

Program Integrity: Data Mining, Sharing, and Analytics

Lead Agencies are responsible for having internal controls in place that ensure program integrity and accountability while maintaining continuity of services. The internal controls strategies should include elements related to the detection, prevention and prediction of fraud, waste, and abuse within the Child Care and Development Fund (CCDF) program. Using information technology (IT), three common program integrity elements can assist states, territories, and tribes in having effective internal controls:



All three elements are achievable through the processes of data mining, sharing, and analytics. These types of technological innovations within government programs are practical strategies that support program integrity and accountability. Integrating collaboration between program leadership, stakeholders, management, and staff on data solutions can build successful technical systems. This brief provides a summary of common ways to use data mining, sharing, and analytics within CCDF.

Data Mining

At a basic level, data mining is the extraction of information from a data set or sets. Given the complexity of eligibility, enrollment, payment, and provider systems, data mining drills down into large data sets and assists the program office in discovering patterns or trends in the data. Data mining may also incorporate the use of algorithms to evaluate those large data sets. During the review, the algorithms detect patterns or red flags as outlined in the logic of the software. Red flag reports are the results of the principles in the programming logic set by an IT, data, or systems specialist.

Data mining can assist the Lead Agency in identifying potential improper payments before the release of the subsidy payment. Data mining may also provide exception reports which are crucial parts of internal controls and support the detection element. These reports give administrators, quality control staff, and supervisors a first-hand look at if the processes and policies

implemented are correctly performed by staff. Here are some types of red flag and exception reports:

Red Flag and Exception Report Examples

Enrollment exceeds licensed capacity	Provider and participant have matching address, phone, or birthdate	Disproportionate eligibility worker authorizations or overrides
Providers consistently report perfect attendance or are never closed	Parents employed by providers where their children also attend	School-age children receiving full-time care during the school year
Children who have enrollment or attendance records outside of the provider's operating hours	Payments sent to a PO box	Provider payment rate or copayment discrepancies
Provider's license has expired	Duration of eligibility less than 12 months	Family co-payment not waived in appropriate circumstances
Same child billed at multiple locations	Application processing delays	Parent income discrepancies

Data Sharing

Data sharing is an effective means to verify applicant eligibility requirements. Through use of a data sharing process, a Lead Agency can review a family's submitted application and verifications against information reported to other agencies. This review can assist a Lead Agency in ensuring that families and children are eligible to receive benefits. In addition, data sharing supports the prevention element of internal controls as it allows the state, territory, or tribe to review the accuracy of information provided from the potential recipient of subsidy funding. Data sharing may incorporate the collaboration of federal and state programs through database linkages or through the ability to view another agency or organization's data system. Examples of eligibility requirements that may be verified through a data sharing process include household composition or size, earned and unearned income, and employment.

Data sharing with other agencies or organizations can be accomplished through a written agreement such as a Memorandum of Understanding or Memorandum of Agreement. Lead Agencies must ensure the confidentiality of personally identifiable information of providers and applicants during the data sharing process. Below are potential data matching options for CCDF:

Data Sharing and Record Matching Options

Child support enforcement	Early Head Start/Head Start	Internal Revenue Service
Medicaid	Motor vehicle registration	Non-public organizations
Pre-kindergarten (pre-K)	Public Assistance Reporting Information System (PARIS)	Secretary of State (verify business records)
Social Security Administration	Supplemental Nutrition Assistance Program (SNAP)	Temporary Assistance for Needy Families (TANF)
Vital statistics	Wage and unemployment compensation benefits (state labor departments)	

Data Analytics

Data analytics, also known as data analysis, is seen as the predictive element in internal controls. The data analytics process includes a review of the administrative data a program possesses and an evaluation of the data for appropriate use. This technology approach includes gathering large data sets, cleaning up the data (i.e. correcting or removing inaccurate records sets, tables or irrelevant parts of the data), executing an analysis through a designed methodology, and then reviewing the results for irregularities, trends, or any predefined risk indicator. Before use of a data analytics tool, the program office can determine the required data sources and prepare a data request that may include specific table names, desired data output, and time frame.

Here is a simple example of data analytics: A Lead Agency wants to identify areas of high risk related to child care payments. First, through data mining, the agency combines all their administrative program data, identified fraud and program violation occurrences, and improper payment data. The agency uses data analytics software to determine that their high-risk areas include improper provider billing, untimely application processing, and altered client wage documentation. In addition, the agency can identify specifically where these occurrences happened and where they are likely to happen in the future. Utilizing the results of data analytics, the Lead Agency implements additional internal controls and accountability measures to reduce the number of high-risk occurrences.

Through data analytics, performing a predictive analysis can provide Lead Agencies with information on where to close gaps in program procedures that could lead to fraud, program violations, and administrative errors. Also, the use of data analytics can provide agencies with official data related to program performance and financial statistics. Data analysis is a useful



Why include an indicator?

Including a fraud risk indicator can enhance a Lead Agency's ability to detect fraud, waste, and abuse.

multi-functional tool for reporting and gaining data on program functions, performance, and areas for improvement.

Data Analytics Methodology Example

An example of data analysis is a risk assessment tool that categorizes agency risks by impact and probability: low impact/low probability, high impact/high probability, low impact/high probability, high impact/low probability. The level of probability assigned is based on historical occurrences of the event within the program. The level of impact assigned is based on qualitative data and may include the financial outcome of an occurrence. Occurrences that create a financial liability to the Lead Agency are areas for process improvement. Data analysis can help Lead Agencies in developing and implementing measures to reduce these financial liability occurrences. Proactive strategies can assist in ensuring the program operations are effective and efficient while maintaining program integrity.

It is important to note that data analytics risk factors may require collaboration between different departments such as IT, fiscal, eligibility, program managers, and leadership.

Data Analytics Results Table Example

Risk Category	Risk Factor Ranking	Qualitative Data	Variable
High Probability/ High Impact	High	<p>FY 16: Agency received 5,775 fraudulent verification documents online. Agency cost - \$59,117 in improper payments.</p> <p>FY 17: Agency received 6,353 fraudulent verification documents online. A 10% increase from FY 16. Agency cost - \$65,029 in improper payments.</p> <p>FY 18: Agency received 6,930 fraudulent verification documents online. A 20% increase from FY 16. Agency cost - \$70,940 in improper payments.</p>	Online verification submission
High Probability/ Low Impact	Low	<p>FY 16: 349 child care applications did not process according to the agency processing time of 30 days. Agency cost - \$6,263 in improper payments.</p> <p>FY 17: 391 child care applications did not process according to the agency processing time of 30 days. A 12% increase from FY 16. Agency cost - \$7,015 in improper payments.</p> <p>FY 18: 412 child care applications did not process according to the agency processing time of 30 days. A 18% increase from FY 16. Agency cost - \$7,390 in improper payments.</p>	Application processing delays

Risk Category	Risk Factor Ranking	Qualitative Data	Variable
Low Probability/ High Impact	High	<p>FY 16: 19 out of 1,300 providers billed and were paid from subsidy funds for hours outside of their child care certified hours. Agency cost - \$159,600 in improper payments.</p> <p>FY 17: 21 out of 1,560 providers billed and were paid from subsidy funds for hours outside of their child care certified hours. A 10% increase from FY 16. Agency cost - \$175,560 in improper payments.</p> <p>FY 18: 27 out of 1,820 providers billed and were paid from subsidy funds for hours outside of their child care certified hours. A 40% increase from FY 16. Agency cost - \$223,440 in improper payments.</p>	Provider billing outside of license hours
Low Probability/ Low Impact	Low	<p>FY 16: 3 out of 8,425 parents are employed at the same provider site where their child attends part time. Of those, investigators verified one parent to also lived with the provider. Agency cost - \$2,340 in improper payments</p> <p>FY 17: 4 out of 8,425 parents are employed at the same provider site where their child attends part time. Of those, investigators verified two of the parents also lived with the provider. A 50% increase from FY 16. Agency cost - \$4,680 in improper payments.</p> <p>FY 18: 5 out of 8,425 parents were employed at the same provider site where their child attends part time. Of those, investigators verified one parent also lived with the provider. A 0% increase from FY 16. Agency cost - \$2340 in improper payments.</p>	Parent working at provider site

Risk factor rankings can be numerical or contextual depending on the level of detail outlined by the Lead Agency.

Summary

Data mining, sharing, and analytics assists the Lead Agency through cost-effective means in ensuring that the program operations are efficient and effective. A Lead Agency can identify areas of risk and make changes in program operations to mitigate those risks by incorporating strategies that include data processes that provide results to the targeted elements of detection, prevention and prediction of program risk and fraud. In addition, the use of data sharing, mining, and analytics may contribute to the reduction of improper payments while helping states to be good stewards of federal and state funds. Finally, Lead Agencies may develop strong partnerships with other agencies serving shared communities.



Other Resources

Lead Agencies may be interested in the following additional resources to help in the development and implementation of data usage techniques.

Public Assistance Reporting Information System (PARIS)

<https://www.acf.hhs.gov/paris>

Interoperability and Data Sharing

<https://www.acf.hhs.gov/opre/research/topic/overview/interoperability-and-data-sharing>

Interoperability

https://www.acf.hhs.gov/about/interoperability?utm_source=OPRE+2018+in+Review+&utm_campaign=OPRE%3A+2018+in+Review+-+%5BEnd+of+Year+Message+12%2F27%2F18%5D&utm_medium=email

Data Analytics for Beginners

<https://www.youtube.com/watch?v=THODdNXOjRw>

Data Exchange Standards for Improved Interoperability of Multiple Human Service Programs

<https://www.federalregister.gov/documents/2018/11/08/2018-24459/data-exchange-standards-for-improved-interoperability-of-multiple-human-service-programs>

Social Security Data Exchange

<https://www.ssa.gov/dataexchange/>

What Can Data Do for Me? Using Data for Decision-Making and Story-Telling: STAM 2015

<https://childcareta.acf.hhs.gov/resource/what-can-data-do-me-using-data-decision-making-and-story-telling-stam-2015>

Use of Technology to Enhance Licensing Administration

<https://childcareta.acf.hhs.gov/resource/use-technology-enhance-licensing-administration>

Virginia's Automation of Child Care Subsidy

https://childcareta.acf.hhs.gov/sites/default/files/public/virginias_automation_of_child_care_subsidy.pdf

Webinar Series-Quality Investments: How to Use Administrative Data and Other Sources to Evaluate What is Working

<https://childcareta.acf.hhs.gov/resource/webinar-series-quality-investments-how-use-administrative-data-and-other-sources-evaluate>