0.93)		

Note: QC = quality control

[a] This column includes all cases with a payment error of any amount, not only those with a payment error amount above the threshold to be counted in the official PER.

3. Rhode Island

Rhode Island began implementing the Pre-Auth Review tool in June 2024. This tool flags households with one or more of the following six characteristics:

1. Five or more members
2. Earned income greater than $1,400 per month

[29] The average number of households enrolled per month in Missouri increased by 14 percent, going from 299,268 in FY 2019 to 340,865 in FY 2020.

3. Unearned income greater than $1,500 per month
4. Any self-employment income
5. Any rental income—owner occupied
6. Benefit amount greater than $200 per month

Rhode Island uses the Pre-Auth Review after determining eligibility and benefit levels as part of an initial application, a recertification, a change report, or a periodic report, but before benefits are issued. Some cases flagged as high risk undergo a second review by an eligibility worker supervisor.

The analysis for Rhode Island has three critical limitations. First, the number of flagged cases Rhode Island reviews depends on the State agency's capacity to conduct the reviews and the time remaining in each case's certification window (the State agency does not want to miss processing timeliness deadlines by holding a case for review). Because Rhode Island does not review all the flagged cases, the study team could not determine how this tool affects households in Rhode Island or Rhode Island's PER; those effects depend on the extent to which Rhode Island acts on the flags the tool produces.

Second, there were some SNAP policy differences between FY 2023, the year of SNAP QC data used for this analysis, and FY 2024, the year Rhode Island began implementing the Pre-Auth Review. These differences include (1) the issuance of emergency allotments under FFCRA, which ended nationwide in February 2023; (2) the suspension of the time limits for ABAWDs under FFCRA, which ended in June 2023; and (3) an increase in the age at which certain adults could be subject to time-limited SNAP benefits under the Fiscal Responsibility Act of 2023 starting in September 2023 (Leftin et al., 2025). If these policy differences resulted in differences in the SNAP caseload between FY 2023 and FY 2024 or differences in payment accuracy between FY 2023 and FY 2024, the estimates of RA tool performance shown here may not be fully applicable to how the RA tool performed in practice.

Third, one of the six case characteristics used in Rhode Island's tool—rental income—is not reported separately from other types of income in the public SNAP QC data. If any cases with rental income did not meet any of the other five criteria in Rhode Island's tool, they were labeled in this analysis as "low risk" instead of "high risk." Therefore, the estimated sensitivity may be lower than the true sensitivity, and the estimated specificity may be higher than the true specificity.

Rhode Island's tool has an estimated sensitivity of 0.86 and specificity of 0.06 (table 4.4). Unlike Missouri and Kansas, Rhode Island's tool flags most cases with payment errors for review, making it more likely to improve the PER. However, because it flags so many cases without payment errors for review, it places a high burden on reviewers' time and attention.

Table 4.4. Confusion matrix for Rhode Island's Pre-Auth Review

RA tool result	SNAP QC result	
	Payment error (N)[a] **Positives:** cases with payment errors	**No payment error (N) Negatives:** cases without payment errors
High risk **Predicted positives:** cases predicted to have payment errors	406	483
Low risk **Predicted negatives:** cases predicted not to have payment errors	64	31
Sensitivity: 0.86 (95 percent confidence interval: 0.83–0.89)		
Specificity: 0.06 (95 percent confidence interval: 0.04–0.08)		

Note: QC = quality control

[a] This column includes all cases with a payment error of any amount, not only those with a payment error amount above the threshold to be counted in the official PER.

B. Evaluating RA Tool Effects on the Payment Error Rate

Another approach to evaluating the effectiveness of State agency RA tools is to measure the effect of implementing an RA tool on PERs using SNAP QC data while controlling for other sources of change. The study team used two analytic techniques for this: run charts and difference-in-difference modeling. Both techniques require pre- and postimplementation data to examine the relationship between implementation and outcomes. For more detail on the methods, see appendix A.

The data available for both analytic techniques are limited by changes to SNAP QC methods, data availability (see chapter 1), and the timing of RA tool implementation. These data challenges limit the availability of usable pre- and post-RA tool implementation data needed to properly evaluate effectiveness. Among the 18 State agencies that currently use or have ever used an RA tool, only one, Minnesota, has an RA tool implementation date that aligns well enough with the available SNAP QC data to provide enough usable pre- and postimplementation data.[30] Minnesota implemented its tool in January 2016 (FY 2016, quarter 2). The study team evaluated the effectiveness of Minnesota's RA tool, examining PERs quarterly between October 2014 (FY 2015, quarter 1) and December 2019 (FY 2020, quarter 1, after which Minnesota discontinued the use of its tool). The study team used a set of six State agencies that did not report having RA tools but that did have available SNAP QC data during the same timeframe as a comparison group.[31]

Because of data limitations, the study team suggests treating the analyses from both techniques as exploratory. The findings from both techniques conform with one another and a plain reading

[30] Minnesota did not provide enough detail on its tool for the study team to evaluate the tool's performance.

[31] The comparison group comprises Arizona, Guam, North Dakota, New York, South Dakota, and the U.S. Virgin Islands. North Dakota reported implementing an RA tool after the timeframe of this analysis.

of Minnesota's quarterly PERs over time. These findings suggest it is possible, though inconclusive, that Minnesota's RA tool decreased quarterly PERs.

1. Run Charts

Before conducting the run chart analysis, the study team determined that a threshold of eight or more consecutive quarterly PERs below the median PER would indicate RA tool success in lowering PERs (see appendix A for more information on the selection of the threshold). Table 4.5 shows the median quarterly PERs between October 2014 and December 2019 for Minnesota and the comparison State agencies. Minnesota's median PER during this time was nominally higher than the median PER of the comparison State agencies. There was no substantial difference in the average PER difference from the median between Minnesota and the comparison State agencies.

Table 4.6 shows the findings from the run chart analysis for Minnesota. Figure 4.1 displays Minnesota's quarterly PERs between October 2014 and December 2019. Based on the threshold of eight quarters, Minnesota did not experience a shift in PERs after it implemented its RA tool. After RA tool implementation, Minnesota saw only three consecutive quarters with a PER below the median PER. During these three quarters, Minnesota experienced PERs that were approximately two percentage points (or about 21 percent) less than the median PER.

Table 4.5. Median PERs October 2014 to December 2019

	N[a]	Median Quarterly PER (%)	Average PER difference from median (% points)
Minnesota (used an RA tool)	1	10.42	0.59
State agencies without an RA tool	6	7.78	0.45

Note: FY = fiscal year; PER = payment error rate

[a] Minnesota and the State agencies that never implemented an RA tool and have published FY 2015 PERs are included. Nine State agencies did not respond to the SNAP Risk Assessment Study Online Survey (including Massachusetts, which has a published FY 2015 PER). They are excluded from these calculations. Although Kansas, West Virginia, and Wyoming have published FY 2015 PERs, they implemented RA tools before October 2014 and are therefore excluded from this analysis. North Dakota implemented an RA tool after December 2019.

Data sources: SNAP Quality Control data; SNAP Risk Assessment Study Online Survey

Table 4.6. Run chart analysis in Minnesota

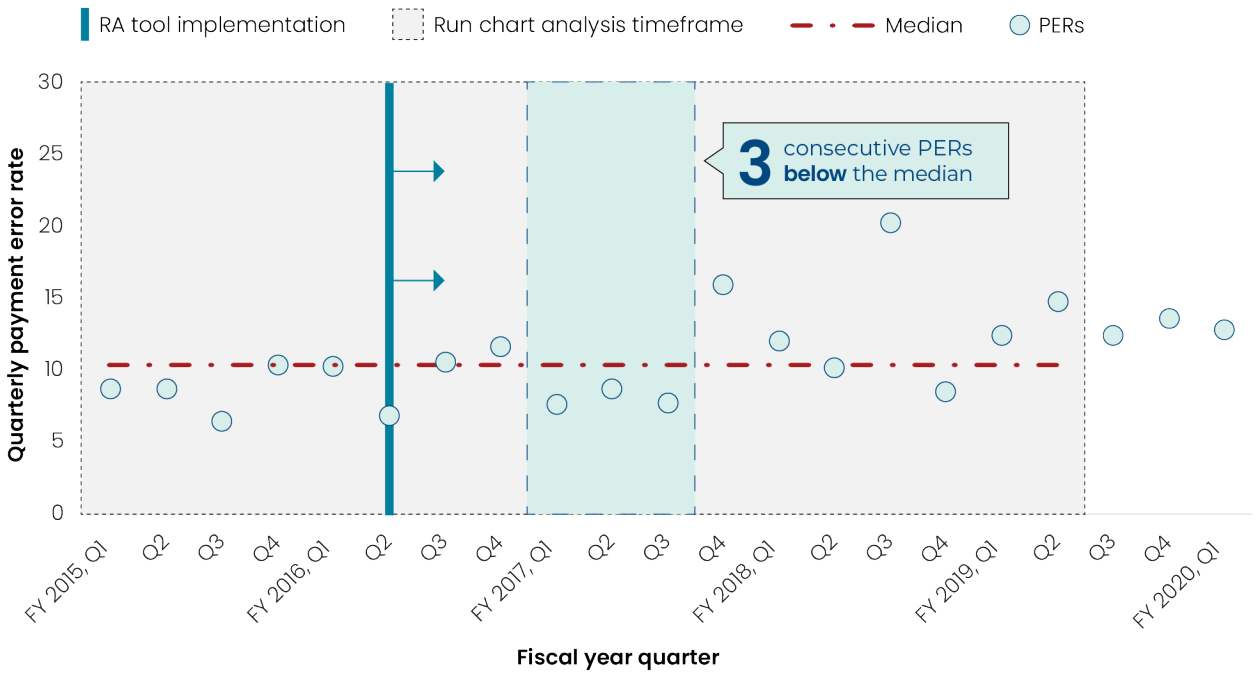
	Available quarters of data	Median PER (%)	Consecutive quarters with PER below the median following RA tool implementation	
			Number of quarters	Average PER difference from median (% points)
Minnesota	18	10.24	3	2.19

Note: The run chart analysis timeframe includes 12 quarters before and 12 quarters after the quarter of RA tool implementation. However, given the SNAP Quality Control data challenges before FY 2015, the study team has only five quarters of data before RA tool implementation for Minnesota. A shift occurs when the quarterly PER is below the median quarterly payment error of the run chart analysis timeframe for eight or more consecutive quarters after the quarter of RA tool implementation. A negative difference value indicates an average PER difference below the median.

PER = payment error rate

Data sources: SNAP Quality Control data; SNAP Risk Assessment Study Online Survey

Figure 4.1. Minnesota's quarterly payment error rates, October 2014 (FY 2015, quarter 1) to December 2019 (FY 2020, quarter 1)



Note: Ideally, the run chart analysis timeframe would include 12 quarters before and 12 quarters after the quarter of RA tool implementation. However, given the SNAP Quality Control data challenges before FY 2015, the study team has only five quarters of data before RA tool implementation for Minnesota. A shift occurs when the quarterly PER is below the median quarterly payment error of the run chart analysis timeframe for eight or more consecutive quarters after the quarter of RA tool implementation. Minnesota did not experience a shift.
FY = fiscal year; PER = payment error rate; Q = quarter
Data sources: SNAP Quality Control data; SNAP Risk Assessment Study Online Survey

Minnesota's run chart analysis provides useful information on the trends in its quarterly PERs during the analysis timeframe. Quarterly PERs were relatively stable between FY 2015 and FY 2017, with some rhythmic variation. By FY 2018, there are more permanent increases in the PER. The run chart analysis for Minnesota does not provide any suggestive evidence of a beneficial effect of the RA tool implementation on the State agency's quarterly PERs. These analyses are limited by the relatively small number of preimplementation quarters relative to the number of postimplementation quarters. They also fail to account for factors that may be influencing quarterly PERs beyond RA tool implementation itself. Finally, Minnesota's consecutive below-median quarterly PERs are plausibly within the prior pattern of regular decreases and increases.

2. Difference-in-Differences Model

Difference-in-differences modeling can estimate an overall average change in an outcome measure in response to a treatment. Difference-in-differences modeling is a quasi-experimental methodology that relies primarily on the assumption that an outcome would continue on the same trend absent a treatment, typically known as the parallel trends assumption. This

assumption enables difference-in-differences models to produce causal estimates when specified correctly. The difference-in-differences model presented in this section examines changes in quarterly PERs following RA tool implementation, assumes PERs would have continued on the same trend absent the implementation of the RA tool, and accounts for unique State agency characteristics.[32]

As with the run chart analysis, the study team pooled quarterly data from Minnesota and the six comparison State agencies across all quarters of October 2014 to December 2019 to fit a longitudinal difference-in-differences model. The study team compared quarterly PERs from Minnesota to PERs among the comparison State agencies, using all six comparison State agencies as the control group to maximize sample size and statistical power of the model. Appendix A provides additional technical details on the difference-in-differences methodology the study team used for this study.

Table 4.7 shows the estimated impact of Minnesota's RA tool on its quarterly PERs. The model detected that Minnesota's tool reduced its quarterly PERs by just over a percentage point (1.37 percentage points), a result that was highly significant. This result is not necessarily indicative of the impact of RA tools generally. This result should also be considered in the context of the limited available data and feasible model specifications.

Table 4.7. Impact of Minnesota's RA tool on its PERs

Minnesota	n	Estimate	SE	CI	p-value
Average change in the quarterly PER following RA tool implementation	140	−1.37	0.22	−1.80, −0.94	< 0.001

Note: CI = confidence interval; PER = payment error rate; SE = standard error
Data sources: SNAP Quality Control data; SNAP Risk Assessment Study Online Survey

Unlike the run chart analysis, the difference-in-differences model for Minnesota provides some suggestive evidence of a beneficial effect of its RA tool implementation on the State agency's quarterly PERs. The model found a highly significant but modest reduction in Minnesota's quarterly PERs after implementing its RA tool. If this is a true impact, it represents a 12-percent decrease in Minnesota's quarterly PER (based on Minnesota's average quarterly PER of 11 percent between October 2014 and December 2019).[33]

Although the difference-in-differences model specifications and framework would generally provide a robust identification strategy and robust causality, the study team cautions against considering this a true impact because of weaknesses in the available data. The study team had only one treatment unit (Minnesota) and six comparison units (the comparison State agencies). The study team also had only four pretreatment quarterly PERs (and one treatment quarter PER) compared with 15 posttreatment quarterly PERs. This produced a model with only 140

[32] See appendix A for details on the covariates the study team used in the difference-in-differences model to account for unique State agency characteristics.
[33] The study team used the average quarterly PER here as a relative comparison point instead of the median because the difference-in-differences model calculates an average change rather than a median change.

observations (20 treatment unit observations and 120 comparison unit observations). The model may have lacked appropriate statistical power. The study team also conducted various sensitivity analyses that suggested the results are likely biased. Appendix A provides additional technical details on the sensitivity analyses used for the difference-in-differences model.

Chapter 5

RA Tool Promising Practices

This chapter draws on findings from all components of the study—the literature review, survey, key informant interviews, and SNAP QC data analysis—to present promising practices for RA tool development, implementation, and evaluation. Thoughtful development of RA tools involves the following steps:

- Robust data analysis
- Collaborative design process
- Transparent programming and design
- Substantial testing
- Continual monitoring
- In-depth and ongoing training for end users

For State agencies considering building an RA tool, a list of considerations and suggestions for monitoring and evaluating RA tools can be found in appendix E. Although the practices in this chapter and appendix E are described in the context of State agencies using RA tools for SNAP administration, these strategies could be used by any health or human services program.

A. Robust Data Analysis

State agencies should analyze available data to identify the characteristics of cases at risk for a payment error to inform the design of RA tools. Among State agencies with RA tools currently in use, survey findings revealed that most (80.0 percent) used SNAP Quality Control data to inform the development of their tools and over half (66.7 percent) used other SNAP data (see chapter 2 for more details). The findings from these analyses can supplement staff's institutional knowledge and suggest ways to design an RA tool to efficiently and effectively lower a State's PER. Appendix E provides specific questions for State agencies to consider during their analysis

of SNAP data and strategies to identify cases at risk for a payment error that may benefit from additional review.

Both the case studies and the literature (Corbett-Davies & Goel, 2018; The White House, 2022) indicate that State agencies should regularly review available data to amend the tool over time to ensure alignment with (1) shifts in the characteristics of cases at risk for a payment error and (2) current SNAP policies and procedures. The frequency with which an RA tool should be recalibrated would likely be specific to each tool and the data it uses.

B. Collaborative Design Process With Technical and Programmatic Perspectives

The case studies suggest that including a comprehensive range of perspectives in RA tool design, development, and testing may facilitate a smooth implementation process. The purpose and format of the tool can help determine the types of staff to include. For instance, an RA tool that takes the form of a paper checklist may require a different combination of staff involvement compared with an algorithmic tool embedded within a State agency's SNAP eligibility system. However, States should consider involving each of the following types of staff in the design, development, and testing:

- SNAP State agency program staff
- Data and statistical experts (e.g., SNAP QC staff)
- State IT staff (and external contractors, if appropriate)
- Tool's potential end users

The SNAP State agency program staff, data experts, and statistical experts should be able to identify the characteristics of cases at risk for a payment error. Involving State IT staff, external contractors, or both is critical when conceptualizing, programming, and testing an algorithmic tool built within a State agency's eligibility system. The State agency will need to collaborate with those staff to strategize how to program the electronic tool to achieve its goals while also aligning with SNAP policies and procedures. Consistent with this promising practice, the survey findings reveal that most State agencies enlisted SNAP State program or policy staff in the development (93.3 percent) and testing (86.7 percent) of their RA tools. However, among State agencies currently using an algorithmic or other electronic RA tool, less than half (45.5 percent) involved IT staff in designing and/or building the tool; most (63.6 percent) involved IT staff in testing.

It is also critical to involve the potential end users of an RA tool in design and testing because they can help craft the programming logic and procedures for using the tool in a way that users can easily understand, without adding a significant burden to staff workloads. For instance, if eligibility workers and supervisors will use the tool, these types of staff should be involved in designing and testing the RA tool. The case studies illustrate both the benefits of involving end

users in tool development (Wisconsin) and the obstacles that can arise during implementation when end users are not consulted (Virginia).

C. Transparent Programming and Design

When an algorithmic or other electronic RA tool is built by a State IT department or in collaboration with an external contractor, the SNAP State agency needs sufficient visibility into the tool's programming to confirm that the tool functions as desired and aligns with SNAP policy. In several examples described in this report, State agency staff identified their goals for RA tools and specified the data sources and variables to include—and then relied on programmers to build it. This approach sometimes led to issues with a tool's performance. For example, the case studies revealed instances of misunderstandings between the programmers and the State agency staff that led to the tool (1) flagging the wrong cases or (2) missing error-prone cases it should have flagged.

When the State agency has visibility into the programming of an electronic RA tool, it can more easily identify problems and communicate potential solutions to programmers. An added benefit of that visibility is that State agency staff will be better positioned to respond to policy changes or updated QC findings because they will be able to communicate to programmers which features of their RA tool need updating. For example, a State agency may originally build its tool to flag cases with earned income values over $800 per month, but subsequent analyses of SNAP QC data reveal that it should flag all cases over $100 per month. In this example, State agency staff could more efficiently respond to the QC results by either clearly communicating to programmers precisely where to make the adjustment or making the adjustment themselves.

This promising practice echoes a similar recommendation from the literature: be transparent in the design of an RA tool. A 2022 expert panel (The White House, 2022) proposed a list of items for tool developers to disclose, including the following:

- Overview of the system and its goals
- Description of data used, including how data are processed and interpreted
- Summary of what data might be missing, incomplete, or erroneous and data relevancy justifications
- Risk identification and steps taken to mitigate potential harms

For SNAP RA tools, the State agency program staff are likely the ones responsible for crafting the overview of the tool and its goals, and the other items may be written in collaboration with data analysts and IT staff (including contractors, when relevant).

D. Performance Testing Before Implementation

Mirroring the recommendations from the literature (Schwartz et al., 2022; Turner-Lee et al., 2019; The White House, 2022), the case studies illustrate the importance of testing RA tools before launch. Respondents also point to two levels of testing that should take place: accuracy and usability. All six case study agencies tested for accuracy, but only two tested for usability.

In testing the accuracy of an RA tool, State agencies ensure that it flags SNAP cases according to the specifications. For instance, if the State agency stipulates that the tool should flag cases with four or more household members, which the data indicate are more at risk for a payment error, the agency should ensure that the RA tool does not flag smaller households. Testing for accuracy also means verifying the programming logic for electronic tools. One Rhode Island respondent indicated that they learned during the pilot stage of their RA tool that the contractor that built it had incorrectly programmed the tool to flag errors for all cases with the status "no further actions required." Testing for accuracy is important to ensure the tool functions as intended and directs staff to focus their time on the cases truly at risk for a payment error.

Respondents across case study agencies indicated that usability testing was too often overlooked. Only two of the six case study agencies conducted usability testing (Virginia and Wisconsin), and only one of those involved the end users in usability testing. Respondents from Wisconsin spoke about extensive user testing that involved multiple types of staff, the contractor, and the end users (i.e., eligibility workers). The State agency intentionally involved the end users because it wanted to ensure that the people who would ultimately use the tool could understand it. Respondents from Virginia indicated that the State agency staff and its contractor conducted usability testing but did not involve the end users (i.e., eligibility workers and supervisors). Ultimately, this worked against the State during implementation because eligibility workers and supervisors found the tool was time-consuming and occasionally imprecise.

The other State agencies did not conduct usability testing, and some respondents noted it was to their detriment when functionality and usability challenges arose during implementation. A lack of user testing also made it difficult for some State agencies to create comprehensive and coherent training materials on RA tools. Respondents from Kansas, for instance, pointed to inconsistencies across training materials as one cause for the inconsistent use of the tool by end users.

E. Continuous Monitoring of Tool Performance

The literature also highlights the need for ongoing testing and evaluation of RA tools after implementation. Schwartz et al. (2022) note that systems may perform differently in preimplementation testing than they do once they are in the field; continuous testing and monitoring are necessary to catch and mitigate potential issues early. Regardless of how the tools perform when first implemented, RA tools should be recalibrated over time to ensure they

continuously adapt to population and contextual changes to maintain accuracy (Corbett-Davies et al., 2018; The White House, 2022).

The case studies also emphasize the need for continual testing, evaluation, and refinement of an RA tool to catch and mitigate potential issues early and be responsive to new trends in QC findings or staff feedback. To that end, State agencies should consider how they want to facilitate monitoring after implementation. When collaborating with an external contractor to build and program an electronic RA tool, Wisconsin requested that the contractor also create a data dashboard that would enable the State agency to monitor staff use of the tool and its effectiveness at flagging error-prone cases. The State agency also met with eligibility workers and supervisors one year after implementation of the Pre-Certification Review Tool to solicit their feedback and then worked with the contractor to implement the adjustments requested. Creating similar feedback loops between the State agency and the workers using the tool in the field can help identify potential improvements and sources of staff confusion or burden.

Appendix E offers specific considerations for monitoring and evaluating RA tools—tasks that require data State agencies may not already collect. Therefore, when developing an RA tool, State agencies should lay out their plans for monitoring and evaluation and make sure their eligibility systems (or other systems) capture all the required data for monitoring and evaluation.

F. In-Depth and Ongoing Training for End Users

It is critical for staff to be trained on how to properly apply the RA tool to cases and what to do when cases get flagged. According to the survey responses, State agencies' training on RA tools most commonly involved online, prerecorded trainings, and some provided written training materials and/or in-person trainings. Several case studies illustrated that when staff receive inadequate training to use a tool, they may use it inconsistently or find ways to circumvent it entirely. As Brauneis and Goodman (2018) explain, it is also important for staff to understand the flags the tool creates, particularly for individuals who use a tool to guide decision making. In the SNAP context, it could be helpful to share with end users the proportion of cases flagged that are expected to have payment errors and the proportion of cases not flagged that are expected to have payment errors.

Some case study agencies found that the RA tools also helped support staff training efforts. In Utah, after a reviewer used the Case-Profiling Tool to assess the entirety of a case, they emailed the original eligibility worker to notify them of any errors. That notification email listed the specific errors found and the suggested corrections. State agency staff said these details helped the eligibility worker learn from their mistakes and, they suspected, prevented similar mistakes from occurring in the future. Similarly, respondents from Wisconsin reported their RA tool is a useful training mechanism for new eligibility workers learning to review SNAP cases because they learn from the errors found by the staff who review their cases.

References

Anhøj, J. (2015). Diagnostic value of run chart analysis: Using likelihood ratios to compare run chart rules on simulated data series. *PLOS One*, *10*(3), Article e0121349. https://doi.org/10.1371/journal.pone.0121349

Anhøj, J., & Olesen, A. V. (2014). Run charts revisited: A simulation study of run chart rules for detection of non-random variation in health care processes. *PLOS One*, *9*(11), Article e113825. https://doi.org/10.1371/journal.pone.0113825

Borusyak, K., Jaravel, X., & Spiess, J. (2024). Revisiting event-study designs: Robust and efficient estimation. *Review of Economic Studies*, *91*(6), 3253–3285. https://doi.org/10.1093/restud/rdae007

Brady, P. J., Harnack, L., Widome, R., Berry, K. M., & Valluri, S. (2023). Food security among SNAP participants 2019 to 2021: A cross-sectional analysis of current population survey food security supplement data. *Journal of Nutritional Science, 12,* Article e45. https://doi.org/10.1017/jns.2023.32

Brauneis, R., & Goodman, E. P. (2018). Algorithmic transparency for the smart city. *Yale Journal of Law & Technology, 20*, 103–176. https://yjolt.org/algorithmic-transparency-smart-city

Callaway, B., & Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, *225*(2), 200–230. https://doi.org/10.1016/j.jeconom.2020.12.001

Canbek, G., Taskaya Temizel, T., & Sagiroglu, S. (2022). PToPI: A comprehensive review, analysis, and knowledge representation of binary classification performance measures/metrics. *SN Computer Science*, *4*(1), 13. https://doi.org/10.1007/s42979-022-01409-1

Corbett-Davies, S., Gaebler, J. D., Nilforoshan, H., Shroff, R., & Goel, S. (2018*). The measure and mismeasure of fairness*. arXiv:1808.00023v3. http://dx.doi.org/10.48550/arXiv.1808.00023

Food and Nutrition Service. (2020). *Supplemental Nutrition Assistance Program (SNAP) Quality Control Data: Quality control data access*. U.S. Department of Agriculture. https://snapqcdata.net/datafiles

Food and Nutrition Service. (2025). *Supplemental Nutrition Assistance Program: Payment error rates, fiscal year 2024*. U.S. Department of Agriculture. https://fns-prod.azureedge.us/sites/default/files/resource-files/snap-fy24QC-PER.pdf

Hand, D. J. (2012). Assessing the performance of classification methods. *International Statistical Review*, *80*(3), 400–414. https://doi.org/10.1111/j.1751-5823.2012.00183.x

Leftin, J., Monkovic, M., Yang, F., Maula, J. C., & Bonelli, S. (2025). *Technical documentation for the fiscal year 2023 Supplemental Nutrition Assistance Program Quality Control database and the QC Minimodel: Final report*. U.S. Department of Agriculture, Food and Nutrition Service. https://snapqcdata.net/sites/default/files/2025-03/FY%202023%20Tech%20Doc.pdf

National Institute of Justice. (2021). *2021 Review and revalidation of the first step act risk assessment tool*. U.S. Department of Justice. https://www.ojp.gov/pdffiles1/nij/303859.pdf

Oregon Audits Division. (2018). *Supplemental Nutrition Assistance Program (SNAP) fraud investigations: Advisory report*. Secretary of State. https://sos.oregon.gov/audits/Documents/2018-18.pdf

Perla, R. J., Provost, L. P., & Murray, S. K. (2011). The run chart: A simple analytical tool for learning from variation in healthcare processes. *BMJ Quality & Safety, 20*(1), 46–51. https://doi.org/10.1136/bmjqs.2009.037895

Polek, C. & Sandy, S. (2023). The disparate impact of artificial intelligence and machine learning. *Colorado Technology Law Journal, 21*(1), 85–108. https://ctlj.colorado.edu/?p=958

Rambachan, A., & Roth, J. (2023). A more credible approach to parallel trends. *Review of Economic Studies*, *90*(5), 2555–2591. https://doi.org/10.1093/restud/rdad018

Roth, J. (2022). Pretest with caution: Event-study estimates after testing for parallel trends. *American Economic Review: Insights*, *4*(3), 305–322. https://doi.org/10.1257/aeri.20210236

Roth, J., Sant'Anna, P. H. C., Bilinski, A., & Poe, J. (2023). What's trending in difference-in-differences? A synthesis of the recent econometrics literature. *Journal of Econometrics*, *235*(2), 2218–2244. https://doi.org/10.1016/j.jeconom.2023.03.008

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence, 1*(5), 206–215. https://doi.org/10.1038/s42256-019-0048-x

Rudin, C., Chen, C., Chen, Z., Huang, H., Semenova, L., & Zhong, C. (2022). Interpretable machine learning: Fundamental principles and 10 grand challenges. *Statistic Surveys*, *16*, 1–85. https://doi.org/10.1214/21-SS133

Ryan, A. M. (2018). Well-balanced or too matchy–matchy? The controversy over matching in difference-in-differences. *Health Services Research*, *53*(6), 4106–4110. https://doi.org/10.1111/1475-6773.13015

Sant'Anna, P. H. C., Callaway, B., & Quispe, A. (2024). CSDID documentation. https://d2cml-ai.github.io/csdid/index.html

Sant'Anna, P. H. C., & Zhao, J. (2020). Doubly robust difference-in-differences estimators. *Journal of Econometrics*, *219*(1), 101–122. https://doi.org/10.1016/j.jeconom.2020.06.003

Schwartz, R., Vassilev, A., Greene, K., Perine, L., Burt, A., & Hall, P. (2022). *Towards a standard for identifying and managing bias in artificial intelligence* (NIST Special Publication 1270). U.S. Department of Commerce, National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.1270

Turner-Lee, N., Resnick, P., & Barton, G. (2019). *Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms*. Brookings. https://www.brookings.edu/research/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/

U.S. Department of Agriculture. (n.d.). *Non-discrimination statement*. https://www.usda.gov/non-discrimination-statement

Venkatadri, G., Sapiezynski, P., Redmiles, E. M., Mislove, A., Goga, O., Mazurek, M., & Gummadi, K. P. (2019, May 13–17). *Auditing offline data brokers via Facebook's advertising platform* [Conference presentation]. Association for Computing Machinery WWW '19: The World Wide Web Conference, New York, NY, United States. https://doi.org/10.1145/3308558.3313666

The White House. (2022). *Blueprint for an AI bill of rights*. https://bidenwhitehouse.archives.gov/ostp/ai-bill-of-rights/

Appendix A

Detailed Methods

The study team used four methods to answer the research questions and objectives of this study. These methods included an expansive and extensive literature review; a survey (and related analyses) of SNAP State agency directors about RA tools; case studies (and related analyses) of six States that currently use, or previously used, an RA tool; and analyses of SNAP QC data to determine RA tool effectiveness. This appendix provides information on these methods and presents some additional findings for the survey and RA tool effectiveness analyses.

A. Literature Review Methodology

This section details the methodology for the literature review.

- Develop search terms.
  - First, the study team reviewed government websites, news stories, peer-reviewed publications, and other sources to identify terms and phrases used to describe concepts related to RA.
    - Search terms included the following:
      - (screening OR assessment OR analysis OR algorithm OR "machine learning") (bias OR discrimination OR equity) (tool OR score OR model) (risk OR profile OR demographic) (SNAP | CMS | childcare | Federal)
      - (screening OR assessment OR algorithm OR "machine learning") (bias OR discrimination) (risk OR profile OR demographic) (legal | justice | judge | bail)
      - (screening OR assessment OR algorithm OR "machine learning" OR AI) (bias OR discrimination OR equity) (tool OR score OR model) (risk OR profile OR demographic) (data | computer)

- Second, the study team ran searches in Google and Google Scholar, iterating on the search terms to yield relevant results.

- Screen results of final searches for relevance.
- Download relevant results and code full texts using NVivo.
  - The study team developed a preliminary codebook based on the study objectives and research questions.
  - The codebook was tested on two of the selected documents and refined to yield a final codebook. The study team used the final codebook to analyze all the literature.

B. Survey Development, Fielding, and Results

This section provides information about the methods of survey development and data collection for the SNAP Risk Assessment Study Online Survey, including survey results.

1. Survey Development and Pretesting

Survey development involved drafting items designed to measure and address the research objectives outlined in the study plan. Once drafted, the study team selected eight respondents for pretesting, as displayed in table A.1. The respondents were recruited from Westat Insight's Nutrition Assistance Research Advisory Panel, whose members are former SNAP State agency employees who provide insight into the planned data collection from the State perspective and enable the study team to pretest the data collection instruments with experienced State agency staff without burdening potential respondents. Other pretesting respondents included Westat Insight employees and consultants who are former SNAP National Office, State agency, or local agency staff and one former local SNAP agency staff member.

Table A.1. Pretest respondents

Study instrument	Pretest respondents
Online survey	■ Tom Hedderman, former New York State SNAP director ■ Patricia McGinn, former Colorado agency training supervisor and FNS Regional Office QC team lead ■ Michael Ribar, former District of Columbia SNAP policy lead, current senior technical adviser at Westat Insight ■ Brian Jones, former Kentucky SNAP QC director, current senior technical assistance lead at Westat Insight ■ John Knaus, former USDA FNS chief of program design, current senior fellow at Westat Insight

Note: QC = quality control

The survey lead emailed respondents a Word version of the survey. Respondents were asked to review the instrument and provide written feedback regarding survey instructions, wording and terminology, clarity, and flow of questions. After respondents fully reviewed the survey, they were also asked to estimate how long they thought it would take a State respondent to complete it.

2. Survey Fielding

Westat Insight programmed and fielded the survey using Qualtrics. Recruitment involved a sequence of emails, telephone follow-up, and customized targeted email reminders to specific respondent groups (table A.2). Recruitment began on July 8, 2024, and concluded on November 29, 2024.

Table A.2. Recruitment schedule

Date	Recruitment
July 8, 2024	FNS prenotice email
July 8, 2024	Qualtrics invitation email
July 15, 2024	Qualtrics first reminder email
July 29, 2024	Qualtrics second reminder email
August 7, 2024	First telephone follow-up
August 13, 2024	Qualtrics third reminder email
August 19, 2024	Second telephone follow-up
August 23, 2024	Qualtrics fourth reminder email
August 29, 2024	FNS follow-up reminder
September 5, 2024	Qualtrics fifth reminder email
September 10, 2024	Custom Qualtrics reminder email ■ Partial completes ■ Nonrespondents
September 10–12, 2024	Third telephone follow-up
September 18, 2024	Qualtrics sixth reminder email
October 3, 2024	Fourth telephone follow-up
October 8, 2024	Qualtrics seventh reminder email
October 18, 2024	Qualtrics eighth reminder email
October 23–25, 2024	Fifth telephone follow-up
November 5–7, 2024	Sixth telephone follow-up
November 6, 2024	Qualtrics ninth reminder email (last)

3. Survey Results

Forty-four of 53 eligible State agencies[34] completed the SNAP Risk Assessment Study Online Survey—a response rate of 83.0 percent.[35] Following is the breakdown by FNS Region and SNAP caseload size from July 2024.

[34] Includes all 50 States; Washington, DC; Guam; and the Virgin Islands. New York City responded for New York State.
[35] Although Arizona did not click the submit button to complete its survey, the agency did respond to all substantive survey questions, so the Westat Insight and FNS study team decided to include Arizona's response data and categorize them as complete.

Response Rates by Region

Response rates varied greatly by FNS Region (table A.3). All agencies in the Mountain Plains, Western, and Southwest Regions responded to the survey. All but one agency in the Mid-Atlantic and Midwest Regions completed the survey, for a response rate of 85.7 percent in each Region. The Regions with the lowest response rates were the Northeast and Southeast, with only half of their agencies completing the survey.

Table A.3. Survey response rate by FNS Region

FNS Region	Agencies invited	Agencies completed	Response rate (%)
Mid-Atlantic	7	6	85.7
Midwest	7	6	85.7
Mountain Plains	8	8	100.0
Northeast*	8	4	50.0
Southeast	8	4	50.0
Southwest	7	7	100.0
Western	8	8	100.0
Overall response rate	53	44	83.0

* New York City responded instead of New York State. New York is a county-administered SNAP State agency, and local social services districts are responsible for administering SNAP, not New York's Office of Temporary and Disability Assistance. The New York SNAP State director gave the study team the contact information for New York City's social services district so that its staff could respond to this survey.

Response Rates by SNAP Caseload Size

Response rates also varied by the size of each State agency's SNAP caseload[36, 37] (i.e., the number of SNAP participants); see table A.4. For analysis purposes, the study team used the first and fourth quartiles to categorize agencies' caseload sizes as small, medium, and large: Small agencies have fewer than 148,987 (first quartile) participants (n = 13), medium agencies have between 148,988 and 871,087 participants (n = 27), and large agencies have more than 871,088 (fourth quartile) participants (n = 13). Agencies with large caseloads had the lowest response rate (69.2 percent) compared with agencies with medium- and small-sized caseloads (92.6 and 76.9 percent, respectively).

Table A.4. Survey response rate by SNAP caseload size

Caseload size (number of individuals)	Agencies invited	Agencies completed	Response rate (%)
Small (< 148,987)	13	10	76.9
Medium (148,988–871,087)	27	25	92.6

[36] State agency caseload data: Food and Nutrition Service. (n.d.). *SNAP data tables: National and/or State level monthly and/or annual data, FY 69 through FY 24*. U.S. Department of Agriculture. https://www.fns.usda.gov/pd/supplemental-nutrition-assistance-program-snap

[37] New York City caseload data: New York State Office of Temporary and Disability Assistance. (2024). *Temporary and disability statistics: July 2024*. https://otda.ny.gov/resources/caseload/2024/2024-07-stats.pdf

Caseload size (number of individuals)	Agencies invited	Agencies completed	Response rate (%)
Large (> 871,088)	13	9	69.2
Overall response rate	53	44	83.0

Survey Completion Rate

Overall, the survey had a high completion rate among respondents (97.6 percent). The study team calculated the survey question completion rate for each responding agency by dividing the number of completed questions by the total number of questions the agency could have completed.[38] The study team then calculated the average percentage of completed eligible questions for the four groups defined by the survey screener question (table A.5). Through skip logic, the screener question directed each respondent only to the questions that applied to their experiences (or lack thereof) with an RA tool. The four screener groups were the following:

- Our agency currently uses an RA tool.
- Our agency previously used an RA tool but has since discontinued it.
- Our agency developed but never implemented RA tools.
- Our agency has never developed or implemented RA tools.

It is important to note that the number of questions presented to each agency varied greatly depending on its response to the screener question. Agencies that never implemented an RA tool were presented as few as three questions, whereas agencies that currently use a tool were presented as many as 102. Despite the significantly higher respondent burden, agencies that currently use a tool averaged a 97.6-percent question completion rate.

Table A.5. Average survey question completion rate by survey screener group

Survey screener group	N	Average percentage complete
Our agency currently uses an RA tool	15	97.6
Our agency previously used an RA tool but has since discontinued it	3	100.0
Our agency developed but never implemented RA tools	1	100.0
Our agency has never developed or implemented RA tools	25	97.1
Average completion rate for individual respondents	N/A	97.6

Note: N/A = not applicable

Some survey questions were identified as critical to the study and programmed to require a response, whereas other questions were optional, allowing respondents to skip them. Analysis of skipped survey questions revealed that only one question (A11: economic data elements) exhibited a pattern in item nonresponse, with four agencies not responding to several income-related data elements. This finding may suggest that some agencies struggle to answer questions

[38] The number of questions each respondent was able to answer varied because of skip logic.

related to income and assets in relation to their RA tools, processes, or both. Table A.6 provides more detail on the variation in completion rates specific to each respondent.

Table A.6. Survey question completion rate for 44 responding State agencies

State agency	Eligible survey questions	Completed survey questions	Completion rate
Alabama	3	3	100.0
Alaska	3	3	100.0
Arizona	7	3	42.9
Arkansas	102	102	100.0
California	6	6	100.0
Colorado	5	5	100.0
Connecticut	77	75	97.4
Guam	3	3	100.0
Hawaii	3	3	100.0
Idaho	74	69	93.2
Illinois	5	5	100.0
Indiana	85	84	98.8
Iowa	7	7	100.0
Kansas	76	76	100.0
Kentucky	83	83	100.0
Louisiana	3	3	100.0
Maine	3	3	100.0
Maryland	4	4	100.0
Minnesota	8	8	100.0
Mississippi	3	3	100.0
Missouri	72	72	100.0
Montana	3	3	100.0
Nebraska	78	73	93.6
Nevada	3	3	100.0
New Jersey	5	5	100.0
New Mexico	93	93	100.0
New York City*	5	5	100.0
North Carolina	4	4	100.0
North Dakota	71	67	94.4
Ohio	3	3	100.0
Oklahoma	4	4	100.0
Oregon	5	5	100.0
Pennsylvania	7	6	85.7
Rhode Island	80	75	93.8
South Dakota	6	6	100.0
Texas	4	4	100.0
Utah	9	9	100.0

State agency	Eligible survey questions	Completed survey questions	Completion rate
Virgin Islands	3	3	100.0
Virginia	83	77	92.8
Washington	71	71	100.0
Washington, DC	5	5	100.0
West Virginia	89	89	100.0
Wisconsin	98	98	100.0
Wyoming	8	8	100.0
Total	1,364	1,331	97.6

* New York City responded instead of New York State.

Lastly, the study team tallied the number of agencies that reported currently using an RA tool and noted whether they uploaded supporting documents in the survey. Of the 15 State agencies that reported current use of an RA tool, eight did not upload supporting documents to describe their tools.

Nonresponse Bias Analysis

Table A.7 shows the results of the survey nonresponse bias analysis.

Table A.7. Survey nonresponse bias analysis results

Variable	Value		Significantly different ($p \leq 0.05$)
	Respondents (n = 44)	Nonrespondents (n = 9)	
Average administrative cost per SNAP case per month ($)	33	26	No ($p = 0.21$)
FNS Region (%)			
Mid-Atlantic	14	11	Yes ($p = 0.03$)
Midwest	14	11	
Mountain Plains	18	0	
Northeast	11	33	
Southeast	9	44	
Southwest	16	0	
Western	18	0	

Variable	Value		Significantly different (p ≤ 0.05)
	Respondents (n = 44)	Nonrespondents (n = 9)	
Program administration (%)			
County administered	23	0	No (p = 0.26)
State administered	77	100	
Average State agency annual payment error rate (%)	12	13	No (p = 0.57)
Average SNAP caseload size (number, thousands)	405	513	No (p = 0.56)
Average State or territory population (number, thousands)	6,112	7,362	No (p = 0.61)

C. Case Studies

This section details the methodology used to conduct the case studies.

1. State Agency Selection

The study team sent FNS the Active Tool Summary Memo on September 9, 2024, that outlined preliminary results from the SNAP Risk Assessment Study Online Survey to support FNS's selection of State agencies for case studies.

The State agencies Westat Insight originally recommended for case studies were Arkansas, Connecticut, Kansas, New Mexico, and Wisconsin. These agencies represented a mixture of FNS Region, SNAP administration type, monthly caseloads, benefits issued, PERs, and tool use timeline.

After a thorough FNS review of Westat Insight's recommendations and current State agency availabilities, FNS approved the following State agencies for case study outreach: Connecticut, Kansas, New Mexico, Utah, Virginia, and Wisconsin. The State agencies Westat Insight interviewed were Connecticut, Kansas, Rhode Island, Utah, Virginia, and Wisconsin; New Mexico was unable to participate.

2. Key Informant Interviews

The study team created and pretested the interview protocols.

The team identified up to six positions for interviews per State agency: RA development lead, quality assurance lead, QC staff, IT staff, local office staff, and local office supervisor. The team held introductory calls with each State agency that agreed to participate to introduce the study, explain staff roles identified, and ask for help reaching out to appropriate staff.

After the State agencies provided names and contact information, the study team reached out to these individuals to schedule interviews.

For each State agency, the team aimed to complete all interviews within a 2-week period. Interviews were conducted in Microsoft Teams and recorded with participants' consent. Table A.8 summarizes the types of respondents and the number of interviews the study team held with each State agency.

Table A.8. Interviews conducted by case study State agency

State agency	Number of interviews conducted				Total
	RA tool development lead	State QA/QC staff	State IT/data analysis staff	Local office staff	
Connecticut	1*	1*	1	1	3*
Kansas	1	2	0	2	5
Rhode Island	1	1	0	0	2
Utah	1	0	2	0	3
Virginia	1	1	0	1	3
Wisconsin	1	2	0	1	4

Note: QA = quality assurance; QC = quality control
*The team conducted a joint interview with the RA tool development lead and State QA/QC staff in Connecticut.

All interviews were professionally transcribed. Analysis began with the development of the coding scheme, which was driven by interview protocols. The study team coded broad topics. After coding the transcripts, the study team reviewed the codes within and across State agencies to identify cross-State agency themes.

D. RA Tool Effectiveness Methodologies

1. Run Chart Analysis Methodology

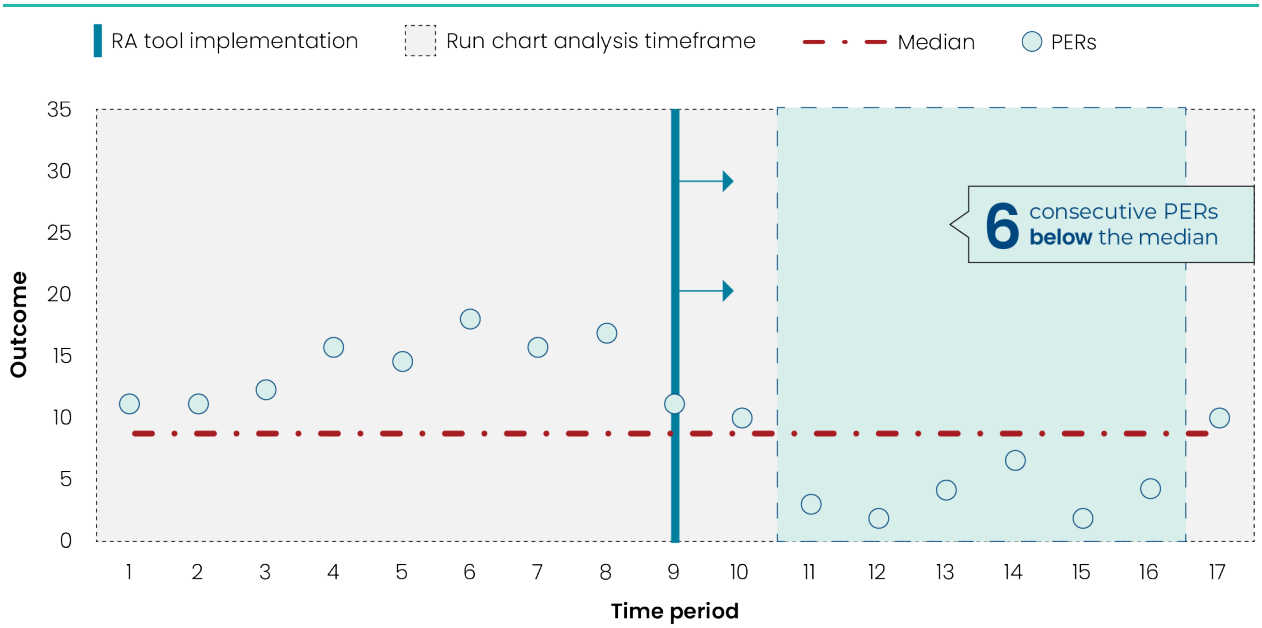
Run charts serve as initial diagnostic tools to show outcomes over time. They are designed to quickly evaluate "shifts" in an outcome after treatment and should be considered exploratory. This section uses a run chart to explore whether Minnesota's implementation of an RA tool in January 2016 is associated with a decrease in the quarterly PER across 18 quarters of data.

Run Chart Methodology

The study team uses run chart analysis as described in Perla et al. (2011), Anhøj and Olesen (2014), and Anhøj (2015). This analysis technique looks for a threshold consecutive number of outcomes above or below the median (depending on the hypothesis) after some treatment. This consecutive number of outcomes is called a shift. Perla et al. (2011), Anhøj and Olesen (2014), and Anhøj (2015) provide some guidance on choosing a shift (e.g., at least six) but largely say the choice is contextual and should be informed by the policy specifics. Run chart analysis also

requires a timeframe for the analysis. It must include sufficient time before the treatment to provide a baseline outcome in the absence of the treatment and sufficient time after the treatment implementation to observe shifts in the outcome that could be caused by the treatment. The median outcome within this timeframe is used when calculating potential shifts. Figure A.1 provides an example of a run chart analysis.

Figure A.1. Example of run chart analysis



Note: The run chart analysis timeframe includes eight time periods before and eight time periods after the treatment. A shift occurs when the outcome is below the median outcome of the run chart analysis timeframe for six or more consecutive time periods after the time period of treatment implementation.
PER = payment error rate

In this study, the outcome of interest is the quarterly PER, and the treatment is RA tool implementation. The study team defined the shift as eight or more consecutive quarterly PERs below the median PER within the analysis timeframe. The team chose eight as the threshold for the shift to—

- Allow State agencies time to gain proficiency in using their tool.
- Reduce the likelihood that any changes in quarterly PERs around the time of RA tool implementation were not coincidental.
- Capture RA tool performance under changing SNAP populations (whether from new certifications, recertifications, or case churn).

The study team defined the analysis timeframe as 12 quarters before and 12 quarters after the quarter of RA tool implementation (for a total of 25 quarters, including the treatment quarter). The team chose this timeframe to allow for full certification periods of SNAP cases and SNAP case churn and to reduce the quarter-specific effects on PERs (e.g., an unexpected case spike or

staffing shortage in a quarter). However, given the SNAP QC Data challenges before FY 2015, only five quarters of data before RA tool implementation for Minnesota are available.

2. Difference-in-Differences Model Methodology

This section details additional methodological details related to the difference-in-differences modeling used to estimate an overall average change in the quarterly PER following RA tool implementation.

Model Specification

Difference-in-differences compares the outcomes of a treatment and control groups before and after implementation of a treatment. As Roth et al. (2023) explain, traditional difference-in-differences models assume the following:

- All units receive treatment (in this case, RA tool implementation) at the same time (identical treatment timing).
- The average change in an outcome for treated and untreated units will be the same in the absence of treatment (parallel trends).

However, not all State agencies implemented RA tools at the same time. The study team hypothesized that parallel trends likely occur within State agencies but not across State agencies (because of circumstances specific to a State agency). That is, there are State agency–specific time trends in PERs. Failing to account for staggered treatment timing leads to "forbidden comparisons" between treated units in the present time period and already-treated units in earlier time periods. When the parallel trends assumption is violated, treatment effects are biased for failing to account for the counterfactual no-treatment outcome.

To account for staggered treatment timing and violated parallel trends, the study team used the Callaway and Sant'Anna (CS) estimator (2021) to estimate the difference-in-differences model. The CS estimator does not make any of the forbidden comparisons. Treated units are compared with never-treated units and not-yet-treated units. The study team used both never-treated and not-yet-treated units as the comparison group because there are State agencies that implemented an RA tool after the timeframe of the usable data (October 2014 to December 2019). The CS estimator also uses Sant'Anna and Zhao's (2020) doubly robust (DR) method to condition parallel trends on relevant covariates.[39] Although staggered treatment timing is irrelevant because only one treatment unit (Minnesota) was used, the study team continued to use the CS difference-in-differences estimator because of how it accounts for parallel trends.

An alternative estimator, the Borusyak, Jaravel, and Spiess (BJS) estimator (Borusyak et al., 2024), exists for robust difference-in-differences estimation. However, the study team did not use it

[39] DR methods use a two-stage model, with the first stage being a propensity score model of the treatment and the second stage being an outcome model. DR methods are (statistically) consistent if at least one of these models is specified correctly.

because of concerns of PER serial correlation and parallel trend violations over a longer time horizon.[40]

The study team assumes no anticipatory effects. Although State agencies and their staff would likely anticipate the adoption of any RA tool (and possibly influence behaviors and subsequently PERs), the team reasons that State agency workers actually conducting QC reviews likely have little to no incentive to alter their behaviors, knowing that RA tool adoption is coming. Although a tool may improve worker productivity, workers, to the study team's knowledge, face no benefit or cost by having their work before tool implementation prospectively mirror their work after tool implementation.

The study team also clustered error terms at the State agency level. Standard regression models assume outcomes are independent of one another. When those outcomes are not independent, specialized regression methods are required. The team has repeated outcomes (PERs) on State agencies that are not independent of one another. For example, State agency *i*'s PER in quarter *x* is not independent from State agency *i*'s PER in quarter *x* + 1. PERs in one quarter are dependent both on current and past conditions (e.g., current State agency capabilities and unresolved prior payment errors). Clustered error terms account for this non-independence of outcomes.

Model Estimand

The CS estimator calculates a series of average treatment on the treated (ATT) effects (the parameter of interest) for all appropriate treatment-comparison pairings. These various ATT effects can then be averaged or otherwise calculated to produce estimands of interest. Generally, the CS estimator can be thought to produce ATT(x), where *x* are the inputs to the estimator. The study team estimates the following:

$$ATT(x) = (Y_{i,t}, X_{i,t}, W_{i,t}, Q, T_i, e_i)$$

—where:

- $Y_{i,t}$ is the quarterly PER for State agency *i* at quarter *t*.
- $X_{i,t}$ is a series of covariates for State agency *i* at quarter *t*, including the lagged quarterly PER and an indicator variable for FY 2018 to FY 2020 versus FY 2015 to FY 2017.
- $W_{i,t}$ is the weight for State agency *i* at quarter *t* (number of completed cases used to calculate the quarterly PER).
- Q is the fiscal year quarter.
- T_i is the fiscal year quarter when State agency *i* implemented its RA tool (0 for State agencies that never implemented).
- e_i is an error term clustered at State agency *i*.

[40] These are reasons to avoid using the BJS estimator, according to Roth et al. (2023).

To calculate the estimate of the average change in the quarterly PER due to RA tool implementation (the estimand), the study team used the weighted average of the treatment effects by State agency. Callaway and Sant'Anna (2021) refer to this type of estimand as the "group" estimand, which returns an average of all ATTs by the "group" of interest (in this case, State agencies). Callaway and Sant'Anna (2021) suggest then averaging the individual by-group estimands, which they characterize as "the average effect of participating in the treatment experienced by all units that ever participated in the treatment" (p. 211). They note that the interpretation of this averaged group estimand is similar to the interpretation of the ATT in the traditional difference-in-differences method (as outlined in Roth et al., 2023). Sant'Anna et al. (2024) recommend using this averaged group estimand as the "overall summary effect of participating in the treatment."

Model Covariates

The study team considered many likely relevant State agency characteristics to include in the model. Some examples were measures of caseload size and administrative costs. However, because the final model had only seven State agencies, variation in these State agency characteristics was insufficient to include them in the model. The study team also had to exclude other State agency–level characteristics that did not vary over time, such as FNS Region and State versus non-State administration of SNAP. Notwithstanding lagged quarterly PER (for the reasons discussed below), the study team excluded any covariates affected by the treatment (RA tool implementation), such as the number of SNAP cases with errors, to avoid endogeneity issues.

Given the limited data, the study team included the following covariates in the model:

- An indicator variable for FY 2018 to FY 2020 versus FY 2015 to FY 2017. The payment error threshold was $37 for FY 2018 to FY 2020 versus $38 for FY 2015 to FY 2017. Although this difference is small, it could account for some variation in error rates. The time variable in the model (fiscal year quarter) is continuous and assumes a linear trend in PER over time. The time variable itself would not accurately account for changes in the threshold, which are not linear and therefore would not be expected to have linear effects on the PER.
- Lagged quarterly PER. The study team used the previous quarter's PER as a covariate. Lagged outcomes can reduce bias from unobserved confounders (Ryan, 2018)—for example, State staffing changes. Roth et al. (2023) note that including the lagged outcome makes sense if one is confident in the "conditional unconfoundedness assumption"—that is, treatment is as good as randomly assigned, conditional on the lagged outcome and other relevant covariates. Roth et al. (2023) also note that including the lagged outcome may be sensible in settings where treatment takeup decisions are made because of lagged outcomes. The lagged quarterly PERs (and State-level error terms) have the potential to capture changes in State agencies and aberrations in SNAP caseloads over time that may affect PERs. For example, if caseloads spike in a previous quarter,

causing additional errors due to "crunch" on State agency employees, the lagged PER could capture that.[41]

When reviewing the SNAP QC data, the study team had more confidence in PERs calculated with larger samples because they were at less risk of small sample bias and sampling variability. For example, a SNAP case with an extremely large payment error would have more influence on a quarterly PER than an annual PER. The team weighted observations by the number of completed SNAP QC cases in the quarter to account for this confidence.

Sensitivity Analyses

The study team faced limitations when constructing the difference-in-differences model because of weaknesses in the available data. Data from only seven State agencies did not offer enough variation in some possible useful State agency characteristics (e.g., caseload as a percentage of population size or administrative cost per case). This lack of variation caused collinearity issues whereby State agency characteristics effectively acted as duplicate State-agency fixed effects (the model already controlled for State agency as the panel variable). The model therefore could only reasonably control for within-quarter-invariant State characteristics (the State-agency fixed effects), lagged quarterly PER, and the change in payment error threshold. The study team believes this makes the model highly susceptible to unobserved variable bias. For example, when using the Callaway and Sant'Anna (2021) difference-in-differences estimator, the foundational parallel trends assumption of difference-in-differences holds when controlling for relevant variables (e.g., lagged quarterly PER).

Given these limitations, the study team conducted a series of sensitivity analyses to test the robustness of the difference-in-differences model results. The study team tested two additional models and conducted a series of tests to evaluate the robustness of the parallel trends assumption.

Additional Models

The main difference-in-differences model used both "never-treated" and "not-yet-treated" observations as the control group. With only six comparison State agencies available to use, the study team chose this approach for the control group to maximize sample size and statistical power of the model. North Dakota, one of the available comparison State agencies, implemented an RA tool in February 2023. Although the usable data ended in December 2019, there is a possibility that conditions in North Dakota that led to its RA tool implementation may have been present during the analysis timeframe (October 2014 to December 2019) and may have been substantively different from the conditions in the other comparison State agencies. If true, this could bias the results of the main difference-in-differences model. The study team ran an additional difference-in-differences model using just the "never-treated" observations as the

[41] The sensitivity analyses discussed in appendix E suggest that lagged quarterly PER and the indicator variable for FY 2018 to FY 2020 versus FY 2015 to FY 2017 are not sufficient to control for all differences in State agency–specific time trends in PERs. These covariates likely control only for some differences in trends.

control group. The results from this model are qualitatively similar to the main model results (see table 4.7).

The study team ran a second additional difference-in-differences model using the “never-treated" and “not-yet-treated” observations as the control group and the annual PER instead of the quarterly PER as the outcome variable. The annual PER is a more policy-relevant outcome variable than the quarterly PER. FNS officially publishes annual PERs, and official FNS regulations and rules are based on the annual PER. However, the study team chose to use the quarterly PER as the main outcome variable to have more time periods and, therefore, maximize the sample size and statistical power of the model. The study team also had concerns that with fewer time periods, it may not have been possible to capture the effects of RA tools on PERs as precisely. When using the annual PER as the outcome variable, the study team found no impact of Minnesota’s RA tool (see table A.9). This may suggest some measurement bias in the quarterly PER as the outcome variable. However, the annual PER model has a very small sample size, which may be biasing its results.

Table A.9. Impact of Minnesota’s RA tool on their PERs, sensitivity analysis models

Average change in Minnesota’s PER following RA tool implementation	N	Estimate	SE	CI	p-value
Main model: “never-treated" and “not-yet treated” observations as the control group, quarterly PERs	140	−1.37	0.22	−1.80, −0.94	< 0.001
Sensitivity analysis model 1: “never-treated" observations as the control group, quarterly PERs	131	−1.33	0.20	−1.71, −0.94	< 0.001
Sensitivity analysis model 2: “never-treated" and “not-yet treated” observations as the control group, annual PERs	35	0.54	0.38	−0.21, 1.29	0.16

Note: CI = confidence interval; PER = payment error rate; SE = standard error
Data sources: SNAP Quality Control data; SNAP Risk Assessment Study Online Survey

Parallel Trends Robustness

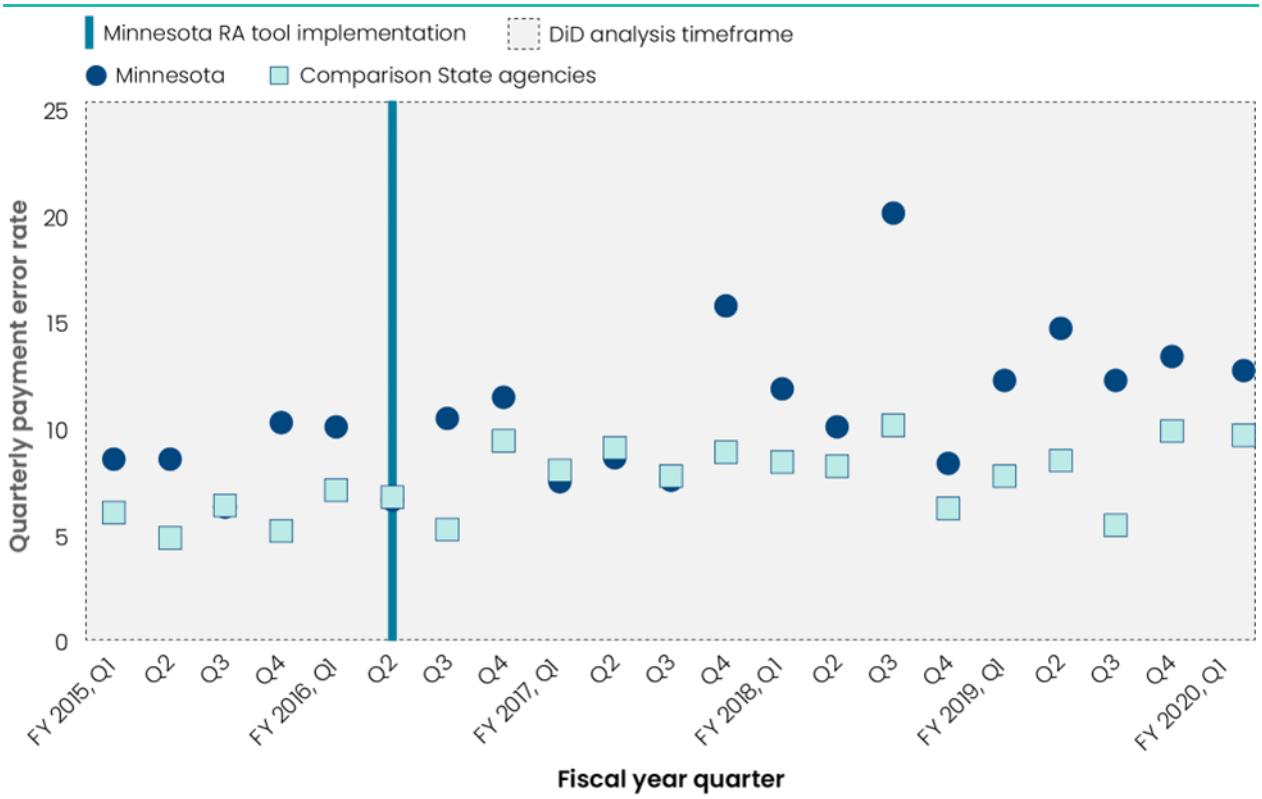
Difference-in-differences is a quasi-experimental methodology that relies primarily on the “parallel trends” assumption that an outcome would continue on the same trend absent a treatment. This feature enables difference-in-differences to produce causal estimates when specified correctly. If this assumption is violated, then any results will be biased and noncausal. In this study, the trends in State agency PERs need to be approximately equivalent prior to any RA tool implementation to comply with the “parallel trends” assumption. Upon review of the by-State agency quarterly PERs in figure A.2, the study team observed a lot of nominal variation in quarterly PER trends between Minnesota and the comparison State agencies before Minnesota’s RA tool implementation. To approximate parallel trends in the pretreatment period, the study team used

the Callaway and Sant'Anna (2021) difference-in-differences estimator to condition parallel trends on relevant covariates.

For the main difference-in-differences model results, the study team tested whether all pretreatment differences between the treatment and control groups were zero. With highly significant results (chi-square = 214, *p*-value = < 0.001), the study team rejected this hypothesis. This suggests that the main difference-in-differences model violated the parallel trends assumption and that the results are biased.

Testing parallel trends in this way does not account for whether the test has enough statistical power to detect relevant violations of the parallel trends assumption. For example, very small variations in quarterly PERs between the treatment and control groups may be acceptable. FNS regulations include a payment error threshold to allow very small payment errors, so it may be reasonable in this study to allow small variations between the treatment and control group pretreatment period outcomes. Using the methodology of Roth (2022), the study team tested for the size of the parallel trends violation that could be detected 99 percent of the time. The results from this test suggest that if there were a linear parallel trends violation of at least 0.29 percentage points, there would be significant pretreatment differences between the treatment and control groups 99 percent of the time (i.e., the parallel trends assumption is violated). Given how small this effect size is, this suggests the main difference-in-differences model does indeed violate the parallel trends assumption.

Figure A.2. Quarterly payment error rates, October 2014 (FY 2015, quarter 1) to December 2019 (FY 2020, quarter 1)



Note: The quarterly payment error rates of the comparison State agencies are the weighted averages of the quarterly payment error rates of the six State agencies that comprise the comparison State agencies group. The study team weighted these rates by the number of completed SNAP QC cases in the quarter.
FY = fiscal year; Q = quarter; QC = quality control
Data sources: SNAP Quality Control data; SNAP Risk Assessment Study Online Survey

The study team also used the methodology discussed in Rambachan and Roth (2023) to estimate the "magnitude" of a parallel trends violation that would invalidate the conclusion of a significant treatment effect. For example, Rambachan and Roth's (2023) methodology could find that to invalidate the conclusion of a significant treatment effect, the study team would need to allow for a posttreatment violation of parallel trends two times larger than the maximal pretreatment violation. This methodology formalizes the pretreatment trend testing intuition: the counterfactual posttreatment trends cannot be "too different" from the pretreatment trends. Rambachan and Roth's (2023) methodology makes clear assumptions about parallel trends violations that must be made as part of accepting a treatment effect as conclusive. The study team found that the "breakdown value" for a significant treatment effect is 0.40, meaning that a significant treatment result is robust to allowing for violations of parallel trends up to 0.40 times the maximum violation in the pretreatment period. Given how small this effect size is, this also suggests the main difference-in-differences model does indeed violate the parallel trends assumption.

Finally, using Rambachan and Roth's (2023) methodology, the study team tested how much the posttreatment violation in parallel trends can vary from the linear extrapolation of the

pretreatment trend. To do this, the study team can impose a constant, C, the maximum amount by which the slope of the pretreatment trend can change across consecutive periods. A value of $C = 0$ implies that the posttreatment trend must be perfectly linear, whereas positive nonzero values of C allow for some nonlinearity in the posttreatment trend. The "breakdown value" for a significant treatment effect is $C \approx 0.0085$, meaning there is not a null pretreatment trend effect (i.e., the parallel trends assumption is violated) without allowing for the linear extrapolation across consecutive periods to be off by more than 0.0085 percentage points. Like the other parallel trends robustness tests, this small effect size suggests that the posttreatment trend must be nearly perfectly linear, a more restrictive condition that makes it more likely that the main difference-in-differences model does indeed violate the parallel trends assumption.

Appendix B

Data Collection Instruments

This appendix contains the survey instrument and each of the key informant interview protocols.

State Agency Director Web Survey

A. Introduction

The U.S. Department of Agriculture's Food and Nutrition Service (FNS) is conducting a study to learn about tools used to identify Supplemental Nutrition Assistance Program (SNAP) cases likely to have a payment error. These tools may be known by different names, such as case-profiling tools, risk assessment tools, or error-prone profiling. After cases are flagged as high risk, they undergo a rigorous process to ensure accurate benefit decisions. FNS hired Westat to conduct this research study to identify SNAP State agencies that use case-profiling tools and to understand the development, effectiveness, and perceived value of the tools.

The information collected during the study will be used to highlight best practices, challenges, and solutions for case-profiling tools. These results will inform the development of case-profiling tools at the Federal level and will help FNS provide resources and technical assistance.

Please complete the survey by *[date]*. You may share the survey link, and other staff in your agency may log in to complete the survey. The survey should take between 15-45 minutes to complete depending on the characteristics of your State. You may log in to the survey as many times as you wish. There is no cost to you to participate apart from the time you spend to complete the survey, and there is no compensation.

Use the << Previous and Next >> buttons at the bottom of the screen to navigate through the survey. Do not use the back button on your web browser as doing so could result in the loss of your answers. If you need to step away from the survey, simply exit your browser. Your answers will save automatically as you progress through the survey. You can resume the survey by following the link we sent to your email inbox.

Participation in this study is voluntary. Refusal to participate will not have any impact on your position, your State agency, or your programs. You may also skip questions you do not wish to answer.

Any personally identifying information obtained will be kept private to the extent provided by law. We will use the data we collect only for the purposes described above. Please note that the final report will present the survey results both in the aggregate and at the individual State level. The raw survey data will be submitted to FNS at the end of the study for research purposes. Any personally identifying information will be removed from the raw survey data submitted to FNS.

If you need additional information, please contact the FNS Project Officer, Eric Williams, at eric.williams@usda.gov or please email us at *SNAPCPSurvey@westat.com* or call *[study phone]*.

Thank you.

B. What Are SNAP Case-Profiling Tools?

Before you fill out the survey, we want to define what we mean when we talk about SNAP case-profiling tools. For the purposes of this study, we define anything that meets the following three criteria as a case-profiling tool:

- The tool is used during the application or recertification process or on active cases.
- The tool identifies the risk of a case having a payment error (i.e., high/low risk, high/medium/low risk, etc.).
- Cases may be treated differently based on their risk category (e.g., high-risk applications are subject to additional scrutiny prior to benefit determination; high-risk active cases are prioritized for a quality assurance [QA] review).

What do case-profiling tools look like? SNAP case-profiling tools can range in sophistication from a simple checklist that helps eligibility workers identify cases likely to have a payment error to complex analytics tools incorporated into a State agency's eligibility system. Three hypothetical examples follow:

Example 1. Checklist of Criteria to Flag Cases for Review

- State agency A determined payment errors were concentrated in large households of six or more people and in households with no reported income. State agency leadership requires a second review for all applications meeting either criterion prior to eligibility and benefit determination. When processing applications, eligibility workers flag all such cases for supervisors to review.

Example 2. Integrated Case-Profile Tool

- State agency B used several years of Quality Control data to develop a machine-learning model that predicts the likelihood a case has a payment error. The model is integrated into the eligibility system, which automatically assigns a risk score to each case as data are entered. If the score is above a certain threshold, the case is flagged as high risk and undergoes a QA review within 30 days of becoming an active case.

Example 3. QA Review Screener and Procedures Manual Update

- State agency C requires supervisors at local offices to conduct at least 10 QA case reviews per month. These reviews are entirely separate from the SNAP QC process. Instead of choosing the 10 cases for QA review at random, the State agency identifies 10 cases at high risk for payment error by passing all active cases through a computerized case-profiling tool. The State agency updated its QA review procedures manual to formally integrate this process into its QA approach.

If you want to ask a clarifying question on what constitutes a case-profiling tool, please email us at *[study email]* or call *[study phone]*.

Screener1. Based on the definition and examples provided, which of the following describe your agency's experience developing and implementing case-profiling tools? *Select all that apply*.

- ☐ Our agency currently uses a case-profiling tool(s). (1)
- ☐ Our agency previously used a case-profiling tool but has since discontinued it. (2)
- ☐ Our agency developed but never implemented case-profiling tools. (3)
- ☐ Our agency has never developed or implemented case-profiling tools. (4)

[PROGRAMMER: If SCREENER1_3 OR SCREENER1_4 are CHECKED, NO OTHER OPTION CAN BE SELECTED]

[PROGRAMMER: Ask SECTION A If SCREENER1_1 = checked]

1. Section A. State Agencies With Current Case-Profiling Tools

A1. To the best of your knowledge, how many case-profiling tools does your State agency currently use to identify SNAP cases at risk of payment error?

_____#_____

[PROGRAMMER: IF 1, GO TO A1a; IF 2 OR MORE, GO TO A1b]

A1a. Please provide the name of the tool so that we may refer to it throughout the survey. If the tool does not have an official name, a two- to four-word description may be used (e.g., Essex County's Tool).

_____

A1b. Of the [INSERT # FROM A1] tools your State agency currently uses, we will ask follow-up questions about one of them. Please provide a name for one of the tools, preferably the one with which you are most familiar, so that we may refer to it throughout the survey. If the tool does not have an official name, a two- to four-word description may be used (e.g., Essex County's Tool).

_____

A2a. Please briefly describe what the [INSERT TOOL NAME FROM A1a or A1b] was designed to do.

A2b. Can local agencies/counties customize (or change) the criteria used by the [INSERT TOOL NAME FROM A1a or A1b] to flag cases at risk of a payment error?

- ☐ Yes (1)
- ☐ No (2)
- ☐ Don't know (3)

A3. In what month and year was [INSERT TOOL NAME FROM A1a or A1b] first implemented?

Month: _____

Year: _____

A4. What was the motivation for creating the [INSERT TOOL NAME FROM A1a or A1b]? *Select all that apply*.

- ☐ To concentrate resources (e.g., staff, funds, time) on only those SNAP cases suspected as being at high risk of payment error (1)
- ☐ To create a formal process for identifying SNAP cases at risk of payment error (2)
- ☐ To address high payment error rates (3)
- ☐ To address the findings or recommendations of an audit or management evaluation (4)
- ☐ Other; explain: _____ (5)
- ☐ Don't know (6)

[PROGRAMMER: IF A4_6 = CHECKED, NO OTHER OPTION CAN BE SELECTED]

A5. Who helped design and/or build the [INSERT TOOL NAME FROM A1a or A1b] for SNAP? *Select all that apply*.

- ☐ SNAP State agency program/policy staff (1)
- ☐ State IT staff (2)
- ☐ Local SNAP office staff (3)
- ☐ Vendor/contractor (4)
- ☐ FNS Regional or National Office (5)
- ☐ Other; specify: _____ (6)
- ☐ Don't know (7)

[PROGRAMMER: IF A5_4 = CHECKED, GO TO A6, ELSE GO TO A7]

[PROGRAMMER: IF A5_7 = CHECKED, NO OTHER OPTION CAN BE SELECTED]

A6. When, if at all, was the SNAP State agency involved in developing the [INSERT TOOL NAME FROM A1a or A1b] with the vendor/contractor? *Select all that apply*.

- ☐ Conceptualization or design phase (1)
- ☐ Development phase (2)
- ☐ Testing phase (3)
- ☐ Monitoring and evaluation (4)
- ☐ Other phase; specify: _____ (5)
- ☐ The SNAP State agency was not involved in developing the tool (6)
- ☐ Don't know or don't recall (7)

[PROGRAMMER: IF A6_6 | A6_7 = CHECKED, NO OTHER OPTION CAN BE SELECTED]

A7. Who helped test the [INSERT TOOL NAME FROM A1a or A1b] to make sure it worked as intended? *Select all that apply*.

- ☐ SNAP State agency program/policy staff (1)
- ☐ State IT staff (2)
- ☐ Local SNAP office staff (3)
- ☐ Vendor/contractor (4)
- ☐ Other; specify: _____ (5)
- ☐ Not applicable, the tool was not tested (6)
- ☐ Don't know (7)

[PROGRAMMER: IF A7_6 | A7_7 = CHECKED, NO OTHER OPTIONs CAN BE SELECTED]

A8. When designing the [INSERT TOOL NAME FROM A1a or A1b], what data were analyzed to determine which SNAP cases are at high risk of payment error? *Select all that apply*.

- ☐ SNAP Quality Control data (1)
- ☐ Other SNAP data (2)
- ☐ Proprietary vendor or contractor data (3)
- ☐ Other source, describe: _____ (4)
- ☐ Don't know (5)

[PROGRAMMER: IF A8_5 = CHECKED, NO OTHER OPTION CAN BE SELECTED]

The next five questions ask about the data elements the SNAP State agency **considered** and eventually **selected** for inclusion in the [INSERT TOOL NAME FROM A1a or A1b]. The State agency's consideration of a data element could mean the agency reviewed the policy related to that data element (e.g., policy related to household size), asked staff members whether that data element is associated with payment error, and/or analyzed SNAP QC or other data to assess the relationship between that data element and payment error.

Each question asks about a different category of data elements, which include (1) household composition, (2) demographic characteristics of the household members, (3) economic characteristics, (4) household expenses, and (5) case characteristics.

[PROGRAMMER: Include a hover box definition for "ABAWD" that reads as follows: "ABAWD stands for Able-Bodied Adult Without Dependents"]

A9. For the first category, ***household composition***, which of the following data elements were **considered**, and which were **selected**, for the [INSERT TOOL NAME FROM A1a or A1b]? *Use the checkboxes to the right of each household composition data element to indicate whether the data element was considered and/or selected for your State agency's case-profiling tool.*

Household Composition Data Elements	Considered	Selected	Neither	Don't Know
Total number of household members (1)	☐	☐	☐	☐
Presence/number of children in the household (2)	☐	☐	☐	☐
Presence/number of elderly in the household (3)	☐	☐	☐	☐
Presence of disabled household member (4)	☐	☐	☐	☐
Presence of ABAWD household member (5)	☐	☐	☐	☐
Presence of household member ineligible for SNAP (6)	☐	☐	☐	☐
Child-only unit (7)	☐	☐	☐	☐

Note: ABAWD = able-bodied adults without dependents

[PROGRAMMER: Display A9a on the same web page as A9]

A9a. If any other ***household composition*** data elements were included in the [INSERT TOOL NAME FROM A1a or A1b], please list them below.

_____

_____

A10. The next data category is the ***demographic characteristics*** *of the household members*. Which of the following demographic data elements were **considered**, and which were **selected**, for the [INSERT TOOL NAME FROM A1a or A1b]? *Use the checkboxes to the right of each demographic data element to indicate whether the data element was considered and/or selected for your State agency's case-profiling tool.*

Demographic Data Elements	Considered	Selected	Neither	Don't Know
Age (1)	☐	☐	☐	☐
Race (2)	☐	☐	☐	☐
Ethnicity (3)	☐	☐	☐	☐
Sex (4)	☐	☐	☐	☐
Student status (5)	☐	☐	☐	☐
Level of education (6)	☐	☐	☐	☐
Employment status (7)	☐	☐	☐	☐
Marital status (8)	☐	☐	☐	☐
Homeless (9)	☐	☐	☐	☐
Residency status (10)	☐	☐	☐	☐
Citizenship status (11)	☐	☐	☐	☐

[PROGRAMMER: Display A10a on the same web page as A10]

A10a. If any other ***demographic*** data elements were included in the [INSERT TOOL NAME FROM A1a or A1b], please list them below.

_____

_____

[PROGRAMMER: Include a hover box definition for "Assets/resources" that reads as follows: "Assets/Resources include sums of money owned by the household and usually in an easily accessible financial instrument like savings accounts or stocks. In addition, assets/resources may include the monetary value of household property such as vehicles or land."

A11. The next data category is ***economic characteristics*** of the SNAP household. Which of the following economic data elements were **considered**, and which were **selected**, for the [INSERT TOOL NAME FROM A1a or A1b]? *Use the checkboxes to the right of each economic data element to indicate whether the data element was considered and/or selected for your State agency's case-profiling tool.*

Economic Data Elements	Considered	Selected	Neither	Don't Know
Presence of earned income (1)	☐	☐	☐	☐
Presence of unearned income (2)	☐	☐	☐	☐
Zero income (3)	☐	☐	☐	☐
Gross income (4)	☐	☐	☐	☐
Net income (5)	☐	☐	☐	☐
Self employment income (6)	☐	☐	☐	☐
Assets/resources (7)	☐	☐	☐	☐

[PROGRAMMER: Display A11a on the same web page as A11]

A11a. If any other ***economic*** data elements were included in the [INSERT TOOL NAME FROM A1a or A1b], please list them below.

_____

_____

[PROGRAMMER: Include a hover box definition for "SUAs" that reads as follows: "Standard utility allowance"]

A12. The next data category is ***household expenses***. Which of the following household expense data elements were **considered**, and which were **selected**, for the [INSERT TOOL NAME FROM A1a or A1b]? *Use the checkboxes to the right of each household expense data element to indicate whether the data element was considered and/or selected for your State agency's case-profiling tool.*

Household Expense Data Elements	Considered	Selected	Neither	Don't Know
Medical expenses (1)	☐	☐	☐	☐
Non-utility shelter expenses (2)	☐	☐	☐	☐
Utility expenses (including SUAs) (3)	☐	☐	☐	☐
Dependent care expenses (4)	☐	☐	☐	☐
Homeless shelter deduction (5)	☐	☐	☐	☐
Excess shelter deduction (6)	☐	☐	☐	☐
Legally obligated child support (7)	☐	☐	☐	☐

[PROGRAMMER: Display A12a on the same web page as A12]

A12a. If any other ***household expense*** data elements were included in the [INSERT TOOL NAME FROM A1a or A1b], please list them below.

_____

_____

A13. The next data category is ***case characteristics***. Which of the following case characteristics were **considered**, and which were **selected**, for the [INSERT TOOL NAME FROM A1a or A1b]? *Use the checkboxes to the right of each case characteristic to indicate whether the data element was considered and/or selected for your State agency's case-profiling tool.*

Case Characteristics	Considered	Selected	Neither	Don't Know
Length of certification period (1)	☐	☐	☐	☐
Reporting requirements (e.g., change reporting, simplified reporting) (2)	☐	☐	☐	☐
Benefit amount (3)	☐	☐	☐	☐
New applicant (4)	☐	☐	☐	☐

[PROGRAMMER: Display A13a on the same web page as A13]

A13a. If any other ***case characteristics*** data elements were included in the [INSERT TOOL NAME FROM A1a or A1b], please list them below.

_____

_____

[PROGRAMMER: ASK A14 IF A5_4 OR A7_4 = CHECKED]

A14a. Of the ***household composition*** data elements you indicated were selected for the [INSERT TOOL NAME FROM A1a or A1b] (shown below), were any recommended by a vendor/contractor to flag cases at risk of payment error? *Select "Vendor gave a recommendation on this variable," "Vendor did not comment on this variable," or "Don't recall" for each data element you selected.*

[PROGRAMMER: POPULATE RESPONSE OPTIONS FROM VARIABLES CHECKED "Selected" IN A9]

Data Elements Selected	Vendor gave a recommendation on this variable	Vendor did not comment on this variable	Don't Recall
Total number of household members (1)	1	2	3
Presence/number of children in the household (2)	1	2	3
Presence/number of elderly in the household (3)	1	2	3
Presence of disabled household member (4)	1	2	3
Presence of ABAWD household member (5)	1	2	3
Presence of household member ineligible for SNAP (6)	1	2	3
Child-only unit (7)	1	2	3

A14b. Of the ***demographic*** data elements you indicated were selected for the [INSERT TOOL NAME FROM A1a or A1b] (shown below), were any recommended by a vendor/contractor to flag cases at risk of payment error? *Select "Vendor gave a recommendation on this variable," "Vendor did not comment on this variable," or "Don't recall" for each data element you selected.*

[PROGRAMMER: POPULATE RESPONSE OPTIONS FROM VARIABLES CHECKED "Selected" IN A10]

Data Elements Selected	Vendor gave a recommendation on this variable	Vendor did not comment on this variable	Don't Recall
Age (1)	1	2	3
Race (2)	1	2	3
Ethnicity (3)	1	2	3
Sex (4)	1	2	3
Student status (5)	1	2	3
Level of education (6)	1	2	3
Employment status (7)	1	2	3
Marital status (8)	1	2	3
Homeless (9)	1	2	3
Residency status (10)	1	2	3
Citizenship status (11)	1	2	3

A14c. Of the ***economic*** data elements you indicated were selected for the [INSERT TOOL NAME FROM A1a or A1b] (shown below), were any recommended by a vendor/contractor to flag cases at risk of payment error? *Select "Vendor gave a recommendation on this variable," "Vendor did not comment on this variable," or "Don't recall" for each data element you selected.*

[PROGRAMMER: POPULATE RESPONSE OPTIONS FROM VARIABLES CHECKED "Selected" IN A11]

Data Elements Selected	Vendor gave a recommendation on this variable	Vendor did not comment on this variable	Don't Recall
Presence of earned income (1)	1	2	3
Presence of unearned income (2)	1	2	3
Zero income (3)	1	2	3
Gross income (4)	1	2	3
Net income (5)	1	2	3
Self employment income (6)	1	2	3
Assets/resources (7)	1	2	3

A14d. Of the ***household expense*** data elements you indicated were selected for the [INSERT TOOL NAME FROM A1a or A1b] (shown below), were any recommended by a vendor/contractor to flag cases at risk of payment error? *Select "Vendor gave a recommendation on this variable," "Vendor did not comment on this variable," or "Don't recall" for each data element you selected.*

[PROGRAMMER: POPULATE RESPONSE OPTIONS FROM VARIABLES CHECKED "Selected" IN A12]

Data Elements Selected	Vendor gave a recommendation on this variable	Vendor did not comment on this variable	Don't Recall
Medical expenses (1)	1	2	3
Non-utility shelter expenses (2)	1	2	3
Utility expenses (including SUAs) (3)	1	2	3
Standard utility allowance (4)	1	2	3
Dependent care expenses (5)	1	2	3
Homeless shelter deduction (6)	1	2	3
Excess shelter deduction (7)	1	2	3
Legally obligated child support (8)	1	2	3

A14e. Of the ***case characteristics*** data elements you indicated were selected for the [INSERT TOOL NAME FROM A1a or A1b] (shown below), were any recommended by a vendor/contractor to flag cases at risk of payment error? *Select "Vendor gave a recommendation on this variable," "Vendor did not comment on this variable," or "Don't recall" for each data element you selected.*

[PROGRAMMER: POPULATE RESPONSE OPTIONS FROM VARIABLES CHECKED "Selected" IN A13]

Data Elements Selected	Vendor gave a recommendation on this variable	Vendor did not comment on this variable	Don't Recall
Length of certification period (1)	1	2	3
Reporting requirements (e.g., change reporting, simplified reporting) (2)	1	2	3
Benefit amount (3)	1	2	3
New applicant (4)	1	2	3

[PROGRAMMER: ASK A15 IF A5_4 = CHECKED]

A15. Does the [INSERT TOOL NAME FROM A1a or A1b] include any variables from datasets maintained by a private company (e.g., *LexisNexis LexID)?*

- ☐ Yes (1)
- ☐ No (2)
- ☐ Don't know (3)

[PROGRAMMER: Ask A15a if A15 = Yes]

A15a. Please describe the variables included in [INSERT TOOL NAME FROM A1a or A1b] from a private company.

_____

_____

A16a. Briefly describe how the decision was made to incorporate the data elements shown below in the [INSERT TOOL NAME FROM A1a or A1b] used to flag cases at high risk of payment error.

[PROGRAMMER: POPULATE RESPONSE OPTIONS FROM VARIABLES CHECKED "Selected" IN A9, A10, A11, A12, or A13 ORGANIZED BY DATA ELEMENT CATEGORY]

_____

_____

A16b. Briefly describe how the decision was made to incorporate the other data elements you listed in the [INSERT TOOL NAME FROM A1a or A1b] used to flag cases at high risk of payment error.

[PROGRAMMER: POPULATE RESPONSE OPTIONS FROM VARIABLES CHECKED "Selected" IN A9a, A10a, A11a, A12a, A13a, or A15a ORGANIZED BY DATA ELEMENT CATEGORY]

_____

_____

[PROGRAMMER: Include a hover box definition for "Descriptive statistics" that reads as follows: "Descriptive statistics include analysis of individual variables like calculating the average, maximum, minimum, and standard deviation. Descriptive statistics also include analyses that involve more than one variable like crosstabulations and correlation."]

[PROGRAMMER: Include a hover box definition for "Regression modeling" that reads as follows: "Regression models include linear regression, logistic regression, and others. They can be typically written out in the form of an equation where the dependent variable is a function of one or more dependent variables that have coefficients estimated by the model (e.g., $Y_i = \beta_0 + \beta_1 X_i + e_i$)."]
[PROGRAMMER: Include a hover box definition for "Machine learning" that reads as follows: "Machine learning includes a broad set of analysis methods. For the purposes of case profiling tools, machine learning approaches could include decision trees (sometimes referred to as classification and regression trees, or CART), support vector machines, and neural networks."]

A17. What type of data analysis was used to develop the [INSERT TOOL NAME FROM A1a or A1b]? *Select all that apply*.

- ☐ Descriptive statistics (1)
- ☐ Regression modeling (2)
- ☐ Machine learning (3)
- ☐ Other; describe: _____ (4)
- ☐ No data analysis was done when developing the tool (5)
- ☐ Don't know (6)

[PROGRAMMER: IF A17_5 | A17_6 = CHECKED, NO OTHER OPTION CAN BE SELECTED]

The next few questions ask about how and when the [INSERT TOOL NAME FROM A1a or A1b] is used.

[PROGRAMMER: Use pop up informational window for "Hover here for examples of case profiling tools" to show the following text: "

What do case-profiling tools look like? SNAP case-profiling tools can range in sophistication from a simple checklist that helps eligibility workers identify cases likely to have a payment error to complex analytics tools incorporated into a State agency's eligibility system.

For example, a state agency may determine payment errors were concentrated in large households of six or more people with no reported income. In this instance, the eligibility workers use a checklist to flag all such cases for supervisors to review.

For a second example, a State agency may have used a machine learning model to predict the likelihood a case has a payment error and integrated the results into the eligibility system. This integrated case-profiling tool would automatically assign a risk score to each case as data are

entered. If the score is above a certain threshold, the case is flagged as a high risk of payment error and undergoes a QA review.

A18. Which of the following options best describe the format of the [INSERT TOOL NAME FROM A1a or A1b]? Hover here for examples of case profiling tools. *Select all that apply.*

- ☐ Written instructions (1)
- ☐ Paper checklist (2)
- ☐ Electronic checklist (3)
- ☐ Algorithm programmed into the eligibility system (4)
- ☐ Algorithm programmed into other systems or databases (5)
- ☐ Other format; describe: _____ (6)
- ☐ Don't know (7)

[PROGRAMMER: IF A18_7 = CHECKED, NO OTHER OPTION CAN BE SELECTED]

A19. When is the [INSERT TOOL NAME FROM A1a or A1b] used to identify SNAP cases at risk of payment error? *Select all that apply.*

- ☐ At the point of application submission for new cases but before the interview (1)
- ☐ After the interview for new cases but before eligibility determination (2)
- ☐ After the eligibility determination for new cases but before benefits have been issued (3)
- ☐ After initial benefit issuance and before recertification (active cases) (4)
- ☐ At the point of application submission for recertification but before the interview (5)
- ☐ After the interview for recertification but before eligibility determination (6)
- ☐ After the recertification determination but before benefits have been issued (7)
- ☐ After benefit issuance for recertified cases (active cases) (8)
- ☐ Other time period; please explain: _____ (9)
- ☐ Don't know (10)

[PROGRAMMER: IF A19_10 = CHECKED, NO OTHER OPTION CAN BE SELECTED]

A20. What data in the SNAP case file, if any, does the [INSERT TOOL NAME FROM A1a or A1b] examine to determine whether the household is at risk of payment error? *Select all that apply.*

- ☐ Data from the household application (1)
- ☐ Data from the household interview (2)
- ☐ Data from data matches (3)
- ☐ Other; explain:_____ (4)
- ☐ Don't know (5)

[PROGRAMMER: IF A20_5 = CHECKED, NO OTHER OPTION CAN BE SELECTED]
[PROGRAMMER: IF A20_5 = CHECKED, SKIP TO A22, ELSE CONTINUE TO A21]

A21. Please describe how the [INSERT TOOL NAME FROM A1a or A1b] uses data from the SNAP case file to determine whether the household is at risk of payment error.

_____

_____

A22. Which staff use the [INSERT TOOL NAME FROM A1a or A1b] or the results for any purpose? *Select all that apply. If county-level staff are the same as local-level staff in your State, please select local-level staff.*

- ☐ Local-level staff (1)
- ☐ County-level staff (2)
- ☐ State-level staff (3)

[PROGRAMMER: ASK A23 IF A22_1 = Checked]

A23. Which local-level staff use the [INSERT TOOL NAME FROM A1a or A1b] to flag cases at risk of payment error? *Select all that apply.*

- ☐ Administrative staff (e.g., receptionist) (1)
- ☐ Frontline eligibility workers (2)
- ☐ Eligibility worker supervisors (3)
- ☐ Other local office staff; specify:_____ (4)
- ☐ Not applicable (e.g., the tool is fully automated, so no staff need to take any action to flag cases at risk of payment error) (5)
- ☐ Don't know (6)

[PROGRAMMER: IF Q23_5 | Q23_6 = CHECKED, NO OTHER OPTION CAN BE SELECTED]

[PROGRAMMER: ASK A24 IF A22_1 = CHECKED]

A24. Which local-level staff use the [INSERT TOOL NAME FROM A1a or A1b] to follow up on cases at risk of payment error? *Select all that apply.*

- ☐ Administrative staff (e.g., receptionist) (1)
- ☐ Frontline eligibility workers (2)
- ☐ Eligibility worker supervisors (3)
- ☐ Other local office staff; specify:_____ (4)
- ☐ Not applicable (5)
- ☐ Don't know (6)

[PROGRAMMER: IF A24_5 | A24_6 = CHECKED, NO OTHER OPTION CAN BE SELECTED]

[PROGRAMMER: ASK A25 IF A22_1 = Checked]

A25. How are local-level staff trained on using the [INSERT TOOL NAME FROM A1a or A1b] to flag cases?

- ☐ In-person (1)
- ☐ Live virtual session (2)
- ☐ Online training without a live presenter or facilitator (3)
- ☐ Written tutorial (4)
- ☐ Other: _____ (5)
- ☐ Not applicable; staff are not trained on this (6)

[PROGRAMMER: IF A25_6 = CHECKED, NO OTHER OPTION CAN BE SELECTED]

[PROGRAMMER: ASK A26 IF A22_2 = Checked]

A26. Which county-level staff use the [INSERT TOOL NAME FROM A1a or A1b] to flag cases at risk of payment error? *Select all that apply.*

- ☐ Quality assurance staff (1)
- ☐ Quality control staff (2)
- ☐ Eligibility workers (3)
- ☐ Eligibility supervisors (4)
- ☐ Other county-level staff; specify:_____ (5)
- ☐ Not applicable (e.g., the tool is fully automated, so no staff need to take any action to flag cases at risk of payment error) (6)
- ☐ Don't know (7)

[PROGRAMMER: IF A26_6 | A26_7 = CHECKED, NO OTHER OPTION CAN BE SELECTED]
[PROGRAMMER: ASK A27 IF A22_2 = Checked]

A27. Which county-level staff use the [INSERT TOOL NAME FROM A1a or A1b] to follow up on cases at risk of payment error? *Select all that apply.*

- ☐ Quality assurance staff (1)
- ☐ Quality control staff (2)
- ☐ Eligibility workers (3)
- ☐ Eligibility supervisors (4)
- ☐ Other county-level staff; specify:_____ (5)
- ☐ Not applicable (6)
- ☐ Don't know (7)

[PROGRAMMER: IF A27_6 | A27_7 = CHECKED, NO OTHER OPTION CAN BE SELECTED]

[PROGRAMMER: ASK A28 IF A22_2 = Checked]

A28. How are county-level staff trained on using the [INSERT TOOL NAME FROM A1a or A1b] to flag cases?

- ☐ In-person (1)
- ☐ Live virtual session (2)
- ☐ Online training without a live presenter or facilitator (3)
- ☐ Written tutorial (4)
- ☐ Other: _____ (5)
- ☐ Not applicable; staff are not trained on this (6)

[PROGRAMMER: IF A28_6 = CHECKED, NO OTHER OPTION CAN BE SELECTED]

[PROGRAMMER: ASK A29 IF A22_3 = Checked]

A29. Which State-level staff use the [INSERT TOOL NAME FROM A1a or A1b] to flag cases at risk of payment error? *Select all that apply.*

- ☐ Quality assurance staff (1)
- ☐ Statisticians (2)
- ☐ Quality control staff (3)
- ☐ Other State-level staff; specify:_____ (4)
- ☐ Not applicable (e.g., the tool is fully automated, so no staff need to take any action to flag cases at risk of payment error) (5)
- ☐ Don't know (6)

[PROGRAMMER: IF A29_5 | A29_6 = CHECKED, NO OTHER OPTION CAN BE SELECTED]

[PROGRAMMER: ASK A30 IF A22_3 = Checked]

A30. Which State-level staff use the [INSERT TOOL NAME FROM A1a or A1b] to follow up on cases at risk of payment error? *Select all that apply.*

- ☐ Quality assurance staff (1)
- ☐ Statisticians (2)
- ☐ Quality control staff (3)
- ☐ Other State-level staff; specify:_____ (4)
- ☐ Not applicable (5)
- ☐ Don't know (6)

[PROGRAMMER: IF A30_5 | A30_6 = CHECKED, NO OTHER OPTION CAN BE SELECTED]

[PROGRAMMER: ASK A31 IF A22_3 = Checked]

A31. How are State-level staff trained on using the [INSERT TOOL NAME FROM A1a or A1b] to flag cases?

- ☐ In-person (1)
- ☐ Live virtual session (2)
- ☐ Online training without a live presenter or facilitator (3)
- ☐ Written tutorial (4)
- ☐ Other: _____ (5)
- ☐ Not applicable; staff are not trained on this (6)

[PROGRAMMER: IF A31_6 = CHECKED, NO OTHER OPTION CAN BE SELECTED]
A32. Once the [INSERT TOOL NAME FROM A1a or A1b] flags a SNAP case as at risk of payment error, what is supposed to happen to that case next? *Select all that apply.*

- ☐ It undergoes a second review by an eligibility worker (1)
- ☐ It undergoes a second review by an eligibility worker supervisor (2)
- ☐ It undergoes a quality assurance review (3)
- ☐ Other; specify:_____ (4)
- ☐ No other action is taken (5)
- ☐ Don't know (6)

[PROGRAMMER: IF A32_5 | A32_6 = CHECKED, NO OTHER OPTION CAN BE SELECTED]

A33. Is the [INSERT TOOL NAME FROM A1a or A1b] used for any other purpose besides flagging SNAP cases at risk of payment error?

- ☐ Yes (1)
- ☐ No (2)
- ☐ Don't know (3)

[PROGRAMMER: ASK A33a IF A33 = 1]

A33a. Please describe the other purpose(s) of the [INSERT TOOL NAME FROM A1a or A1b]?

_____

_____

The next few questions ask about evaluating the performance of the [INSERT TOOL NAME FROM A1a or A1b].

A34. In general, have you been able to ascertain whether the [INSERT TOOL NAME FROM A1a or A1b] has an impact on error rates, either good or bad?

- ☐ Yes (1)
- ☐ No (2)
- ☐ Unsure (3)

A35. How often do you evaluate the effectiveness of the [INSERT TOOL NAME FROM A1a or A1b] to flag cases with payment errors?

- ☐ Annually (1)
- ☐ More frequently than annually (2)
- ☐ Less frequently than annually (3)
- ☐ We have never evaluated effectiveness (4)

[PROGRAMMER: ASK A35a IF A35 = 1 | A35 = 2 | A35 = 3]

A35a. Please describe the evaluation process.

_____

_____

A36. Please upload up to five materials that you think may help the study team understand how the [INSERT TOOL NAME FROM A1a or A1b] works to identify SNAP cases at risk of payment error.

If available, please include the following materials:

- ☐ Tool itself (e.g., a PDF of a checklist tool, an .RDS file with a machine-learning model developed in R)
- ☐ Documentation on the tool
- ☐ Procedural manual for using the tool
- ☐ Evaluations of the tool

Click this link to upload the materials.

Important: Do *not* include any SNAP participant data in any of your files.

[PROGRAMMER: Ask SECTION B If SCREENER1_2 = checked]

2. Section B. State Agencies That Previously Used Case-Profiling Tools

B1. You indicated your State agency previously used a case-profiling tool to identify SNAP cases at risk of payment error but discontinued using that tool. Please provide the name of the tool so that we may refer to it throughout the survey.

If the tool does not have an official name, a two- to four-word description may be used (e.g., Essex County's Tool). If your State agency has previously used and discontinued multiple tools, please tell us about the most recent tool.

_____

B2. Who helped design and/or build the [INSERT TOOL NAME FROM B1] for SNAP? *Select all that apply*.

- ☐ SNAP State agency program/policy staff (1)
- ☐ State IT staff (2)
- ☐ Other State-level staff (3)
- ☐ Local SNAP office staff (4)
- ☐ Vendor/contractor (5)
- ☐ FNS Regional or National Office (6)
- ☐ Other; specify: _____ (7)
- ☐ Don't know/don't recall (8)

[PROGRAMMER: IF B2_8 = CHECKED, NO OTHER OPTION CAN BE SELECTED]

B3. Please briefly describe what the [INSERT TOOL NAME FROM B1] was designed to do.

_____

_____

B4. What years was the [INSERT TOOL NAME FROM B1] in use? *Estimates are fine*.

From _____ [year] to _____ [year]

B5. Please briefly describe the reason(s) the [INSERT TOOL NAME FROM B1] is no longer in use.

_____

_____

[PROGRAMMER: Ask SECTION C If SCREENER1_3 = checked]

3. Section C. State Agencies That Developed but Never Implemented Case-Profiling Tools

C1. You indicated your State agency developed a case-profiling tool to identify SNAP cases at risk of payment error but never implemented the tool. Please provide the name of the tool so that we may refer to it throughout the survey.

If the tool does not have an official name, a two- to four-word description may be used (e.g., Essex County's Tool). If your State agency has developed but not implemented multiple tools, please tell us about the most recently developed tool.

_____

C2. Who helped design and/or build the [INSERT TOOL NAME FROM C1] for SNAP? *Select all that apply.*

- ☐ SNAP State agency program/policy staff (1)
- ☐ State IT staff (2)
- ☐ Other State-level staff (3)
- ☐ Local SNAP office staff (4)
- ☐ Vendor/contractor (5)
- ☐ FNS Regional or National Office (6)
- ☐ Other; specify: _____ (7)
- ☐ Don't know (8)

[PROGRAMMER: IF C2_8 = CHECKED, NO OTHER OPTION CAN BE SELECTED]

C3. Please briefly describe what the [INSERT TOOL NAME FROM C1] was designed to do.

_____

_____

C4. Please briefly describe the reason(s) the [INSERT TOOL NAME FROM C1] was never implemented.

_____

_____

[PROGRAMMER: Ask SECTION D If SCREENER1_4 = checked]

4. Section D. State Agencies That Neither Developed nor Implemented Case-Profiling Tools

D1. Are you familiar with case-profiling tools used to identify which SNAP cases are at risk of payment error?

- ☐ Yes (1)
- ☐ No (2)

[PROGRAMMER: IF D1_2 = checked, SKIP TO SECTION E]

D2. Has your State agency ever considered developing a case-profiling tool to identify SNAP cases at risk of payment error?

☐ Yes (1)

☐ No (2)

[PROGRAMMER: ASK D2a IF D2 = 2]

D2a. Please share the reasons your State agency has not considered developing a case-profiling tool for SNAP.

_____

_____

[PROGRAMMER: ASK D2b IF D2 = 2]

D2b. Does your State agency plan to develop a case-profiling tool? Why or why not?

_____

_____

5. Section E. Context

[PROGRAMMER: ASK E1 IF SCREENER1 = 1]

E1. Do you currently have enough funding to properly administer the [INSERT TOOL NAME FROM A1a or A1b]?

☐ Yes (1)

☐ No (2)

☐ Unsure (3)

E2. Has the COVID-19 pandemic affected your State agency's ability to administer SNAP QC?

☐ Yes (1)

☐ No (2)

☐ Unsure (3)

[PROGRAMMER: ASK E2a IF E2 = 1]

E2a. Please describe how the COVID-19 pandemic affected your State agency's ability to administer SNAP QC.

_____

_____

[PROGRAMMER: ASK E3 IF SCREENER1 = 1]

E3. If the study team has follow-up questions, who is the best person to contact?

Name: (1) _____

Title: (2) _____

Email: (3) _____

Phone: (4) _____

You have reached the end of the survey! Thank you for taking the time to respond.

As you know, the final phase of the study involves case studies with six State agencies, which will be selected in collaboration with FNS. We will email you in the coming months if your State agency is selected.

Interview Guide: State-Level Data Analysis Staff

OMB Number: 0584-0696
Expiration Date: 05/31/2027

A. Introduction

Good morning/afternoon. Thank you for taking the time to talk with me today. My name is *[interviewer's name],* and I work for Westat, a private research company based in Rockville, Maryland. Joining me is my colleague, *[name]*.

Purpose: The U.S. Department of Agriculture's Food and Nutrition Service, or FNS, is interested in understanding tools used to identify cases likely to have a payment error. These tools may be known by different names, such as case-profiling tools, risk assessment tools, or error-prone profiling. After cases are flagged, they undergo a rigorous process to ensure accurate benefit decisions. FNS hired Westat to conduct a study to learn more about the development and implementation of these tools. The findings from the study will be used to inform the development of case-profiling tools FNS and State agencies use, identify best practices, and develop resources and technical assistance.

How you were selected: We first conducted an online survey of all SNAP State directors. We then worked with FNS to select six State agencies among survey respondents for more in-depth case studies on their use of case-profiling tools. After we discussed with the SNAP State Director what we hope to learn from your State agency, the State Director identified you as someone who would have valuable input and should be interviewed.

Risks and privacy: We use all data we collect only for the research purposes we describe. FNS knows which State agencies were asked to participate in each case study but does not know the names or job titles of the individuals interviewed. We will report the results of these interviews for each State agency, but your name will not be linked to your responses. In our reports, we may include direct quotes, but they will be presented without the speaker's name or title. FNS will receive a redacted copy of the transcript of this interview that has been stripped of identifying information, except for the name of your State agency.

Study costs and compensation: There is no cost to you to participate apart from the time you spend with us for this interview, and there is no compensation. The interview takes 60 minutes.

Voluntary participation: Your participation is entirely voluntary. Refusal to participate will not have any impact on your position, your State agency, or nutrition programs. You may take a break, skip questions, say something off the record, or stop participating at any time.

Questions: If you have questions about your rights and welfare as a research participant, please call the Westat Human Subjects Protections office at 1.888.920.7631. Please leave a message with your first name; the name of the research study you are calling about, which is the SNAP Risk Assessment study; and a phone number, beginning with the area code. Someone will return your call as soon as possible.

We have planned for this discussion to last 60 minutes, until *[time]*. Is that still okay?

With your permission, I would like to record this discussion to help us fill any gaps in our written notes. The recordings, transcripts, and any notes we have will be stored on our secure server and will be destroyed after the project is complete. FNS will not receive any audio recordings.

Do you have any questions? *[Answer all questions]*

May I turn on the audiorecorder now? *[Turn on audiorecorder if gives consent]*

Now that the audiorecorder is on, do you agree to participate? *[Pause for response]*

And do you consent to be audiorecorded? *[Pause for response]*

B. Warmup and Context

1. To start, please tell me how long you have worked at the agency and what your responsibilities are.
2. For the rest of this discussion, we will be talking mostly about the *[tool name]*, which your State agency provided information about in the online survey. What was the nature of your involvement with *[tool name]*?

   [Probe: Designed it, tested it? Is familiar with tool but was not involved in development?]
3. In the survey, the State agency said *[tool name]* is a *[read survey response A18]* that flags SNAP cases at risk of payment error using data from *[read survey response A20]*. Does that description still seem accurate?

   [If no] How would you revise the description?

C. Data Analysis [if relevant]

1. To figure out which case characteristics to have the tool focus on, how did you know where the majority of payment errors were coming from?

   a. Tell me about any data you analyzed to determine where the errors were coming from.

   [Probe: SNAP Quality Control data, vendor/contractor data, local agency data, other?]

   [If they analyzed data] What sort of analysis did you do on those data?

[Probe: Descriptive statistics, modeling, machine learning?]

a. How did you decide on your analytical approach?

b. What outcome or outcomes did you examine?

[Probe: Presence of a payment error? Payment error as a continuous measure?]

c. How, if at all, did you test the accuracy of the analysis?

d. What challenges arose when using those data to inform how the tool would be designed?

   i. How did you overcome those challenges?

D. Factors and Variables

Thank you. Those details gave me a really helpful background. Now I'd like to ask about the specific information the *[tool name]* looks for.

1. My understanding from the survey is that, when the *[tool name]* tries to identify which cases are at high risk of payment error, it looks at *[read survey responses A9–A13, A15a]*. Did I capture that information correctly?

   a. *[If no]* What information does the *[tool name]* look at to flag SNAP cases at high risk of payment error?

   b. Is there any documentation you could share on each factor in terms of whether they are categorical, continuous, or measured in some other way?

   [Note: Follow up on this at the end, and request any documentation]

2. Can you recall how the decision was made to focus on those factors?

   a. Who made the decision?

   b. What, if anything, would you change about the factors the *[tool name]* focuses on?

   c. Can you recall how the decision was made to measure each factor categorically, continuously, or some other way?

      i. Who made the decisions?

      ii. What, if anything, would you change about how the factors are measured?

3. Were any factors considered for the tool that you didn't end up using?

[If yes]

   a. Which ones?

   b. Why was it decided not to use them?

4. Have the factors the *[tool name]* focuses on changed over time? If yes, how?

[Probe to understand whether changes were to focus on different factors altogether or on how those factors were measured—continuous, categorical, etc.]

   a. Why were the changes made?

[Probe to understand whether staff learned they needed to make adjustments as a result of monitoring or testing of the tool]

E. Testing the Tool [if relevant]

1. *[If survey response A7 ≠ NA]* Tell me about how the *[tool name]* was tested before going live.

   a. What was being tested?

   [Probe: Were they looking at the overall accuracy of the tool? Equity of the tool across subgroups? User-friendliness?]

   b. Who did the testing?

   c. Did those early tests reveal anything that needed to be fixed?

   [If yes]

      i. Tell me a little about what needed to be fixed.

      ii. How long did that take to resolve?

2. Was the tool also tested after going live? If yes, how?

[Probe: What were they looking for with those tests—accuracy, equity, user-friendliness?]

   a. Did anything need to be fixed when testing the tool after it went live?

   [*If yes*]

      i. Tell me a little about what needed to be fixed.

      ii. How long did that take to resolve?

F. Tracking Implementation

1. What data, if any, are tracked on the outcomes of the cases the *[tool name]* flags?

[Note: We are asking whether they track any follow-up steps taken for these cases, such as efforts to find additional documentation on the household or calls to the household to ask follow-up questions]

[If track data]

   a. Who tracks those data?

2. *[If tool is used **after** benefit determination]* Do the data indicate whether the flagged cases were actually found to have payment errors?

[If do not track data]

3. What data would you like to track, if any?
4. What, if anything, makes it difficult to collect data on the *[tool name]*?
5. What, if anything, do you report on the cases *[tool name]* flags?
   a. What information is shared in these reports?
6. Who receives these reports?

   [Probe: Local office staff, State-level staff, FNS staff?]
7. What do the recipients do with this information?
8. Would you be able to share the latest of these reports with us?
   [Note: Ask for this report again at the end of the interview]
9. Aside from reporting, what other steps are you required to take based on the results of those reports?

   [Probe to understand how these steps are carried out, when, and by whom.]
10. We know this can be hard to determine, but have you been able to ascertain whether the tool has had an impact on payment error rates, either good or bad?

    [Probe to understand whether the impact is their perception or based in data]

G. Wrap-Up

This discussion has been very helpful.

1. If you were to talk to other State agencies considering implementing a similar tool, what advice would you give them?
2. Before we wrap up, are there any key challenges to building or implementing case-profiling tools like *[tool name]* that we haven't already discussed?

We've reached the end of the interview. Thank you so much for taking the time to talk with us and share your experiences. The information you provided gave us valuable insights into how tools like *[tool name]* work.

[If applicable] I recall you mentioning that you would be willing to share *[documents]* with me. Those would be really helpful to see, so thank you for offering to send them. You can send them to me at *[email address]*. I can also set up a secure FTP site to receive the materials if the documents contain identifying information.

Interview Guide: Risk Assessment Tool Development Lead, Quality Assurance Director

OMB Number: 0584-0696
Expiration Date: 05/31/2027

A. Introduction

Good morning/afternoon. Thank you for taking the time to talk with me today. My name is *[interviewer's name]*, and I work for Westat, a private research company based in Rockville, Maryland. Joining me is my colleague, *[name]*.

Purpose: The U.S. Department of Agriculture's Food and Nutrition Service, or FNS, is interested in understanding tools SNAP State agencies use to identify cases likely to have a payment error. These tools may be known by different names, such as case-profiling tools, risk assessment tools, or error-prone profiling. After cases are flagged as high risk, they undergo a more rigorous process to ensure accurate benefit decisions. FNS hired Westat to conduct a study to learn more about the development and implementation of these tools. The findings from the study will be used to inform the development of case-profiling tools FNS and State agencies use, identify best practices, and develop resources and technical assistance.

How you were selected: We conducted an online survey of all SNAP State agencies and then worked with FNS to select six State agencies for more in-depth case studies on their use of case-profiling tools. Your SNAP State Director identified you as someone who would have valuable input about your State agency's case-profiling tool.

Risks and privacy: We use all data we collect only for the research purposes we describe. FNS knows which State agencies were asked to participate in each case study but does not know the names or job titles of the individuals interviewed. We will report the results of these interviews for each State agency, but your name will not be linked to your responses. In our reports, we may include direct quotes, but they will be presented without the speaker's name or title. FNS will receive a redacted copy of the transcript of this interview that has been stripped of identifying information, except for the name of your State agency.

Study costs and compensation: There is no cost to you to participate apart from the time you spend with us for this interview, and there is no compensation. The interview takes 90 minutes.

Voluntary participation: Your participation is entirely voluntary. Refusal to participate will not have any impact on your position, your State agency, or nutrition programs. You may take a break, skip questions, say something off the record, or stop participating at any time.

Questions: If you have questions about your rights and welfare as a research participant, please call the Westat Human Subjects Protections office at 1.888.920.7631. Please leave a message with your first name; the name of the research study you are calling about, which is the SNAP Risk Assessment study; and a phone number beginning with the area code. Someone will return your call as soon as possible.

We have planned for this discussion to last 90 minutes, until *[time]*. Is that still okay?

With your permission, I would like to record this discussion to help us fill any gaps in our written notes. The recordings, transcripts, and any notes we have will be stored on our secure server and will be destroyed after the project is complete. FNS will not receive any audio recordings.

Do you have any questions? *[Answer all questions]*

May I turn on the audiorecorder now? *[Turn on audiorecorder if gives consent]*

Now that the audiorecorder is on, do you agree to participate? *[Pause for response]*

And do you consent to be audiorecorded? *[Pause for response]*

B. Warmup and Context

1. To start, please tell me how long you have worked at your *[agency/organization/office]* and what your responsibilities are.
2. For the rest of this discussion, we will be talking mostly about the *[tool name]* that your State agency provided information about in the online survey. What was the nature of your involvement with *[tool name]*?

   [Probe: Designed it, built it, tested it, promoted it, other?]
3. In the survey, the State agency said the *[tool name]* is a *[read survey response A18]* that flags SNAP cases at risk of payment error using data from *[read survey response A20]*. Does that description still seem accurate?
   a. *[If no]* How would you revise the description?
4. Has the *[tool name]* been modified in any way since it was first implemented?

[If yes]

   a. Please explain how it evolved.

   [Probe: When and why it evolved; who initiated those changes?]

   b. What were the reasons for those changes?

[Probe: Prompted by staff feedback, review of data, civil rights complaint, other?]

C. Developing the Tool

I want to understand why and how the *[tool name]* was designed and built.

1. What motivated your State agency to develop the *[tool name]*?
2. The survey indicates that the following types of staff were involved in designing the tool: *[read survey response A5]*. How was it decided who would design it?
    a. *[If A5 = multiple responses]* How did the collaboration go between those different staff?
    b. What suggestions do you have for similar teams trying to create a tool like this?
    c. What challenges arose during the design phase?
3. *[If A5 = vendor/contractor]* How much input did the State agency have in how the tool was developed?
    a. Did the vendor offer a premade case-profiling tool that they already had available or did they have to create your State's tool from scratch?
        i. What were the pros and cons of that?
    b. How much input did the State agency have on the final algorithm for the tool?
4. To figure out which case characteristics to have the tool focus on, how did you identify the sources of payment errors?

    [Probe: analyzed SNAP Quality Control data or vendor/contractor data, other?]
5. *[If they analyzed data to inform those decisions]* What sort of analysis did you do on those data?

    [Probe: Descriptive statistics, modeling, machine learning?]
    a. Where were the data pulled from?

        [Probe: Centralized SNAP State database, local agency databases, other?]
    b. What challenges arose when using those data to inform how the tool would be designed?

Now I'd like to ask about the specific information the *[tool name]* looks for.

6. My understanding from the survey is that, when the *[tool name]* tries to identify which cases are at high risk of payment error, it looks at *[read survey response A9–A13, A15a]*. Did I capture that information correctly?
    a. *[If no]* What information does the *[tool name]* look at to flag SNAP cases at high risk of payment error?
    b. Is there any documentation you could share on how each variable is operationalized in terms of whether it's categorical, continuous, or measured in some other way?

        [Note: Make a note to follow up on this question at the end, and request that documentation]

7. Can you recall how the decision was made to focus on those variables?
    a. Who made the decision?
    b. What, if anything, would you change about the variables the *[tool name]* focuses on?
    c. Can you recall how the decision was made to operationalize each variable, in terms of whether they're categorical, continuous, or measured some other way?
        i. Who made the decisions?
        ii. What, if anything, would you change about how the variables are operationalized?
8. Were any variables considered for the tool that you didn't end up using?

[If yes]

    a. Which variables?
    b. Why did you decide not to use them?

9. Have the variables the *[tool name]* focuses on changed over time? If yes, how?

    [Probe to understand whether changes were to focus on different variables altogether or how those variables were measured—continuous, categorical, etc.]

    a. Why were the changes made?

        [Probe to understand whether staff learned they needed to make adjustments as a result of monitoring or testing of the tool]

10. *[If survey A7 ≠ NA]* Tell me about how the *[tool name]* was tested before going live.

    [Probe: Were they looking at the overall accuracy of the tool? Equity of the tool across subgroups? User-friendliness?]

    a. Who did the testing?
    b. Did those early tests reveal anything that needed to be fixed?
    c. Was the tool also tested after going live? If yes, how?

        [Probe: What were they looking for with those tests—accuracy, equity, user friendliness?]

    d. Did anything need to be fixed when testing the tool after it went live?

D. Implementing the Tool

My next set of questions will help me better understand how the *[tool name]* was actually implemented.

1. How did you develop the procedures for using the tool?

   [Note: If the tool was a checklist, these procedures may relate to using the checklist to flag a case and conduct any follow-up steps. If the tool was an algorithm, these procedures may have been a written explanation of how and when the tool flags cases and any follow-up steps for staff on the flagged cases.]

   a. Who was responsible for developing those procedures?

2. Was any training conducted to help staff understand the *[tool name]* and how to use it?

   [If yes]

   a. Who led the training?

   b. Who attended the training?

      [If mention local office staff, clarify whether it was at the supervisor/manager-level or the frontline worker level.]

   c. What did the training cover?

3. My understanding from the survey is that the *[tool name]* flags SNAP cases thought to be at risk of payment error at *[read survey response A19]*. Is that correct?

   a. *[If no]* When does *[tool name]* flag a SNAP case thought to be at risk of payment error?

      [Probe: During certification process, after certification but before benefits are issued, after certification and before recertification, during recertification, other?]

4. If you could, would you adjust the *[tool name]* to flag cases at a different point in the process?

   [If yes]

   a. When would you prefer a case be flagged?

   b. Why would that change be helpful?

5. After the tool flags a case as being at risk of payment error, what is **supposed to** happen next?

   a. What is the timeframe in which those steps have to occur?

   b. What kinds of staff are involved?

      [Probe: Local office staff, State-level staff?]

c. What percent of the time would you estimate that the staff are able to complete those steps exactly as they are spelled out?

   i. [*if not 100% of the time*] What makes it difficult for staff to complete those follow-up steps on cases that are flagged?

6. [*If vendor created tool*] If an aspect of the tool needs to be updated, is the vendor responsible for doing that or is it someone else?

   a. How quickly are those updates typically made?

   b. What challenges arise when making those updates?

   c. What helps that process go smoothly?

7. What data, if any, are tracked on what happens to the cases the *[tool name]* flags?

   [Note: We are asking whether they track any follow-up steps taken for these cases, such as efforts to find additional documentation on the household or calls to the household to ask follow-up questions]

   a. Who tracks those data?

   b. *[If tool is used after benefit issuance, per survey question A19]* Do the data indicate whether the flagged cases were actually found to have payment errors?

8. What, if anything, do you report on the cases the *[tool name]* flags?

   a. What information is shared in these reports?

   b. Who receives these reports?

      [Probe: Local office staff, State-level staff, FNS staff?]

   c. What do the recipients do with this information?

   d. Would you be able to share the latest of these reports with us?

      [Note: Ask for this report again at the end of the interview]

9. Apart from what you just mentioned, do you know of any other ways the State agency uses the data on the cases flagged as being at risk of payment error?

   [Probe to understand if the data are used to identify local offices that need additional training.]

10. What do you believe are the biggest challenges to implementing the *[tool name]*?

    [Probe: IT issues, staffing challenges, training, other?]

11. We know this can be hard to determine, but have you been able to ascertain whether the tool has had an impact on payment error rates, either good or bad?

    [Probe to understand whether the impact is their perception or based in data.]

    a. *[If tool had positive impact on error rates]* Thinking back on the work involved in developing, testing, and implementing the tool, to what extent might those costs be balanced out by improvements to payment accuracy?

12. While the tool *[is/was]* in use, what other strategies, if any, has the State agency used to try to reduce payment error rates?
13. If you were to talk to another SNAP State agency considering implementing a similar tool, what advice would you give them before they roll it out?

[if applicable] Discontinuing the Tool

Now I'd like to ask a few questions about discontinuing the *[tool name]*.

1. In what year was the tool discontinued?
2. Tell me how the decision was made to stop using the *[tool name]*.
   a. Who made the decision?
   b. Were any data considered when making the decision? If yes, explain.
3. Have you been able to ascertain whether discontinuing the tool had an impact on payment error rates, either good or bad?

   [Probe to understand whether the impact is their perception or based in data.]

E. State/Local Context and Wrap-Up

This information has been very helpful. These last few questions will give me a little more understanding of the State and local context before we wrap up.

[Ask questions 32–35 if respondent is a State-level staff person]

1. Do you feel that your local offices have enough qualified staff to review cases and make eligibility determinations? Why or why not?
   a. *[If not enough staff]* How long have local offices been short staffed?

      [Probe: Since before COVID or more recently?]
2. Do you feel your local offices have enough funding to properly review cases and make eligibility determinations? Why or why not?
3. We know that COVID-19 drastically affected what were formerly standard practices when moving a SNAP case through the eligibility determination process. For instance, SNAP offices largely stopped conducting certification interviews. How else did COVID-19 change SNAP processes in ways that may have affected the State agency's payment error rate?

   [Note: Be clear on whether the impact on error rates was positive or negative]

4. *[Ask whether tool was implemented between 2017 and 2019, per survey questions A3 and B4]* We see from the official payment error rates FNS released that your State agency's payment error rate ranged from *[X]* to *[Y]* between 2017 and 2019. Can you think of anything happening at your State agency during that time that may have affected the payment error rate?

   [Probe: New management information system, policy change, corrective action plans, significant staffing changes, other?]

   [Note: Be clear on whether the impact on error rates was positive or negative]

5. Are there any key challenges to successfully building or implementing tools like *[tool name]* that we haven't already discussed?

6. Can you think of anything that makes these tools more accurate in identifying SNAP cases at risk of payment error?

We've reached the end of the interview. Thank you so much for taking the time to talk with us and share your experiences. The information you provided gave us valuable insights into how tools like *[tool name]* work.

[If applicable] I recall you mentioning that you would be willing to share *[documents]* with me. Those would be helpful to see, so thank you for offering to send them. You can send them to me at *[email address]*. I can also set up a secure FTP site to receive the materials if the documents contain identifying information.

Interview Guide: State-Level IT Systems Staff

OMB Number: 0584-0696
Expiration Date: 05/31/2027

A. Introduction

Good morning/afternoon. Thank you for taking the time to talk with me today. My name is *[interviewer's name],* and I work for Westat, a private research company based in Rockville, Maryland. Joining me is my colleague, *[name]*.

Purpose: The U.S. Department of Agriculture's Food and Nutrition Service, or FNS, is interested in understanding tools used to identify cases likely to have a payment error. These tools may be known by different names, such as case-profiling tools, risk assessment tools, or error-prone profiling. After cases are flagged, they undergo a rigorous process to ensure accurate benefit decisions. FNS hired Westat to conduct a study to learn more about the development and implementation of these tools. The findings from the study will be used to inform the development of case-profiling tools FNS and State agencies use, identify best practices, and develop resources and technical assistance.

How you were selected: We first conducted an online survey of all SNAP State directors. We then worked with FNS to select six State agencies among survey respondents for more in-depth case studies on their use of case-profiling tools. After we discussed with the SNAP State Director what we hope to learn from your State agency, the State Director identified you as someone who would have valuable input and should be interviewed.

Risks and privacy: We use all data we collect only for the research purposes we describe. FNS knows which State agencies were asked to participate in each case study but does not know the names or job titles of the individuals interviewed. We will report the results of these interviews for each State agency, but your name will not be linked to your responses. In our reports, we may include direct quotes, but they will be presented without the speaker's name or title. FNS will receive a redacted copy of the transcript of this interview that has been stripped of identifying information, except for the name of your State agency.

Study costs and compensation: There is no cost to you to participate apart from the time you spend with us for this interview, and there is no compensation. The interview takes 45 minutes.

Voluntary participation: Your participation is entirely voluntary. Refusal to participate will not have any impact on your position, your State agency, or nutrition programs. You may take a break, skip questions, say something off the record, or stop participating at any time.

Questions: If you have questions about your rights and welfare as a research participant, please call the Westat Human Subjects Protections office at 1.888.920.7631. Please leave a message with your first name; the name of the research study you are calling about, which is the SNAP Risk Assessment study; and a phone number, beginning with the area code. Someone will return your call as soon as possible.

We have planned for this discussion to last 45 minutes, until *[time]*. Is that still okay?

With your permission, I would like to record this discussion to help us fill any gaps in our written notes. The recordings, transcripts, and any notes we have will be stored on our secure server and will be destroyed after the project is complete. FNS will not receive any audio recordings.

Do you have any questions? *[Answer all questions]*

May I turn on the audiorecorder now? *[Turn on audiorecorder if gives consent]*

Now that the audiorecorder is on, do you agree to participate? *[Pause for response]*

And do you consent to be audiorecorded? *[Pause for response]*

B. Warmup and Context

1. To start, please tell me how long you have worked at the agency and what your responsibilities are.
2. For the rest of this discussion, we will be talking mostly about the *[tool name]*, which your State agency provided information about in the online survey. What was the nature of your involvement with *[tool name]*?

[Probe: Designed it, built it, tested it? Is familiar with tool but was not involved in development?]

3. In the survey, the State agency said *[tool name]* is a *[read survey response A18]* that flags SNAP cases at risk of payment error using data from *[read survey response A20]*. Does that description still seem accurate?
4. *[If no]* How would you revise the description?
5. Has *[tool name]* been modified in any way since it was first implemented?

[If yes]

   a. Please explain how it evolved.

      [Probe: When and why it evolved; who initiated those changes?]

   b. What were the reasons for those changes?

      [Probe: Prompted by staff feedback, policy change, review of data, civil rights complaint, other?]

6. *[For electronic tools/algorithms]* To what extent does the *[tool name]* interface with State agency data systems?

[Probe: State eligibility system, statewide online application system, other?]

   a. What data are exchanged through the interface?
   b. What had to happen to make that interface successful?
   c. What suggestions do you have for similar teams trying to integrate a tool like this with existing data systems?

      [Probe: Challenges encountered and solutions?]

C. Developing the Tool [if relevant]

I want to understand how *[tool name]* was developed and what it was designed to do.

1. The survey indicates the following types of staff were involved in designing the tool: *[read survey response A5]*. How was it decided who would design the tool?
   a. *[If A5 = vendor/contractor]* How much input did the State agency have in how the tool was developed?
      i. How much input did the State agency have on the final algorithm for the tool?
2. *[If A5 = multiple responses]* How did the collaboration go between those different staff?
   a. What suggestions do you have for similar teams trying to create a tool like this?

D. Testing and Reporting [if relevant]

1. *[If survey response A7 ≠ NA]* Tell me about how the *[tool name]* was tested before going live.
   a. What was being tested?

      [Probe: Were they looking at the overall accuracy of the tool? Equity of the tool across subgroups? User-friendliness?]
   b. Who did the testing?
   c. Did those early tests reveal anything that needed to be fixed?

      [*If yes*]
      i. Tell me a little about what needed to be fixed.
      ii. How long did that take to resolve?

2. Was the tool also tested after going live? If yes, how?

[Probe: What were they looking for with those tests—accuracy, equity, user-friendliness?]

   a. Did anything need to be fixed when testing the tool after it went live?

      [*If yes*]

      i. Tell me a little about what needed to be fixed.

      ii. How long did that take to resolve?

3. Thinking broadly about the testing phase, what challenges arose during that time?

   a. How were those challenges overcome?

4. Were any reporting tools programmed into your systems to provide information on [tool name]?

[Note: We are asking whether they track any information on the number or types of cases flagged, the follow-up steps taken for these cases, such as efforts to find additional documentation on the case or calls to the household to ask questions]

[*If yes*]

   a. What are those reports programmed to track with regard to *[tool name]*?

   b. Who can access those reports?

   c. Does the system automatically generate and disseminate reports or is it more that those with access can run the reports anytime they wish?

E. Wrap-Up

1. If you were to talk to other State agencies considering implementing a similar tool, what advice would you give them?

2. Before we wrap up, are there any key challenges to building case-profiling tools like *[tool name]* that we haven't already discussed?

We've reached the end of the interview. Thank you so much for taking the time to talk with us and share your experiences. The information you provided gave us valuable insights into how tools like *[tool name]* work.

[If applicable] I recall you mentioning that you would be willing to share *[documents]* with me. Those would be really helpful to see, so thank you for offering to send them. You can send them to me at *[email address]*. I can also set up a secure FTP site to receive the materials if the documents contain identifying information.

Interview Guide: Local Office Staff

OMB Number: 0584-0696
Expiration Date: 05/31/2027

A. Introduction

Good morning/afternoon. Thank you for taking the time to talk with me today. My name is *[interviewer's name]*, and I work for Westat, a private research company based in Rockville, Maryland.

Purpose: The U.S. Department of Agriculture's Food and Nutrition Service, or FNS, is interested in understanding more about tools used to identify cases likely to have a payment error. These tools may be known by different names, such as case-profiling tools, risk assessment tools, or error-prone profiling. After cases are flagged as high risk, they undergo a rigorous process to ensure accurate benefit decisions. FNS hired Westat to conduct a study to learn more about the development and implementation of these tools. The findings from the study will be used to inform the development of case-profiling tools FNS and State agencies use, identify best practices, and develop resources and technical assistance.

How you were selected: We first conducted an online survey of all SNAP State Directors. We then worked with FNS to select six State agencies among survey respondents for more in-depth case studies on their use of case-profiling tools. After we discussed with the SNAP State Director what we hope to learn from your State agency, the State Director identified you as someone who would have valuable input and should be interviewed.

Risks and privacy: We use all data we collect only for the research purposes we describe. FNS knows which States were asked to participate in each case study but does not know the names or job titles of the individuals interviewed. We will report the results of these interviews for each State as a whole, but your name will not be linked to your responses. In our reports, we may include direct quotes, but they will be presented without the speaker's name or title. FNS will receive a redacted copy of the transcript of this interview that has been stripped of identifying information, except for the name of your State.

Study costs and compensation: There is no cost to you to participate apart from the time you spend with us for this interview, and there is no compensation. The interview takes 60 minutes.

Voluntary participation: Your participation is entirely voluntary. Refusal to participate will not have any impact on your position, your State agency, or nutrition programs. You may take a break, skip questions, say something off the record, or stop participating at any time.

Questions: If you have questions about your rights and welfare as a research participant, please call the Westat Human Subjects Protections office at 1.888.920.7631. Please leave a message

with your first name; the name of the research study you are calling about, which is the SNAP Risk Assessment study; and a phone number, beginning with the area code. Someone will return your call as soon as possible.

We have planned for this discussion to last 60 minutes, until *[time]*. Is that still okay?

With your permission, I would like to record this discussion to help us fill any gaps in our written notes. The recordings, transcripts, and any notes we have will be stored on our secure server and will be destroyed after the project is complete. FNS will not receive any audio recordings.

Do you have any questions? *[Answer all questions]*

May I turn on the audiorecorder now? *[Turn on audiorecorder if gives consent]*

Now that the audiorecorder is on, do you agree to participate? *[Pause for response]*

And do you consent to be audiorecorded? *[Pause for response]*

B. Warmup and Background

1. To start, please tell me how long you have worked at this particular office and what your responsibilities are.
2. For the rest of this discussion, we will be talking mostly about the *[tool name]*. What *[is/was]* the nature of your involvement with *[tool name]*?

   [Probe: Designed it, built it, tested it? Is familiar with tool but was not involved in development?]
3. *[If not involved in designing/building it]* When did you first hear about the *[tool name]*?
   a. What were you told about what it was designed to do?
   b. What were your initial thoughts about it?

C. Using an RA Tool

1. Before I ask more specific questions, please describe how *[tool name]* works.
2. My understanding from the survey is that the *[tool name]* flags SNAP cases thought to be at risk of payment error at *[read survey response A19]*. Is that correct?
   a. *[If no]* When does *[tool name]* flag a SNAP case thought to be at risk of payment error?
3. Which staff use the *[tool name]?*
4. *[For e-tools]* How do staff know that *[tool name]* has flagged a SNAP case as being at risk of payment error?

   [Probe: Is there a note in the case file? Is an automatic notification sent to you? Is there color coding to indicate whether a SNAP case is at risk?]

5. When *[tool name]* flags a SNAP case, what is supposed to happen next?

   [Probe: Undergoes a second review? Household asked to provide additional information?]

   [Note: For the following subquestions, multiple teams of local staff—eligibility staff, interviewers, QC staff, supervisors—may touch a case. Make sure you are clear on which types of staff the respondent is talking about.]

   a. Are staff required to complete these follow-up steps, or are they considered suggested practices?

   b. What percentage of the time would you estimate that the local staff are able to complete those steps exactly as they are spelled out?

      i. *[If not 100 percent of the time]* What makes it difficult for staff to complete those follow-up steps on cases that are flagged?

   c. *[If next steps involve collecting additional information on the case flagged]* How often has collecting that additional information on a SNAP case revealed new information that changes a household's eligibility or benefits?

6. What is the timeframe for completing these steps after a case has been flagged?

   a. Is that enough time to complete the follow-up steps needed? Why or why not?

7. How are these steps different from what would be done for a regular case that was not flagged?

   [Probe: Are the cases flagged reviewed more times than those not flagged? Do staff collect additional or different data for the cases thought to be at risk of payment error than they do for cases not flagged?]

   a. Are different staff involved with the cases flagged than with cases not flagged?

      i. *[If yes]* Which staff are only involved with cases that are flagged by the tool?

8. *[For e-tools]* Do staff enter any information manually when using the *[tool name]*?

   [Note: For instance, staff may enter information gleaned from a head-of-household interview that the tool then uses to determine whether the case is high risk]

   a. Does the *[tool name]* use information from any of your local office systems or databases? If so, please explain.

      [Probe: Online applications, eligibility system, other?]

9. For the households flagged by *[tool name]*, what are they told about this process, if anything?

   a. What are they told about why their case was flagged as being at risk for payment error?

   b. For cases whose SNAP benefits changed because they were flagged by *[tool name]*, what are households told about the reasons for the change?

   c. How could a household correct any inaccurate information that was used to flag their case?

10. Tell me about any training you or other staff received on *[tool name]*.

    [Note: Supervisors/managers may have received training or technical assistance from the SNAP State agency; other local office staff would have probably received training from the supervisor]

    a. Which staff received the training?

    b. Who led the training?

    c. What did the training cover?

    d. Should the training have covered any other topics? If yes, what?

    e. *[If supervisor received training or technical assistance from State]* To what extent did the training prepare you to explain the *[tool name]* to your staff?

    f. *[If no training provided]* What kind of training would have been useful?

        i. Which staff should be trained?

11. What kind of impact, if any, has *[tool name]* had on your daily work?

    [Probe to understand impact on workload, on workflow between staff, other]

12. What do you believe are the biggest challenges to using *[tool name]*?

    [Probe: IT issues, staffing challenges, training, other?]

13. If you could, what would you change about how *[tool name]* works?

    a. Why would that change be beneficial?

14. If you were to hear that another local SNAP office was considering using a similar tool, what advice would you give the office?

D. Data Tracking [Local Office Supervisors and Managers Only]

1. What data, if any, do you receive from your SNAP State agency on payment error rates?

    a. What do you do with that information?

2. Tell me about any data your office tracks with regard to *[tool name]*.

    [Probe: Number of cases flagged, number of cases followed up on, number of cases with payment error, other?]

    a. How are those data tracked?

        [Probe: Shared database? Staff track these data on their own in various modes?]

    b. What reports or other information do you receive?

    c. Do the data indicate whether the cases flagged using *[tool name]* were actually found to have payment errors?

3. We know this can be hard to determine, but what impact, if any, has *[tool name]* had on payment error rates?

   [Probe to understand whether this is their perception only or whether they have data to point to]

E. Local Context and Wrap-Up

This information has been very helpful. These last few questions will give me a little more understanding of your local office before we wrap up.

1. One thing I'm curious about is staff availability. In general, do you feel there are enough staff to process SNAP cases?

   [Yes, enough staff]

   a. And are there enough staff to follow up on the SNAP cases that *[tool name]* flags?

   [No, not enough staff]

   b. What types of staff do you feel are needed in your office to process SNAP cases?

      i. *[Supervisors/managers]* Which of those staff are most critical to keeping payment error rates low?

   c. What types of staff are needed to follow up on the SNAP cases that *[tool name]* flags?

      i. Are there enough of those staff in your office?

   d. How long has the office needed additional staff?

   e. [Probe: Since before COVID-19 or more recently?]

      How has the staff shortage affected the use of [tool name]?

   f. [For supervisors/managers] How has the staff shortage affected payment error rates?

2. *[For supervisors/managers]* Do you feel you have enough funding to process SNAP cases? Why or why not?

   a. Do you feel you have enough funding to follow up on the cases that *[tool name]* flags? Why or why not?

   b. *[Supervisors/managers]* How might the funding level affect the office's payment error rate?

3. *[For supervisors/managers]* We know that COVID-19 drastically affected what were formerly standard practices when moving a SNAP case through the eligibility determination process. For instance, offices largely stopped conducting certification interviews. How else did COVID-19 change SNAP processes in ways that may have affected payment error rates?

   [Note: Be clear on whether the impact on error rates was positive or negative]

4. Are there any key challenges to successfully building or implementing tools like *[tool name]* that we haven't already discussed?
5. *[For supervisors/managers]* Is there anything else about the *[tool name]*, or payment error rates in general, that you would like to share?

We've reached the end of the interview. Thank you so much for taking the time to talk with us and share your experiences. The information you provided gave us valuable insights into how tools like *[tool name]* work.

Appendix C

Supplemental Tables

This appendix contains the supplemental data tables created from SNAP State agencies' responses to the SNAP Risk Assessment Online Survey.

Table C.1. State agency use of RA tools

State agency	State agency use of RA tools			
	Currently uses an RA tool(s)	Previously used an RA tool but discontinued use	Developed but never implemented RA tools	Never developed or implemented RA tools
Alabama				●
Alaska				●
Arizona				●
Arkansas	●			
California				●
Colorado				●
Connecticut	●			
Guam				●
Hawaii				●
Idaho	●			
Illinois			●	
Indiana	●			
Iowa				●
Kansas	●			
Kentucky	●			
Louisiana				●
Maine				●
Maryland				●
Minnesota		●		

State agency	State agency use of RA tools			
	Currently uses an RA tool(s)	Previously used an RA tool but discontinued use	Developed but never implemented RA tools	Never developed or implemented RA tools
Mississippi				●
Missouri	●			
Montana				●
Nebraska	●			
Nevada				●
New Jersey				●
New Mexico	●			
New York				●
North Carolina				●
North Dakota	●			
Ohio				●
Oklahoma				●
Oregon				●
Pennsylvania				●
Rhode Island	●			
South Dakota				●
Texas				●
Utah		●		
Virgin Islands				●
Virginia	●			
Washington	●			
Washington, DC				●
West Virginia	●			
Wisconsin	●			
Wyoming		●		
Total (*n*)	15	3	1	25
Percent of State agencies	28.3%	5.7%	1.9%	47.2%

Note: Survey question SCREENER1
N = 53
Data source: SNAP Risk Assessment Study Online Survey

Table C.2. Adequate funding for RA tool

State agency	State agency has enough funding to administer tool
Arkansas	Yes
Connecticut	No
Idaho	Yes
Indiana	Unsure
Kansas	Yes
Kentucky	Unsure
Missouri	Unsure
Nebraska	Unsure
New Mexico	No
North Dakota	Unsure
Rhode Island	No
Virginia	Yes
Washington	Yes
West Virginia	Unsure
Wisconsin	Yes

Note: Survey question E1
n = 15
Data source: SNAP Risk Assessment Study Online Survey

Table C.3. Dates of use for SNAP RA tools

State agency	Implementation start date	Implementation end date
Arkansas	July 2012	Still in use
Connecticut	June 2021	Still in use
Idaho	October 2021	Still in use
Indiana	September 2024	Still in use
Kansas	November 2011	Still in use
Kentucky	June 2005	Still in use
Minnesota	2016	2019
Missouri	August 2020	Still in use
Nebraska	June 2023	Still in use
New Mexico	September 2023	Still in use
North Dakota	February 2023	Still in use
Rhode Island	June 2024	Still in use
Utah	2009	2024
Virginia	June 2023	Still in use
Washington	June 2012	Still in use
West Virginia	July 2012	Still in use
Wisconsin	April 2022	Still in use
Wyoming	2008	2023

Note: Survey questions A3 and B4
n = 18
Data source: SNAP Risk Assessment Study Online Survey

Table C.4. Motivations for developing RA tools currently in use

State agency	Motivation(s) for developing the RA tool				
	Concentrate resources on suspected high-risk cases[a]	Create formal process for identifying at-risk cases[b]	Address high PERs	Address audit findings[c]	Other
Arkansas	●	●	●		
Connecticut			●		●
Idaho					●
Indiana		●	●	●	
Kansas		●			
Kentucky	●		●		●
Missouri	●	●	●		
Nebraska	●		●		
New Mexico	●				
North Dakota					●
Rhode Island	●		●		
Virginia		●	●	●	
Washington	●		●		
West Virginia			●		
Wisconsin		●			
Total (*n*)	7	6	10	2	4
Percentage of State agencies currently using RA tools	46.7%	40.0%	66.7%	13.3%	26.7%

Note: Survey question A4
PER = payment error rate
[a] To concentrate resources (e.g., staff, funds, time) on only those SNAP cases suspected as being at high risk of payment error
[b] To create a formal process for identifying SNAP cases at risk for payment error
[c] To address the findings or recommendations of an audit or management evaluation
n = 15
Data source: SNAP Risk Assessment Study Online Survey

Table C.5. Data analyzed to determine high risk of payment error

State agency	SNAP Quality Control data	Other SNAP data	Proprietary vendor or contractor data	Other source	Don't know
Arkansas	●	●			
Connecticut		●			
Idaho	●	●			
Indiana	●	●			
Kansas	●				
Kentucky		●			
Missouri	●				
Nebraska	●			●	
New Mexico	●	●			
North Dakota				●	
Rhode Island	●	●			
Virginia	●	●			
Washington	●				
West Virginia	●	●			
Wisconsin	●	●			
Total (*n*)	12	10	0	2	0
Percentage of State agencies currently using RA tools	80.0%	66.7%	0.0%	13.3%	0.0%

Note: Survey question A8
n = 15
Data source: SNAP Risk Assessment Study Online Survey

Table C.6. Data analysis used to develop SNAP RA tools

State agency	Descriptive statistics	Regression modeling	Machine learning	Other	No data analysis	Don't know
Arkansas	●					
Connecticut					●	
Idaho						●
Indiana	●					
Kansas				●		
Kentucky				●		
Missouri						●
Nebraska	●			●		
New Mexico	●		●			
North Dakota					●	
Rhode Island	●					
Virginia	●					
Washington	●					
West Virginia						●
Wisconsin				●		
Total (*n*)	7	0	1	4	2	3
Percentage of State agencies currently using RA tools	46.7%	0.0%	6.7%	26.7%	13.3%	20.0%

Note: Survey question A17
n = 15
Data source: SNAP Risk Assessment Study Online Survey

Table C.7. Developers of RA tools currently in use

State agency	Tool developer(s)					
	SNAP State agency program/ policy staff	State IT staff	Local SNAP office staff	Vendor/ contractor	FNS Regional or National Office	Other
Arkansas	●	●	●	●		
Connecticut	●	●	●			
Idaho	●					
Indiana	●			●		
Kansas	●					
Kentucky	●	●		●		
Missouri	●				●	
Nebraska	●		●			
New Mexico				●		
North Dakota	●	●	●			
Rhode Island	●					
Virginia	●			●		
Washington	●	●				
West Virginia	●		●	●		
Wisconsin	●	●	●	●		●
Total (*n*)	14	6	6	7	1	1
Percentage of State agencies currently using RA tools	93.3%	40.0%	40.0%	46.7%	6.7%	6.7%

Note: Survey question A5
n = 15
Data source: SNAP Risk Assessment Study Online Survey

Table C.8. State agency involvement in tools currently in use developed by vendors/contractors

State agency	Phase of State agency involvement				
	Conceptualization or design phase	Development phase	Testing phase	Monitoring and evaluation	Other
Arkansas	●	●	●	●	
Indiana	●	●	●	●	
Kentucky	●	●			
New Mexico		●			
Virginia	●	●	●	●	
West Virginia					●
Wisconsin	●	●	●	●	●
Total (*n*)	5	6	4	4	2
Percentage of State agencies with RA tools developed by vendors/contractors	71.4%	85.7%	57.1%	57.1%	28.6%

Note: Survey question A6
n = 7
Data source: SNAP Risk Assessment Study Online Survey

Table C.9. Testers of implemented RA tools currently in use

State agency	Tool tester(s)					
	SNAP State agency program/policy staff	State IT staff	Local SNAP office staff	Vendor/ contractor	Other	Not applicable; tool was not tested
Arkansas	●	●	●			
Connecticut	●		●			
Idaho	●					
Indiana	●			●		
Kansas	●					
Kentucky	●	●				
Missouri						●
Nebraska	●		●			
New Mexico	●	●	●	●		
North Dakota	●	●	●			
Rhode Island	●	●	●	●		
Virginia	●		●			
Washington	●	●				
West Virginia	●		●			
Wisconsin		●		●	●	
Total (*n*)	13	7	8	4	1	1
Percentage of State agencies that implemented RA tools	86.7%	46.7%	53.3%	26.7%	6.7%	6.7%

Note: Survey question A7
n = 15
Data source: SNAP Risk Assessment Study Online Survey

Table C.10. Number of State agencies offering each type of training

Training type	Local staff		County staff		State agency staff	
	n	%	n	%	n	%
In-person	1	6.7	1	6.7	1	6.7
Live, virtual session	4	26.7	2	13.3	3	20
Online training without a live presenter or facilitator	6	40	4	26.7	6	40
Written tutorial	3	20	1	6.7	3	20
Other	2	13.3	0	0	0	0
Staff are not trained	1	6.7	0	0	2	13.3

Note: Survey questions A25, A28, and A31
n = 15
Data source: SNAP Risk Assessment Study Online Survey

Table C.11. Factors considered and included in currently active RA tools

Variable		State agencies that considered the variable		State agencies that included the variable in their tools		State agencies that don't know	
		n	%	*n*	%	*n*	%
Household composition	Total number of household members	11	73.3	9	60	1	6.7
	Presence/number of children in the household	3	20	3	20	1	6.7
	Presence/number of elderly in the household	2	13.3	2	13.3	2	13.3
	Presence of disabled household member	2	13.3	2	13.3	2	13.3
	Presence of ABAWD household member	3	20	3	20	2	13.3
	Presence of household member ineligible for SNAP	2	13.3	2	13.3	1	6.7
	Child-only unit	1	6.7	1	6.7	2	13.3
Demographic data	Age	1	6.7	1	6.7	2	13.3
	Race	0	0	0	0	2	13.3
	Ethnicity	0	0	0	0	2	13.3
	Sex	0	0	0	0	2	13.3
	Student status	2	13.3	2	13.3	2	13.3
	Level of education	0	0	0	0	2	13.3
	Employment status	6	40	5	33.3	1	6.7
	Marital status	1	6.7	1	6.7	2	13.3
	Homeless	1	6.7	1	6.7	2	13.3
	Residency status	2	13.3	1	6.7	2	13.3
	Citizenship status	0	0	0	0	2	13.3

Variable		State agencies that considered the variable		State agencies that included the variable in their tools		State agencies that don't know	
		n	%	*n*	%	*n*	%
Economic characteristics	Presence of earned income	12	80	10	66.7	0	0
	Presence of unearned income	11	73.3	8	53.3	0	0
	Zero income	8	53.3	7	46.7	1	6.7
	Gross income	9	60	6	40	1	6.7
	Net income	5	33.3	2	13.3	1	6.7
	Self-employment income	12	80	10	66.7	0	0
	Assets/resources	4	26.7	2	13.3	1	6.7
Household expenses	Medical expenses	6	40	5	33.3	0	0
	Non-utility shelter expenses	8	53.3	7	46.7	0	0
	Utility expenses (including SUAs)	10	66.7	7	46.7	0	0
	Dependent care expenses	7	46.7	5	33.3	0	0
	Homeless shelter deduction	6	40	4	26.7	0	0
	Excess shelter deduction	7	46.7	5	33.3	0	0
	Legally obligated child support	7	46.7	5	33.3	0	0
Case characteristics	Length of certification period	3	20	2	13.3	0	0
	Reporting requirements (e.g., change reporting, simplified reporting)	3	20	2	13.3	1	6.7
	Benefit amount	9	60	7	46.7	1	6.7
	New applicant	4	26.7	3	20	1	6.7

Note: Survey questions A9–A13
n = 15
ABAWD = able-bodied adult without dependents; SUA =Standard Utility Allowance
Data source: SNAP Risk Assessment Study Online Survey

Table C.12. Variables used in statewide RA tools

State agency	Variables
Indiana	*
Kansas	Household size 3–4
	Issuance $400 or more (nonprorated)
	Earned income $400 or more
Missouri	Earned income greater than $200
	Issuance of $200 or more
Rhode Island	Household size 5 or greater
	Earned income more than $1,400
	Unearned income more than $1,500
	Any self-employment income
	Any owner-occupied rental income
	Issuance more than $200

* Indiana did not provide details about the variables the State agency uses in its RA tool.
n = 4
Data source: SNAP Risk Assessment Study Online Survey

Table C.13. Formats of RA tools currently in use

State agency	Tool format					
	Written instructions	Paper checklist	Electronic checklist	Algorithm programmed into the eligibility system	Algorithm programmed into other systems or databases	Other format
Arkansas			●			
Connecticut						●
Idaho						●
Indiana				●		
Kansas	●	●	●			
Kentucky			●			
Missouri	●	●	●			
Nebraska	●					
New Mexico				●		
North Dakota	●	●		●		
Rhode Island				●		
Virginia	●		●			●
Washington	●		●		●	
West Virginia						●
Wisconsin				●		
Total (*n*)	6	3	6	5	1	4
Percentage of State agencies with RA tools	40.0%	20.0%	40.0%	33.3%	6.7%	26.7%

Note: Survey question A18
n = 15
Data source: SNAP Risk Assessment Study Online Survey

Table C.14. State agency allows local agencies to customize criteria used by RA tool

State agency	State agency allows local agencies to customize criteria used by RA tool
Arkansas	Yes
Connecticut	Yes
Idaho	Yes
Indiana	No
Kansas	No
Kentucky	Yes
Missouri	No
Nebraska	Yes
New Mexico	Yes
North Dakota	Yes
Rhode Island	No
Virginia	Yes
Washington	No
West Virginia	No
Wisconsin	Yes

Note: Survey question A2b
n = 15
Data source: SNAP Risk Assessment Study Online Survey

Table C.15. Time period when current tools are used to identify cases at risk for payment error

State agency	Time period when tool is used to identify cases at risk for payment error								
	At the point of application submission for new cases but before the interview	After the interview for new cases but before eligibility determination	After the eligibility determination for new cases but before benefits have been issued	After initial benefit issuance and before recertification (active cases)	At the point of application submission for recertification but before the interview	After the interview for recertification but before eligibility determination	After the recertification determination but before benefits have been issued	After benefit issuance for recertified cases (active cases)	Other time period
Arkansas				●				●	
Connecticut				●				●	
Idaho				●				●	
Indiana		●				●			
Kansas			●				●		●
Kentucky									●
Missouri		●	●	●		●	●	●	
Nebraska			●				●		
New Mexico		●	●			●			
North Dakota		●				●			
Rhode Island			●				●		●
Virginia									●
Washington			●				●		
West Virginia			●	●			●	●	
Wisconsin			●				●		
Total (*n*)	0	4	8	5	0	4	7	5	4
Percentage of State agencies with RA tools	0.0%	26.7%	53.3%	33.3%	0.0%	26.7%	46.7%	33.3%	26.7%

Note: Survey question A19
n = 15
Data source: SNAP Risk Assessment Study Online Survey

Table C.16. SNAP case file data used by active tools

State agency	Data from SNAP case file used by tool			
	Data from household application	Data from household interview	Data from data matches	Other data
Arkansas	●	●	●	
Connecticut	●	●	●	
Idaho	●	●	●	
Indiana	●	●	●	
Kansas	●	●		
Kentucky	●	●		
Missouri	●	●	●	
Nebraska	●	●		●
New Mexico	●	●		
North Dakota	●	●	●	
Rhode Island	●	●	●	
Virginia				●
Washington		●	●	
West Virginia	●	●		
Wisconsin	●	●	●	
Total (*n*)	13	14	9	2
Percentage of State agencies with RA tools	86.7%	93.3%	60.0%	13.3%

Note: Survey question A20
n = 15
Data source: SNAP Risk Assessment Study Online Survey

Table C.17. Staff who use RA tool to flag cases at risk for payment error

State agency	Local frontline eligibility workers	Local eligibility worker supervisors	County quality assurance staff	County eligibility workers	County eligibility worker supervisors	Other county-level staff	State quality assurance staff	State statisticians	State quality control staff	Other State-level staff	Not applicable
Arkansas					●					●	
Connecticut		●							●		
Idaho										●	
Indiana	●	●					●	●			
Kansas	●										
Kentucky					●		●			●	
Missouri	●	●					●		●		
Nebraska	●										
New Mexico	●	●					●	●			
North Dakota		●			●						
Rhode Island							●			●	
Virginia	●	●									
Washington											
West Virginia	●	●		●	●		●				
Wisconsin		●	●		●	●	●				
Total (*n*)	7	8	1	1	5	1	7	2	2	4	0
Percentage of State agencies with active RA tools	46.7%	53.3%	6.7%	6.7%	33.3%	6.7%	46.7%	13.3%	13.3%	26.7%	0.0%

Note: Survey questions A23, A26, and A29
n = 15
Data source: SNAP Risk Assessment Study Online Survey

Table C.18. Ways State agencies take action on cases flagged as at risk for payment error

State agency	Next step(s) for cases flagged as at risk for payment error				
	It undergoes a second review by an eligibility worker	It undergoes a second review by an eligibility worker supervisor	It undergoes a quality assurance review	Other	Don't know
Arkansas		●			
Connecticut		●	●		
Idaho			●		
Indiana	●				
Kansas		●			
Kentucky				●	
Missouri			●		
Nebraska		●			
New Mexico			●		
North Dakota			●		
Rhode Island		●			
Virginia	●	●			
Washington				●	
West Virginia					●
Wisconsin	●	●	●		
Total (*n*)	3	7	6	2	1
Percentage of State agencies with active RA tools	20.0%	46.7%	40.0%	13.3%	6.7%

Note: Survey question A32
n = 15
Data source: SNAP Risk Assessment Study Online Survey

Table C.19. Staff who follow up on cases at risk for payment error based on RA tool results

State agency	Local frontline eligibility workers	Local eligibility worker supervisors	Other local office staff	County quality assurance staff	County eligibility workers	County eligibility worker supervisors	State quality assurance staff	State statisticians	State quality control staff	Other State-level staff	Not applicable
Arkansas					●	●				●	
Connecticut	●	●							●		
Idaho										●	
Indiana	●	●					●	●			
Kansas		●									
Kentucky											●
Missouri	●	●					●		●		
Nebraska		●									
New Mexico	●	●					●	●			
North Dakota	●	●			●	●					
Rhode Island							●			●	
Virginia	●	●									
Washington			●								
West Virginia	●	●			●	●	●				
Wisconsin	●	●		●	●	●					
Total (*n*)	8	10	1	1	4	4	5	2	2	3	1
Percentage of State agencies with active RA tools	53.3%	66.7%	6.7%	6.7%	26.7%	26.7%	33.3%	13.3%	13.3%	20.0%	6.7%

Note: Survey questions A24, A27, and A30
n = 15
Data source: SNAP Risk Assessment Study Online Survey

Table C.20. State agency use of RA tool data for other purposes

State agency	Uses RA tool data for other purposes	Other purposes
Arkansas	No	N/A
Connecticut	No	N/A
Idaho	Yes	*"We use this data to make updates in our process manual and possibly training."*
Indiana	No	N/A
Kansas	No	N/A
Kentucky	No	N/A
Missouri	No	N/A
Nebraska	No	N/A
New Mexico	No	N/A
North Dakota	Yes	*"ND has an integrated eligibility system for SNAP, CCAP, LIHEAP, Medicaid and TANF. Error proofing is completed at the time of application and review for all programs in which a determination was made by a new eligibility worker or a worker new to a program."*
Rhode Island	No	N/A
Virginia	No	N/A
Washington	No	N/A
West Virginia	No	N/A
Wisconsin	Yes	*"The data can be analyzed to determine training needs for individuals or training needed overall."*

Note: Survey questions A33 and A33a
CCAP = Child Care Assistance Program; LIHEAP = Low Income Home Energy Assistance Program; N/A = not applicable; TANF = Temporary Assistance for Needy Families
n = 15
Data source: SNAP Risk Assessment Study Online Survey

Table C.21. State agency assessment of impact on error rates

State agency	State agency ascertains tool has impact on error rates (positive or negative)		
	Yes	No	Unsure
Arkansas	●		
Connecticut		●	
Idaho	●		
Indiana		●	
Kansas	●		
Kentucky	●		
Missouri	●		
Nebraska	●		
New Mexico	●		
North Dakota		●	
Rhode Island	●		
Virginia	●		
Washington	●		
West Virginia	●		
Wisconsin	●		
Total (*n*)	12	3	0
Percentage of State agencies with RA tools	80.0%	20.0%	0.0%

Note: Survey question A34
n = 15
Data source: SNAP Risk Assessment Study Online Survey

Table C.22. Frequency of effectiveness evaluations of tools currently in use

State agency	Frequency of tool effectiveness evaluations			
	Annually	More frequently than annually	Less frequently than annually	Has never evaluated effectiveness
Arkansas		●		
Connecticut				●
Idaho		●		
Indiana				●
Kansas			●	
Kentucky			●	
Missouri		●		
Nebraska		●		
New Mexico		●		
North Dakota				●
Rhode Island		●		
Virginia	●			
Washington		●		
West Virginia		●		
Wisconsin			●	
Total (*n*)	1	8	3	3
Percentage of State agencies with RA tools	6.7%	53.3%	20.0%	20.0%

Note: Survey question A35
n = 15
Data source: SNAP Risk Assessment Study Online Survey

Table C.23. Reason developed RA tools were not implemented

State agency	Reason developed tool was not implemented
Illinois	*"Module is still in development and testing phase."*

Note: Survey question C4
n = 1
Data source: SNAP Risk Assessment Study Online Survey

Table C.24. State agency consideration of RA tool development

State agency	Reason State agency has not considered developing RA tool	Does State agency plan to develop an RA tool
Arizona	—	—
California	*"California does not want to impact the QC sample by imposing bias or inequities toward certain household types."*	*"No, not at this time."*
Iowa	*"I am new in this role so unsure if it has been discussed in the past, however, I have never heard of this discussion in my prior role with HHS."*	*"Unknown at this time, I am new to this role. It is not ruled [out] and could/may be considered."*
Pennsylvania	—	*"PA has no plans to do so at this time."*
South Dakota	*"South Dakota has historically maintained a low payment error rate. Our current policies and procedures used to determine eligibility assist us in maintaining this low payment error rate."*	*"No, as stated in the previous question, our current policies and procedures used to determine eligibility assist us in maintaining a low payment error rate."*

Note: Survey questions D2a and D2b
— = State agency did not respond; HHS = Health and Human Services; QC = quality control
n = 5
Data source: SNAP Risk Assessment Study Online Survey

Table C.25. Reasons RA tools are no longer in use

State agency	Reason tool is no longer in use
Minnesota	*"It's been on pause since the pandemic started as we are currently focusing our SNAP-ME review efforts in other areas. We hope to bring it back into use again in the future."*
Utah	*"Utah discovered that the error-prone cases were less error-prone than others. As a result, the Performance Review Team (PRT) recently made changes to minimize using case-profiling tools."*
Wyoming	*"We switched to a universal caseload and at this time staff are focused on learning the steps for this. We do plan to implement tools in place in the future."*

Note: Survey question B5
n = 3
Data source: SNAP Risk Assessment Study Online Survey

Appendix D

Research Objectives and Questions Crosswalk

This appendix includes the research objectives and questions crosswalk. In the main text of this report, the study team consolidated discussion of research questions for objective 2 ("Determine what factors are being used in RA tools") and objective 3 ("Determine what variables States are using in their RA tools"). Factors are concepts or constructs (e.g., household size). Variables are operationalizations of constructs (e.g., four or more household members). The study team found during the study that State agencies made no practical distinction between "factors" and "variables." Furthermore, the study team judged that discussing factors and variables separately when their findings were nearly equivalent was more confusing than useful.

Table D.1. Understanding risk assessment in SNAP payment accuracy crosswalk of research questions and data collection instruments

Research questions	Administrative data request	Online survey	Interview guides			
			RA tool development lead	IT staff	Data analysis staff	Local office staff
Objective 1: Determine which States are using risk assessment (RA) tools to reduce error rates and who developed these tools						
1. Which States have implemented RA tools?		SCREENER1				
2. Who developed the RA tool implemented by the State?		A5–A7, B2, C2	C.2-3		D.1-4	
3. For States that have implemented RA tools, developed by contractors, how much input did the State agency have in the final algorithm for the tool?		A6	C.3	C.1		
4. Which States have developed but not implemented RA tools?		SCREENER1				
5. For States that have chosen not to design RA tools, how did they reach that decision?		D3–D3a				
6. What limitations and challenges has the State found in designing or implementing RA tools?			C.2-5, D.6, D.10	C.1-2, D.1-3	F.4	C.12-14
Objective 2: Determine what factors are being used in RA tools						
1. Do State officials know what factors are used in their RA tools?		A9–A13a, A15a	C.6-9		D.1-4	
2. What factors do States consider important when developing RA tools?		A16	C.7		D.2, D.4	
3. Does the standard SNAP QC data include all factors and variables that States feel are important in developing an RA tool?	●		C.7		D.2, D.4	
4. Do States use factors beyond those found in the standard SNAP QC data when developing their RA tools?	●	A9–A13a, A15a	C.7		D.2	
5. Were these factors recommended by the State or by a contractor?		A14–A15a	C.7		D.2	

Research questions	Administrative data request	Online survey	Interview guides			
			RA tool development lead	IT staff	Data analysis staff	Local office staff
Objective 3: Determine what variables States are using in their RA tools						
1. Do State officials know what variables are used by the RA tools?		A9–A13a, A15a	C.6-9		D.1-4	
2. Which of the standard SNAP QC variables are being used in RA tools?	●	A9–A13a, A15a	C.7		D.1	
3. Why did the State choose to include those variables in their RA tools?		A16	C.7-9		D.2-4	
4. Across States, are there patterns of variables in the standard SNAP QC data that States find effective for their RA tools?	●		C.7-9		D.2	
5. Why do States find these effective?		A16	C.7-9		D.2	
6. Across States, are there patterns of variables in the standard SNAP QC data that States find ineffective for their RA tools?	●		C.9		D.3-4	
7. Why do States find these ineffective?			C.8-9		D.3-4	
8. What variables beyond the standard QC data do States use in their RA tools?	●	A9a–A13a, A15a	C.6-7		D.1	
9. Among States that used a contractor for RA tool development and/or administration, does the tool use proprietary variables?	●	A15–A15a	C.6-7		D.1	
Objective 4: Identify how State agencies act on the results of their RA tools						
1. Do States use RA models/tools to determine what cases to review for payment accuracy? If so, how?		A2, A19–A21, A32	D.1, D.3			C.1-8
2. What other ways do State agencies use the results from RA tools?		A33–A33a	D.7-9		F.1-3	D.1-2
3. What staff are involved in implementing, reviewing, and taking action based on RA tool determinations? How are they trained?		A22–A31	D.2, D.5, D.7	D.1-4	f.1-10	C.1-10

Research questions	Administrative data request	Online survey	Interview guides			
			RA tool development lead	IT staff	Data analysis staff	Local office staff
Objective 5: Using QC data provided by FNS, determine whether States' RA tools are successful in reducing error rates						
1. What were the error rates for each State before the implementation of an RA tool in that specific State?	●					
2. What were the error rates for each State after the implementation of an RA tool?	●					
3. If an RA tool was implemented in different stages, how did error rates change when a State implemented the different parts?	●					
4. What other strategies did States use concurrently with the implementation of RA tools to reduce error rates?			D.12			
5. Overall, does it seem that implementation of RA tools helped reduce error rates?		A34	D.11		F.10	D.3
Objective 6: Determine best practices in development and use of RA tools						
1. Overall, compared with other strategies that have been used to lower error rates, are RA tools effective in lowering error rates?		Determined during analysis				
2. What factors and variables seem to be the most effective in building an RA tool?		Determined during analysis				
3. How can States evaluate whether their RA tool performs in a way that is practicable and cost-effective to States?		Determined during analysis				

QC = quality control

Appendix E

Considerations for Developing, Using, and Monitoring the Performance of RA Tools

Creating and implementing successful RA tools requires collaboration between many people, including, as relevant, program administrators, policy experts, IT staff, external contractors, frontline staff, supervisors, and statisticians. When everyone involved has a foundational understanding of RA tools, they can ask the right questions and identify how their expertise can best contribute to a successful RA tool. This appendix is designed to equip local, State, and Federal staff with an understanding of key RA tool concepts that can help them design, use, and oversee the use of RA tools to improve payment accuracy. Although this appendix uses examples from SNAP, the principles described here apply to other human service agency settings.

A. Key Questions to Ask When Developing an RA Tool

To develop a successful RA tool, agency staff need to ask the following questions:

- Which cases should the tool flag?
  - Should it flag all cases with a payment error?
  - Should it flag only cases with a payment error above a certain dollar amount?
  - Should it flag some other set of cases?
- Which case characteristics best differentiate between the cases that should and should not be flagged?
- What is the most accurate source of data to use for the RA tool?
- What performance metrics can tell you how well the RA tool is working?
- Once an initial tool has been developed, how well is it working?

- During development, how can you test whether the RA tool works better for some protected classes than others?
- Should local agencies be allowed to customize the tool?

1. Which cases should the tool flag?

The question of which cases an RA tool should flag may have more than one "right" answer; it may vary by agency. Administrators, policy experts, and others should collaborate to determine which cases the tool should flag. It may seem natural for RA tools to provide a yes/no flag indicating whether a case is likely to have a payment error. However, there are other options for RA tool output. For instance, an RA tool could be designed to output the probability that a given case has a payment error. If the agency has the capacity to review all the cases with at least a 50-percent probability of having a payment error, it can do so. If that would be too many cases to review, the agency could instead review only the cases with a higher probability of having an error (e.g., 70 percent). Yet another output to consider is the estimated dollar amount of the payment error on a case; that approach could help agencies prioritize reviewing the cases predicted to have the largest payment errors, thereby having the greatest effect on the PER.[42]

When agencies are first starting to explore their data to find an effective algorithm for flagging cases, it may be worthwhile to try algorithms that flag different types of cases and then compare (1) how well those algorithms work overall (refer to the "Once an initial tool has been developed, how well is it working?" section for more information), (2) how well they work for each protected class (refer to the "During development, how can you test whether the RA tool works better for some protected classes than others?" section for more information), and (3) how challenging it will be to implement a full RA tool using each algorithm (e.g., will it be easier to train staff to use one tool than another?).

2. Which case characteristics best differentiate between the cases that should and should not be flagged?

As part of their current efforts to minimize payment errors, agencies already identify the types of cases that most often have payment errors. For example, SNAP QC staff review representative samples of cases to identify those with payment errors and highlight which aspect(s) of the original case files were incorrect. SNAP State agencies can use this information to reduce their payment errors by implementing solutions like improved staff training on specific types of cases.

The type of data analysis required to develop an RA tool is similar to work that agencies already do, but it does have an important difference. In addition to focusing on the types of cases most likely to have payment errors, agencies also need to identify the types of cases least likely to have

[42] In SNAP, larger households have higher maximum benefit levels. Focusing on cases with larger predicted payment errors would likely result in reviewing more cases from larger households, which are the households more likely to have children. Under the USDA non-discrimination statement (USDA, n.d.), USDA programs are prohibited from discriminating on the basis of family/parental status; agencies should be cautious when considering this approach. For more detail on this topic, refer to the "During development, how can you test whether the RA tool works better for some protected classes than others?" section.

payment errors when developing an RA tool. In other words, effective RA tools need to be able to differentiate between cases with and without payment errors; it is not sufficient for them to identify all the cases with payment errors without regard for accurately determining which cases do not have payment errors.

To provide a concrete example of why it is important to examine cases without payment errors, consider a SNAP State agency whose QC data indicate that cases with payment errors tended to be in larger households. The issue is that among all large-household cases, there are likely many without payment errors, so the State agency would waste time reviewing all cases for large households. Household size alone, in this example, is not sufficient to accurately differentiate between cases with and without payment errors. If the State agency could identify the characteristics of cases that rarely have payment errors, the State agency could add that information to its RA tool to improve it. For example, the State agency's QC data might reveal that cases with only unearned income rarely have payment errors. As such, the State agency could modify the tool to flag large households with earned income and skip those with only unearned income. Making sure that an RA tool can (1) accurately flag cases with payment errors and (2) accurately skip those without payment errors ensures staff focus only on the cases most likely to need correction, thereby making efficient use of time.

Consider Using Multivariable Approaches

In the previous example, the State agency is conducting bivariate analyses; it is looking at one case characteristic at a time (variable 1) and assessing how strongly each is associated with payment error (variable 2). This is an important step regardless of the final methods used. However, agencies could consider additional analyses to develop their final RA tools. Statistical and machine learning models can use many variables simultaneously to identify the types of cases most likely to have payment errors. This approach can yield better performance than could be achieved by building an RA tool based on bivariate analyses alone.

Although multivariable methods hold great promise for developing accurate RA tools, this suggestion should not be misunderstood as a recommendation to use the most advanced and complicated methods available. In fact, many would recommend against that approach because those methods do not necessarily yield better results and can be

Not everyone needs to know how to do the data analysis, but everyone needs to understand the results

A statistician or other data expert will be able to conduct specialized analyses that can improve the final RA tool, but developing a good RA tool is not simply a math problem to be solved. Good RA tools also require policy knowledge and frontline program experience. It is crucial for everyone to understand the results of the data analysis and provide feedback based on their expertise. For example, State agency staff might be able to quickly spot results that conflict with their knowledge of the types of cases most and least likely to have payment errors. Raising these concerns can help ensure accurate data analysis and final RA tools that reflect a proper understanding of SNAP policy and practice.

so complicated that the results are difficult to fully understand.[43] Using methods that can readily be interpreted and understood, such as logistic regression models or decision trees (Rudin et al., 2022), can help RA tool developers refine the tools and make it easier to communicate how the final tool works, which is important for transparency (see "Transparent Programming and Design" in chapter 5).

3. What is the most accurate source of data to use for the RA tool?

Agencies gather information from a variety of sources to make payment decisions. It is critical to ensure that all data analyzed to develop an RA tool (and the data the RA tool will use going forward to flag cases at risk for a payment error) are accurate. For example, SNAP State agencies collect written applications, interview households, and conduct data matches. If a State agency knows a particular piece of information collected from written applications is frequently corrected during interviews, it would be advisable not to use that information from the written application for the RA tool. Similarly, agencies should carefully review any data collected by third parties to ensure their accuracy before using them in RA tools.[44]

4. What performance metrics can tell you how well the RA tool is working?

The first step to understanding how well an RA tool works is to identify what positive and negative effects it could have for the agency and its beneficiaries. Then, you can decide on the best way to measure those effects. Choosing appropriate tool performance metrics is critical because (1) tool developers will seek to optimize those performance metrics when creating the tool and (2) those metrics will be used to monitor the tool after it is implemented to ensure it works as well as anticipated.

There is no universal single metric or set of metrics best able to measure the performance of all RA tools. Some general metrics like accuracy (see the performance metrics textbox) can be useful, but other specialized performance metrics may provide more meaningful insight into specific positive and negative effects of an

Example RA Tool Performance Metrics

Accuracy: the number of cases classified correctly (i.e., cases with payment errors flagged as high risk plus cases without payment errors not flagged as high risk) divided by the total number of cases

Sensitivity: the proportion of cases with a payment error that are correctly flagged as high risk

Specificity: the proportion of cases without a payment error that are correctly not flagged as high risk

[43] See Rudin (2019), for example.

[44] Data from data brokers are often contested as inaccurate. For example, in a study of Facebook users, Venkatadri et al. (2019) found that Facebook users described more than 40 percent of data points about them from data brokers as "not at all accurate."

RA tool.[45] To find the best way to measure tool performance, the study team recommends everyone involved in developing an RA tool first work together to identify the potential effects of the tool. Then, the RA tool developers can identify metrics that best capture those effects.

To provide an example of how to first identify the potential effects of a tool and then choose performance metrics that capture those effects, consider a SNAP RA tool designed to flag cases likely to have payment errors. The SNAP State agency would collect a written application and conduct an interview with households applying to the program, and a caseworker would make an initial eligibility and benefit level determination according to standard procedure. Then, a caseworker could apply the RA tool to the case to see whether it flags the case as high risk. If so, a supervisor could review all the case information to ensure the eligibility and benefit level determination is correct before benefits are issued.

To determine the best ways to measure the performance of this SNAP RA tool, the first step is to think through the potential effects of the tool. The RA tool can have four results, each with its own unique potential effects on households and the agency: true positives, false positives, false negatives, and true negatives. These RA tool results are typically laid out in a table called a confusion matrix (see table 4.1). Not only are confusion matrices helpful for thinking through the potential effects of all results of an RA tool, but they also serve as a foundation for measuring tool performance.

Table E.1 describes the potential effects of the example SNAP RA tool on households and the State agency for the four outcomes. Note that for cases with payment errors (i.e., true positives and false negatives), the effects differ between overpayments and underpayments. The effects described in table E.1 are specific to the example RA tool described previously.[46] Effects for RA tools implemented in a different fashion or implemented in other programs with different policies may have different effects.

[45] See Hand (2012) and Canbek et al. (2022) for examples.

[46] Kansas, Missouri, and Rhode Island use their tools after they finish their standard processes to collect information from applicants and participants (e.g., after interviews, after receipt of a periodic report). In other words, these State agencies do not use their risk assessment results to ask additional questions during the interview or require additional verifications for high-risk cases. The reasoning explained here is why the study team chose to use sensitivity and specificity to measure RA tool performance in the analyses presented in chapter 4.

Table E.1. Effects of an example SNAP RA tool on households and the SNAP State agency

Outcome	Effect on household	Effect on State agency
True positive (underpayment)	Household benefits increase relative to the initial underpayment.	The State agency catches and corrects what would have been a payment error.
True positive (overpayment)	Household benefits decrease relative to the initial overpayment, but the household will not be subject to claims, as it would have been had the case not been flagged and the agency identified the overpayment later (e.g., through a QC review).	
False Positive	No effect on the household. The household's benefits will not change, and the household may not even be aware the State agency flagged and reviewed its case.	The State agency will invest staff time in reviewing the case, only to find there was nothing wrong with it.
False negative (underpayment)	The household receives lower benefits than it should. If the State agency reviews the case later (e.g., as part of SNAP QC), the agency will restore benefits for up to 12 months (7 CFR 273.17(b)). If the State agency does not review the case and identify the underpayment, the household will never receive the full benefits for which it qualified.	The case has a payment error. If it is sampled for QC, it will contribute to the State agency's PER.
False negative (overpayment)	If the State agency never identifies the overpayment, the household will receive higher benefits. If the State agency identifies the overpayment later (e.g., through a QC review), the agency will establish a claim on the household (7 CFR 273.18(a)(2)). Paying claims may impose hardship on SNAP participants because they have low incomes and frequently report food insecurity despite receiving SNAP benefits (Brady et al., 2023).	
True negative	No effect on the household or State agency. The State agency takes no further action on this case. If the State agency had taken further action on this case, nothing would have changed about the household's benefits.	

Note: PER = payment error rate; QC = quality control

In this example, from the household's perspective, the RA tool has no effect when there is no payment error (i.e., false positives and true negatives). However, there are meaningful effects for households that would have had payment errors had the RA tool not flagged their cases (i.e., true positives). There are also meaningful effects for any households not flagged that had payment errors (i.e., false negatives); they will be issued benefits in the incorrect amount. In this example, the most relevant metrics of RA tool performance for households should be based on the number of true positives and false negatives. A common measure that meets these criteria is ***sensitivity***, which is the proportion of cases with a payment error the RA tool flags as high risk. The higher the sensitivity, the more households will have the correct benefit amounts.

From the State agency's perspective, true positives are the value of having an RA tool: The greater the number of true positives, the more payment errors the State agency can catch and correct. False negatives are also important to State agencies because they represent missed opportunities. Had the RA tool been able to flag them, the State agency would have caught more payment errors. So, just as sensitivity is a valuable metric of RA tool performance for households, it also captures vital information about the RA tool from the State agency perspective.

One more RA tool outcome is critical for State agencies: false positives. False positives are a cost of having an RA tool because the greater the number of false positives, the more time the State agency spends reviewing cases that already had the correct benefit amount. As such, State agencies also have an interest in metrics of RA tool performance that incorporate information about the number of false positives. A common measure that serves this purpose is ***specificity***, which is the proportion of cases without a payment error that the RA tool classifies as low risk. The higher the specificity, the less time the State agency will spend reviewing cases with no payment errors.

In summary, the study team recommends a two-step process to determine the best way to measure how well an RA tool is working. First, agencies need to rely on their program and policy expertise to make their version of table E.1 specific to the RA tool they are considering. Note that table E.1 is merely an example; if a SNAP State agency implemented an RA tool in a way that resulted in different application, recertification, and/or reporting requirements for households flagged as high risk (e.g., a longer interview, more required verifications), the effect of the tool on the households would be different from what the table shows. Other differences between the example RA tool discussed here and a particular RA tool an agency is considering could also change table E.1.

Second, based on table E.1, State agencies can choose which RA tool performance metrics best measure how well the RA tool is working. Note that the study team encourages anyone involved in developing and overseeing the use of an RA tool to use the approach described here—not the conclusions drawn here. Sensitivity and specificity may not always be the best performance metrics for all RA tools. Weighing the pros and cons of different performance metrics requires policy and programmatic expertise as well as expertise in statistics, machine learning, or related fields. In practice, this step will require collaboration among many people.

5. Once an initial tool has been developed, how well is it working?

Before implementing any RA tool, it is best practice to test how well it will perform (The White House, 2022). Once an agency has developed an initial RA tool algorithm to flag cases, it should apply the initial tool to a set of cases to develop confusion matrices (see tables E.2 and E.3 for examples) and calculate the performance metrics selected to measure how well the tool works (see previous section for a discussion of selecting performance metrics). The ideal source of data to generate a confusion matrix is one that (1) includes a sample of cases representative of the population the RA tool will be used on and (2) has been reviewed by experts to determine which cases had payment errors. For most SNAP State agencies, QC data are likely the best option.

Assessing performance is critical for all State agencies, not only those with RA tools

All State agencies do at least some additional review of some of their cases (e.g., quality assurance reviews). If State agencies assess the performance of their method for selecting cases, they can refine their approach to maximize the benefits of these reviews.

When reviewing RA tool performance, the agency should check the cases the RA tool classified incorrectly (i.e., false positives and false negatives). If the agency can identify why these errors occurred, it can revise the RA tool algorithm to improve performance. Agencies should expect to iterate through many versions of an RA tool during the development stage to optimize the performance of the final tool selected for implementation.

Table E.2. Example confusion matrix for a proposed RA tool

RA tool result	SNAP QC result	
	Payment error (N) **Actual positives:** cases with payment errors	**No payment error (N)** **Actual negatives:** cases without payment errors
High risk **Predicted positives:** cases predicted to have payment errors	8	4
Low risk **Predicted negatives:** cases predicted not to have payment errors	42	46

Note: RA = risk assessment; QC = quality control

6. During development, how can you test whether the RA tool works better for some protected classes than others?

As agencies develop RA tools, they need to be aware of the legal implications of their tools. Tools that were not intended or designed to adversely affect any group of people may still do so, which

could be considered disparate impact (Polek & Sandy, 2023). While the legal standards in disparate impact lawsuits continue to evolve (Schwartz et al., 2022),[47] core to disparate impact claims is demonstrating that individuals in a protected class have been disproportionately harmed. It is therefore important for agencies to proactively test how well their tools work for different protected classes, which can give them the opportunity to adjust their tools as needed to avoid disparate impact. For instance, family/parental status is a protected class listed in the USDA non-discrimination statement (USDA, n.d.). To test for the potential for disparate impact by family/parental status, the State agency from the previous example (see tables E.1 and E.2) would want to calculate separate confusion matrices for households with and without children (table E.3).

Table E.3. Example confusion matrix for a proposed RA tool stratified by the presence of children in the household

RA tool result	SNAP QC result			
	Households with children		Households without children	
	Payment error (N) Actual positives: cases with payment errors	**No payment error (N)** Actual negatives: cases without payment errors	**Payment error (N)** Actual positives: cases with payment errors	**No payment error (N)** Actual negatives: cases without payment errors
High risk Predicted positives: cases predicted to have payment errors	8	1	0	3
Low risk Predicted negatives: cases predicted not to have payment errors	30	23	12	23

Note: RA = risk assessment; QC = quality control

The example in table E.3 illustrates two important points. First, it shows how a tool might perform better for some groups than others; all eight cases with payment errors correctly flagged by the tool (i.e., true positives) were among households with children. Second, it shows how the number of cases in each cell of the confusion matrix can become small when dividing the sample by protected class status. These small samples can make it hard to detect differences in tool performance across protected classes. It is critical to note that just because a statistical test shows no significant difference across protected classes does not mean there is truly no difference; the sample may just be too small to detect a difference.

[47] Similar to the work of Schwartz et al. (2022), this document is not meant to serve as legal guidance, but merely a reminder that RA tools have important legal implications. Agencies should work closely with legal experts to understand the current law in this area.

The study team recommends consulting legal experts and a statistician to help conduct and interpret tests for potential disparate impact.

7. Should local agencies be allowed to customize the tool?

Currently, most SNAP State agencies with an RA tool allow local agencies to customize the tool. When a local agency changes anything about the algorithm used to flag cases, a case flagged by one local agency may not necessarily be flagged by another. Potential benefits to this approach include local agencies having tools that better meet their specific situations, possibly yielding better tool performance and leading to greater payment accuracy. Potential drawbacks of this approach include the increased effort required to make data-driven customizations to a tool at the local agency level and to test and monitor multiple versions of an RA tool rather than a single version used by the entire State agency. Local agency customization may also create unintended consequences if some customized tools have higher performance than others, yielding differing payment accuracy for SNAP populations in different parts of the State. State agencies should weigh these possible benefits against the possible costs of local agency customization.

B. Key Considerations for Monitoring and Evaluating an RA Tool

As discussed in chapter 5, section E, ongoing monitoring and evaluation are critical for State agencies using RA tools. Key questions include the following:

- Once implemented, how can you tell if RA tools are performing as well as anticipated?
- How can you assess the overall effects of RA tools?

1. Once implemented, how can you tell if RA tools are performing as well as anticipated?

To get real-time feedback on how well an RA tool is working, agencies can track the number of cases the RA tool flagged and how many of those flagged cases initially had payment errors that agency staff corrected upon review. To use the example from table E.2, 8 of the 12 flagged cases had payment errors. The agency could therefore expect approximately two-thirds of its flagged cases to have payment errors its staff can correct. When the agency sees a decrease in the proportion of flagged cases that require correction, it is likely time to revisit the tool's algorithm and make updates as needed. Tracking which flagged cases have errors corrected during a review serves a second benefit: helping agencies identify what types of errors are happening in real time so the agency can provide training, make policy updates, or take other actions to improve accuracy.

In addition to real-time tracking, it is best practice to periodically generate updated confusion matrices to provide a full assessment of RA tool performance and update the tool as needed to ensure continued high performance (The White House, 2022). For SNAP State agencies, just as

was true at the tool development stage, QC data will most likely be the best data source. However, SNAP State agencies need to add at least two data points generated during the implementation of the RA tool:[48]

1. The case ID or another identifier for all cases flagged as high risk
2. The original benefit amount determined for all flagged cases (i.e., before any review resulting from the case getting flagged)

The addition of these two data points allows data users to distinguish between (1) cases that initially had an incorrect eligibility or benefit determination that was corrected after the case was flagged and (2) cases that always had a correct eligibility and benefit determination. Once these data points are merged with the QC data, the State agency can calculate the difference between the original benefit amount and the correct benefit amount as determined during the QC review to determine whether the case would have had a payment error had the original benefit amount been issued. The agency can then compare which cases the RA tool flagged as high risk against which cases would have had a payment error in the absence of the RA tool, generating a confusion matrix that accurately shows when the RA tool made correct and incorrect classifications.

Agencies can use the updated confusion matrices to check the performance of their tools and test how changes to their tools would affect performance. Any change in the number or type of errors—because of changes in the population of program beneficiaries, policy changes, system changes, staff turnover, or other reasons—can affect RA tool performance. As a result, when enough time has passed for a large portion of the caseload to have churned or after an important change in policy or practice, it is advisable to generate new confusion matrices and conduct a full assessment of RA tool performance.

2. How can you assess the overall effects of RA tools?

In addition to monitoring tool performance, it may be valuable to assess changes in the PER attributable to RA tool implementation. Agencies should begin by reviewing PERs before and after they introduced their tools. To estimate the impact, agencies can consider using quasi-experimental methods, such as the difference-in-differences model discussed in chapter 4.[49] Accurately estimating the impact of an RA tool would require staff time, technical expertise, and rigorous data collection on a variety of factors aside from RA tool implementation that could affect PERs (e.g., changes in the population of program beneficiaries, program policy changes).

Agencies should also document (or precisely estimate) all costs related to developing, testing, implementing, monitoring, and evaluating their RA tools. They should then compare these costs against the dollar amounts of payment errors corrected and any possible liabilities they might otherwise have needed to pay for high PERs. This type of evaluation can help agencies determine

[48] Other data points can also be valuable for RA tool monitoring, but the utility of other data may depend on the specifics of the State agency and its RA tool.

[49] Agencies could also use these methods for other outcomes of interest, such as the proportion of cases that have payment errors.

whether they are generating a positive return on investment from their RA tool. Agencies can consider tool performance, changes in the PER, and tool cost-effectiveness to create a holistic understanding of the tool's strengths and areas for improvement, informing next steps to improve payment accuracy most effectively.