GAO U.S. GOVERNMENT ACCOUNTABILITY OFFICE

HIGHLIGHTS OF A FORUM

Data Analytics
to Address Fraud and Improper Payments

CONVENED BY THE
COMPTROLLER GENERAL
OF THE UNITED STATES

March 2017 | GAO-17-339SP

Data Analytics to Address Fraud and Improper Payments

Why GAO Convened This Forum

In fiscal year 2016, federal agencies estimated making over $144 billion in improper payments, a number that has increased considerably in recent years. Improper payments, including those resulting from fraud, are a significant and pervasive government-wide issue and have presented a continuing challenge to the federal government. Reducing improper payments is critical to safeguarding federal funds, and could help achieve cost savings and help improve the government's fiscal position.

GAO's previous work has shown that implementing preventive and detective controls, including data-analytics tools and techniques, is one strategy to address improper payments, including fraud.

On September 21, 2016, GAO convened a group of data-analysis experts to discuss: (1) opportunities for collaboration, (2) considerations for establishing a data-analytics program, and (3) considerations for refining a data-analytics program. GAO selected 12 forum panelists from the government, private sector, public–private partnerships, and academia based on their demonstrated ability to provide examples of their work and discuss topics related to using data analytics to address improper payments, including fraud. This report summarizes the discussion by forum panelists. Comments expressed in the report do not necessarily represent the views of all participants, their organizations, or GAO. Panelists reviewed a draft of this report, and GAO incorporated their comments, as appropriate.

What the Participants Said

During the GAO-sponsored forum, panelists discussed the following topics related to analytics to address improper payments, including fraud:

▶ **Opportunities for Collaboration**

Collaboration between the government, private sector, public–private partnerships, and academia may allow entities to share analytics-related resources and knowledge to address the challenge of improper payments, as shown in the figure below. To foster collaboration, entities in a prospective collaboration should identify the defined objectives and desired outcomes of their partnership, according to panelists. Further, entities should work to build trust with each other and demonstrate a willingness to collaborate. Panelists stated that many commonly cited barriers to collaboration, including potential legal and data-sharing barriers, may be overcome through communication and research.

Examples of Partnerships and Resources Provided by the Panelists

The Healthcare Fraud Prevention Partnership (HFPP)

As a voluntary public–private partnership between the federal government, state agencies, law-enforcement organizations, private health-insurance plans, and health-care antifraud associations, the HFPP aims to foster a proactive approach to detect and prevent health-care fraud by exchanging data and information between the public and private sectors, leveraging various analytic tools against data sets provided by HFPP partners, and providing a forum for public and private leaders and subject-matter experts to share successful antifraud practices and effective methodologies.

The Consolidated Data Analysis Center (CDAC)

Housed at the Department of Health and Human Services (HHS) Office of Inspector General (OIG), CDAC started as an interdisciplinary pilot project designed to support the expansion of the Medicare Fraud Strike Force by leveraging knowledge from OIG staff with experience in audits, evaluations, and investigations. CDAC provides the OIG with data analysis, advanced analytics, and data stewardship to support fraud prevention, enforcement, and recovery efforts.

The Institute for Advanced Analytics at North Carolina State University

The institute works with the public and private sectors to address data-analytics challenges. According to a panelist, the institute received 60 requests for assistance within the past year and is currently working on 24.

The Interagency Risk and Fraud Data Mining Group

Composed of federal investigators and auditors, this group shares best practices, raises awareness, and offers a forum for the evaluation of data mining and risk-modeling tools and techniques to detect fraudulent patterns and emerging risks.

Source: GAO. | GAO-17-339SP

▶ **Considerations for Establishing Analytics Programs**

Forum panelists identified several considerations for entities in the early stages of developing an analytics program to address improper payments, including (1) identifying specific objectives and aligning analytics programs with those objectives; (2) conducting an inventory of existing technology, techniques, data, and staff, and identifying new opportunities to leverage existing capabilities; and (3) demonstrating the value of analytics to obtain entity-wide support.

▶ **Considerations for Refining Analytics Programs**

Panelists highlighted the importance of continuously improving analytics programs by revisiting and reassessing analytics tools, techniques, and available data sets. Additionally, panelists stated that entities should identify staff with knowledge of both analytics and the business of the entity to further the entity's mission and meet objectives.

Contents

Abbreviations

CDAC	Consolidated Data Analysis Center
CIA	Central Intelligence Agency
CMS	Centers for Medicare & Medicaid Services
CPI	Center for Program Integrity
DOD	Department of Defense
FEMA	Federal Emergency Management Agency
FHA	Federal Housing Administration
FNS	Food and Nutrition Service
HFPP	Healthcare Fraud Prevention Partnership
HHS	Department of Health and Human Services
HUD	Department of Housing and Urban Development
IDR	Integrated Data Repository
OIG	Office of Inspector General
OMB	Office of Management and Budget
SNAP	Supplemental Nutrition Assistance Program
USDA	United States Department of Agriculture
USPS	United States Postal Service

GAO

March 31, 2017

Congressional Addressees

In fiscal year 2016, federal agencies estimated making over $144 billion in improper payments, which includes payments that should not have been made or that were made in an incorrect amount. Such payments are a significant and pervasive government-wide issue and present a continuing challenge to the federal government. Since fiscal year 2003—when certain agencies were required by statute to begin reporting improper payments—cumulative improper payment estimates have totaled over $1.2 trillion.[1] Reducing improper payments, including those that are the result of fraud, is critical to safeguarding federal funds, and could help achieve cost savings and help improve the government's fiscal position.

As stewards of taxpayer dollars, federal managers have the ultimate responsibility for overseeing how government funds are spent. We have previously reported on ways they might do so more effectively.[2] Our prior work highlights a set of leading practices that may serve as a guide for program managers to use when developing or enhancing efforts to combat fraud in a strategic, risk-based manner.[3] Laws, including the Fraud Reduction and Data Analytics Act of 2015, and guidance by the Office of Management and Budget (OMB), including the revised OMB Circular A-123, have increasingly focused on the need for program

---

[1]GAO has found that recent laws and guidance have focused attention on improper payments, but incomplete or understated estimates and noncompliance with criteria listed in federal law hinder the government's ability to assess the full extent of improper payments and implement strategies to reduce them. See GAO, *Fiscal Outlook: Addressing Improper Payments and the Tax Gap Would Improve the Government's Fiscal Position*, GAO-16-92T (Washington, D.C.: Oct. 1, 2015).

[2]For example, see GAO, *Improper Payments: CFO Act Agencies Need to Improve Efforts to Address Compliance Issues*, GAO-16-554 (Washington, D.C.: June 30, 2016), and GAO-16-92T, among others.

[3]See GAO, *A Framework for Managing Fraud Risks in Federal Programs*, GAO-15-593SP (Washington, D.C.: July 2015).

managers to take a strategic approach to managing improper payments and risks that include fraud.[4]

One element of a strategic approach is implementing preventive and detective controls, including data analytics. Data analytics involves a variety of techniques to analyze and interpret data to facilitate decision making and may be used to identify patterns or trends, determine whether problems are widespread and systemic in nature, and evaluate program performance and outcomes.

On September 21, 2016, we convened a forum to advance the intergovernmental dialogue on implementing data analytics to prevent and detect improper payments, including fraud.[5] We selected 12 panelists, including individuals from across government, the private sector, public–private partnerships, and academia, based on their demonstrated ability to discuss and provide examples of using data analytics to address improper payments.[6] In addition to these panelists, we invited guests, primarily from the federal government, to observe the forum and learn from the panel discussions.

Forum panelists highlighted the following three topics for discussion:

- Establishing an analytics program to address improper payments.
- Refining an analytics program to address improper payments.
- Identifying opportunities for collaboration among analytics programs.

This report summarizes the discussion by forum panelists and guests, highlighting key ideas that emerged during the discussion sessions. It is not intended to represent an exhaustive list of all ideas and themes that

---

[4]The Fraud Reduction and Data Analytics Act of 2015 requires OMB to establish guidelines for federal agencies to establish financial and administrative controls to identify and assess fraud risks and design and implement control activities in order to prevent, detect, and respond to fraud, including improper payments. Additionally, the law states that these guidelines shall incorporate the leading practices identified in GAO's Framework for Managing Fraud Risks in Federal Programs. Pub. L. No. 114-186 (June 30, 2016); 31 U.S.C. § 3321 note.

[5]For purposes of this report, when we use the term "improper payments," we are referring to all improper payments, including those considered fraud.

[6]Although we originally selected 13 panelists for the forum, 1 panelist was unable to attend at the last minute. A total of 12 panelists spoke at the forum.

emerged during the forum. Further, the information presented in this report does not necessarily represent the views of all participants or the views of their organizations, including GAO. We structured the forum so that all panelists could openly comment on issues, although not all panelists commented on all topics. Throughout the report, we refer to "panelists" in order to indicate views expressed by more than one panelist. After the forum, we conducted several follow-up interviews with selected panelists in order to obtain further clarification on statements made during the forum, and provided an opportunity for the panelists to comment on a draft of this report.

Appendix I provides more details on the scope and methodology used to plan and conduct the forum and prepare this report. Appendix II provides a list of forum panelists and their affiliations, and appendix III provides the forum agenda. Appendix IV provides the issue brief sent to panelists prior to the forum.

We conducted our work from December 2015 to March 2017 as part of our ongoing body of work on improper payments, including fraud, and in accordance with all sections of GAO's Quality Assurance Framework that are relevant to our objectives. The framework requires that we plan and perform the engagement to obtain sufficient, appropriate evidence to meet our stated objectives and discuss any limitations in our work. We believe that the information and data obtained, and the analysis conducted, provide a reasonable basis for this product.

This report is available at no charge on the GAO website at http://www.gao.gov.

If you or your staff have any questions about this report, please contact Vijay D'Souza at (202) 512-2700 or dsouzav@gao.gov or Seto Bagdoyan at (202) 512-6722 or bagdoyans@gao.gov. Contact points for our Offices of Congressional Relations and Public Affairs may be found on the last page of this report. GAO staff who made key contributions to this report are listed in appendix V.

I wish to thank all of the forum participants for their time and thoughtful contributions to the forum discussion. The discussion enhanced our understanding of issues related to the use of data analytics to address improper payments, and we will all benefit from these insights as we carry out our work on this issue.

Gene L. Dodaro
Comptroller General of the United States

List of Addressees

The Honorable Greg Walden
Chairman
The Honorable Frank Pallone
Ranking Member
Committee on Energy and Commerce
House of Representatives

The Honorable Lamar Smith
Chairman
Committee on Science, Space, and Technology
House of Representatives

The Honorable Jeff Flake
Chairman
The Honorable Al Franken
Ranking Member
Subcommittee on Privacy, Technology, and the Law
Committee on the Judiciary
United States Senate

The Honorable Barbara Comstock
Chairwoman
Subcommittee on Research and Technology
Committee on Science, Space, and Technology
House of Representatives

The Honorable Christopher Coons
United States Senate

The Honorable Mark Takano
House of Representatives

Considerations for Establishing Analytics Programs

Forum panelists identified several considerations for entities[7] establishing analytics programs to address improper payments including (1) identifying specific objectives and aligning analytics programs with those objectives; (2) conducting an inventory of existing technology, techniques, data, and staff and identifying new opportunities to leverage existing capabilities; and (3) demonstrating the value of analytics to obtain entity-wide support.

Identify Objectives and Align Analytics Programs with Those Objectives

Panelists suggested that entities in the early stages of developing an analytics program should identify how the specific objectives of an analytics program may help to meet an entity's business needs. To help ensure analytics work meets an entity's objectives and business needs, one panelist highlighted the importance of establishing a data-governance structure within an analytics program.[8] A data-governance structure can provide the analytics program with an opportunity to review potential analytics work to determine whether it is aligned with the entity's objectives and to identify the costs and benefits of undertaking the work. Additionally, a data-governance structure can guide analytics to ensure all work is done according to appropriate laws and guidelines, including the Privacy Act and entity policies.

Inventory Existing Analytics Technology, Techniques, Data, and Staff and Identify Opportunities to Leverage Existing Capabilities

In addition to identifying business objectives, panelists stated that entities looking to establish an analytics program should inventory their current resources, including analytics technology, techniques, data, and staff, and look for opportunities to leverage existing capabilities.

- *Technology and Techniques.* Panelists recommended that entities should inventory currently available analytics systems and software tools to avoid duplicating existing software capabilities. During this inventory, entities may identify many types of analytics techniques that can be used to address improper payments. For example, figure

---

[7]For purposes of this report, the term "entity" may describe an organization from the public, private, or nonprofit sector.

[8]Data governance may be defined as an institutionalized system of decision rights and accountabilities for planning, overseeing, and controlling data management. For more information, see GAO, *DATA Act: Progress Made in Initial Implementation but Challenges Must be Addressed as Efforts Proceed*, GAO-15-752T (Washington, D.C.: July 29, 2015), and *DATA Act: OMB and Treasury Have Issued Additional Guidance and Have Improved Pilot Design but Implementation Challenges Remain*, GAO-17-156 (Washington, D.C.: Dec. 8, 2016).

1 shows four techniques that the Centers for Medicare & Medicaid Services (CMS) uses and how each technique may be used to help identify improper payments.

Figure 1: Examples of Analytic Techniques

Technique Type		Medicare Example
	Rules-Based Filter fraudulent claims and behaviors with rules	A provider that bills using a Medicare identification number that was previously stolen and used improperly
	Anomaly Detect individual and aggregated abnormal patterns versus peer group	A provider that bills for more services in a single day than the number of services that 99 percent of similar providers bill in a single day
	Predictive Assess against known fraud cases	A provider that has characteristics similar to those of known bad actors
	Network Discover knowledge with associative link analysis	A provider that is linked to known bad actors through addresses or phone numbers

Source: Centers for Medicare & Medicaid Services. | GAO-17-339SP

After a data-analytics program inventories its current technological capabilities, the program should identify opportunities to leverage additional resources available throughout the entity. For example, the Department of Health and Human Services (HHS) Office of Inspector General (OIG) partnered with CMS to utilize existing HHS investments for the CMS Center for Program Integrity (CPI). Specifically, HHS OIG leveraged access to the CMS Integrated Data Repository (IDR), a high-volume data warehouse integrating Medicare claims, beneficiary and provider data sources, along with advanced SAS software, and the fraud-prevention system to develop predictive analytics capabilities and other analytic tools for use by OIG staff to help ensure

the integrity of HHS programs, especially Medicare.[9] According to one panelist, this centralized approach has enabled CMS and HHS OIG to pursue a more strategic and coordinated set of antifraud policies, avoid incurring additional costs, and improve collaboration.

- *Data.* Panelists encouraged entities to inventory current data sources and consider how the data can be used to address program objectives. Figure 2 shows examples of data sources a state insurance fraud department used to detect improper payments, as presented by one of our panelists.

---

[9]According to the company that produces it, SAS is an integrated system of software solutions that enables one to perform data entry, retrieval, and management; report writing and graphics design; statistical and mathematical analysis; business forecasting and decision support; operations research and project management; and applications development.

Figure 2: Examples of Data Sources Used to Detect Improper Insurance Payments
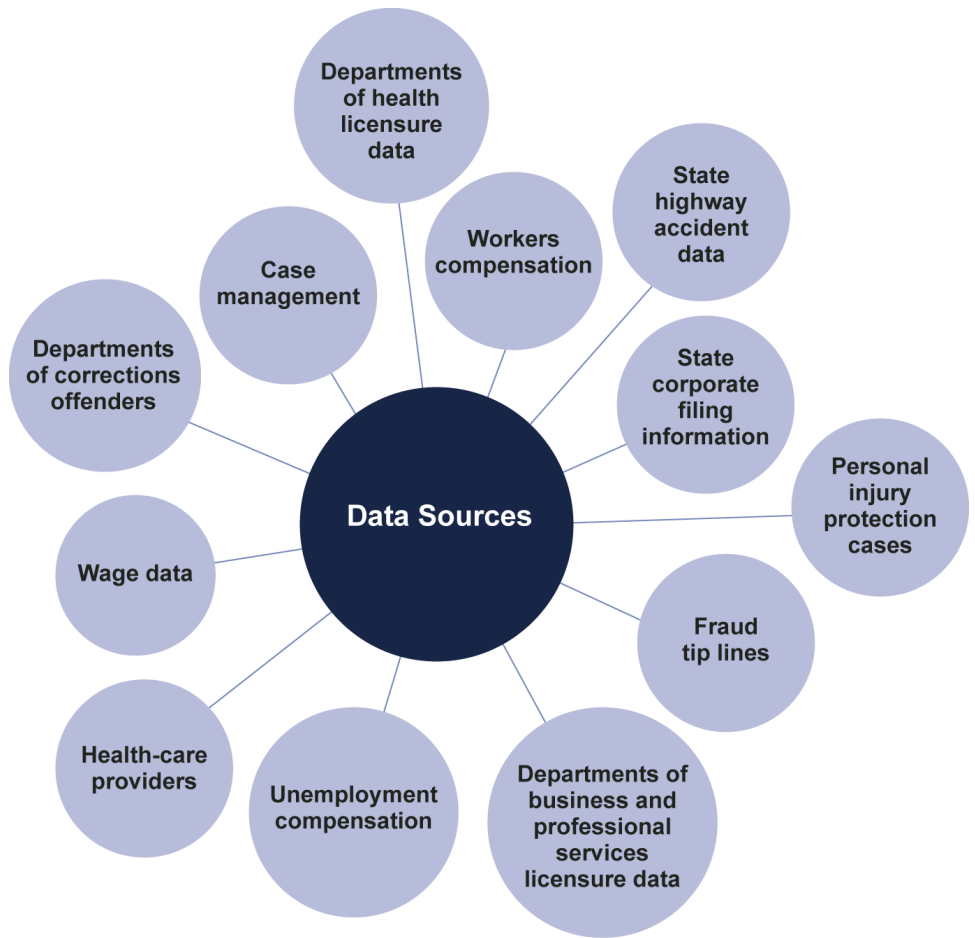
Source: Analysis360.com | GAO-17-339SP

Once an entity has identified its currently available data sources, panelists suggested the entity should assess the quality of its data to identify data-collection and data-processing errors, among other potential problems. For example, entities that rely on many individuals to collect and record data may encounter inconsistencies and errors.[10] Specifically, one panelist mentioned that the Department of Housing and Urban Development (HUD) has tens of thousands of

---

[10]For example, in a follow-up after the forum, one panelist explained that these inconsistencies and errors may range from name variation due to typographical errors, nicknames, and alternative spelling to incomplete addresses and abbreviations.

outside stakeholders who input data about housing authorities and communities that receive Community Development Block Grant funding into various systems. Given the number of individuals responsible for entering data and the potential for human error in data entry, HUD must determine the reliability and accuracy of the data prior to conducting data analytics. To assist with conducting a data-reliability assessment, GAO previously published a guide on assessing the reliability of computer-based data.[11] After an entity determines the reliability of its data, it should acknowledge any limitations when reporting results, according to panelists.

Finally, while reviewing currently available data and assessing the reliability of data, an entity may discover suspicious patterns or unexpected results that may help further define an entity's objectives or lead to a new area of focus, as the sidebar shows.

- *Staff.* Panelists suggested that entities identify current staff who possess experience with, enthusiasm for, and interest in data analytics. To identify these individuals, one panelist recommended looking throughout the entity to find individuals from many backgrounds including auditors, evaluators, investigators, and counsel, to leverage a diverse set of perspectives. Another panelist recommended finding individuals within an entity who are currently working with data to measure performance and consider transferring this experience to a data-analytics program.

DOD's Excessive Shipping Costs

In September 2006, the Department of Defense (DOD) uncovered excessive shipping charges after looking at transaction-level data. DOD identified a purchase of two $0.19 flat washers with a suspicious shipping charge of nearly $1 million. Given that this charge did not seem reasonable, DOD looked into the vendor bill and noticed that the exact shipping charge of $998,798 matched the invoice number, which was likely an attempt to claim plausible deniability if identified, according to a panelist. Upon further investigation, DOD found that this vendor and nine similar vendors had been submitting similar bills for nearly 10 years, amounting to over $20 million in improper payments. The vendor was convicted of conspiracy, sentenced to prison time, and ordered to pay restitution.
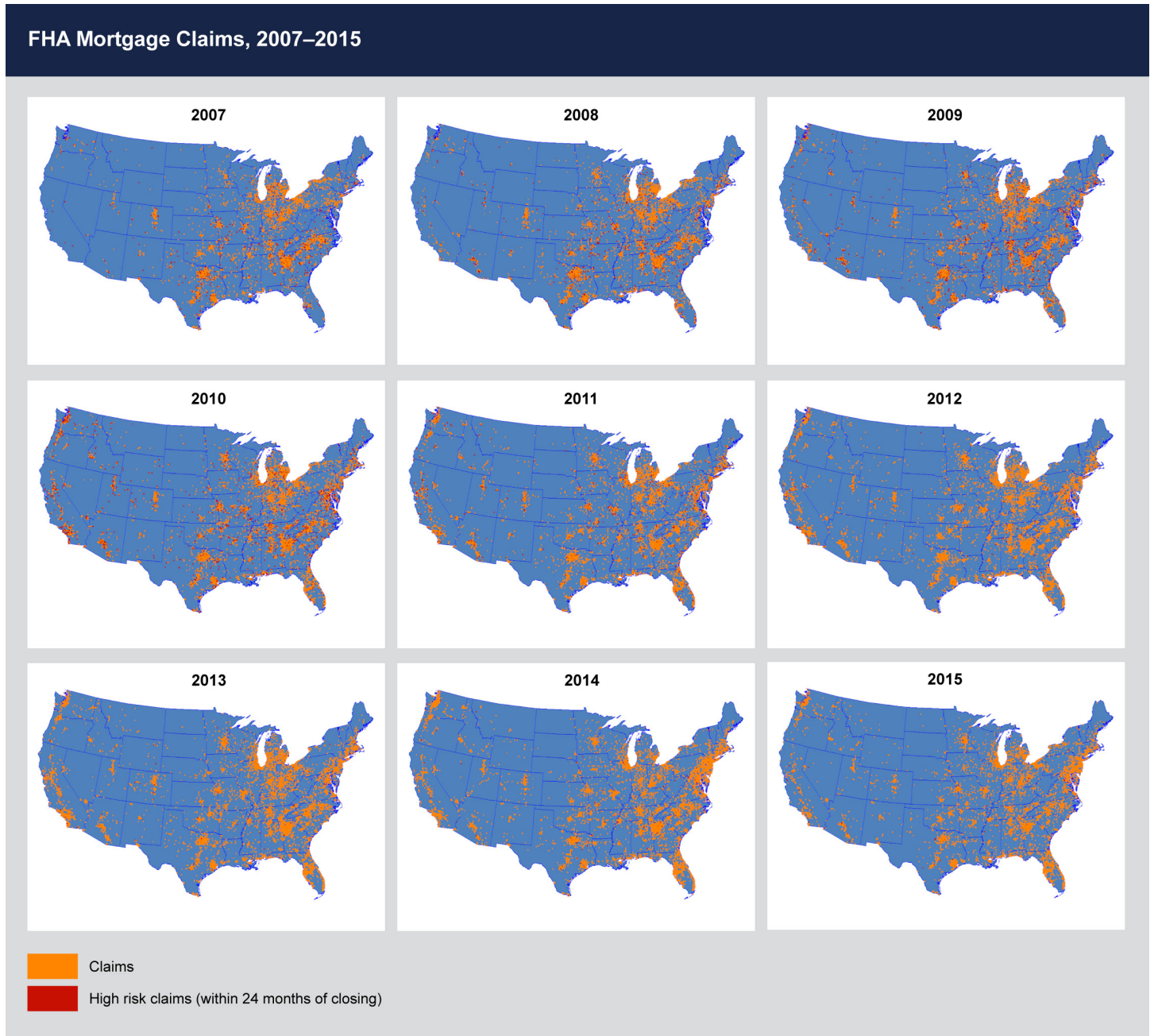

Source: GAO. | GAO-17-339SP

---

[11]See GAO, *Assessing the Reliability of Computer-Processed Data*, GAO-09-680G (Washington, D.C.: July 2009).

Demonstrate the Value of Analytics to Obtain Entity-Wide Support

To help obtain support from both leadership and staff for data analytics, an entity should demonstrate the value of analytics in addressing an entity's objectives, according to panelists. Data visualization is one technique that an entity might use to demonstrate the return on investment of an analytics program. By developing visualizations, an entity can use pictures to present an analytics model, which may help people without experience in data analytics understand an issue. For example, figure 3 depicts a data visualization that the HUD OIG used to illustrate the concentration of Federal Housing Administration (FHA) mortgage-insurance claims over time. The panelist who presented this data visualization explained that the dramatic increase in claims indicated, and may have warned of, the recent financial crisis that was initially triggered by defaults on U.S. mortgage loans. The panelist noted that this type of visualization assisted with briefing high-level government officials and encouraged the HUD OIG to develop an analytical methodology to track default rates as early indicators of trends in the housing-mortgage market.

Figure 3: Visualization of Federal Housing Administration (FHA) Mortgage Insurance Claims

FHA Mortgage Claims, 2007–2015

Claims

High risk claims (within 24 months of closing)

Source: Department of Housing and Urban Development, Office of Inspector General. | GAO-17-339SP

Note: Each map represents FHA mortgage claim data from the first quarter of the year.

In another example of how HUD OIG used visualizations when presenting its work to high-level government officials, figure 4 illustrates the Hurricane Sandy storm surge and the affected HUD public-housing assets. By combining satellite imagery from the Federal Emergency Management Agency (FEMA) with HUD housing data, HUD OIG and high-level government officials could gain a more complete understanding of the number of public-housing assets that were affected by the hurricane. Additionally, these data initiated an effort to determine how many public-housing authorities nationwide are located in flood zones and, of those, how many were insured. According to this panelist, using visual aids when presenting analysis to the highest levels of an entity can help demonstrate the value of an analytics program.

Figure 4: Hurricane Sandy Storm Surge and Department of Housing and Urban Development (HUD) Public-Housing Assets around Atlantic City, New Jersey
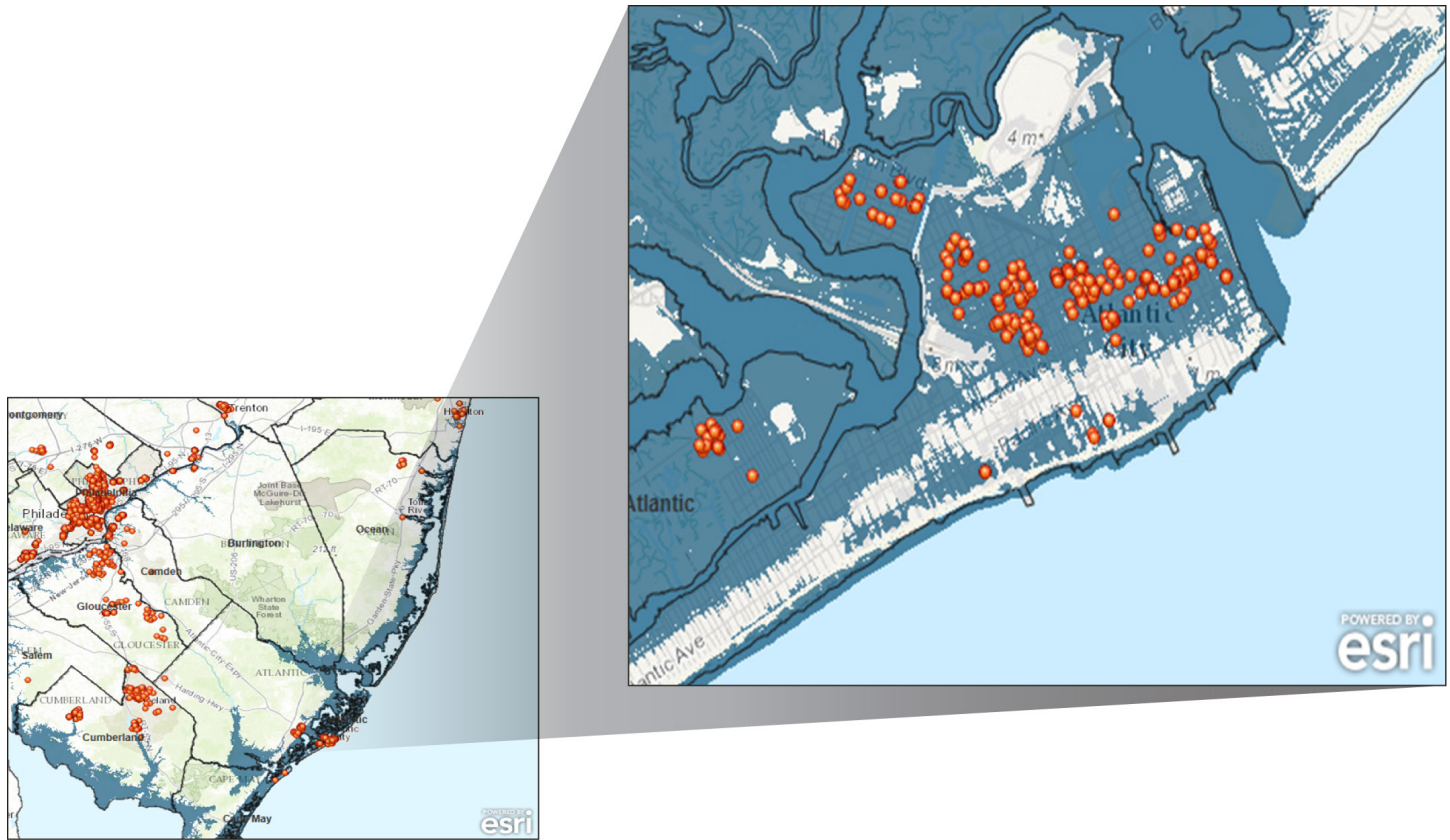
■ Storm surge

● Housing and Urban Development (HUD) public-housing assets

Additionally, panelists stated that it might benefit entities looking to acquire new analytics software to demonstrate the value or return on investment of an analytics tool before buying it. To help determine the usefulness of a tool, panelists recommended that entities work with vendors to understand how software works, whether the tool will meet the objectives of the entity, and how it may interface with an entity's existing network before purchasing the software. For example, one panelist discussed a successful proof-of-concept project for modeling health-care claims data. After looking at many potential vendors, the panelist's entity conducted a 6-month pilot to establish return on investment. After 6

months, the pilot was determined to be a success, and the entity received management approval and purchased the tool. Finally, one participant suggested that an entity ask its data analysts their opinions on analytics software. This participant suggested that data analysts who are excited about a tool may be more motivated to learn and use the software.

Considerations for Refining Analytics Programs

Forum panelists identified considerations for entities refining analytics programs, including seeking to continuously improve analytics to address improper payments and employing individuals who understand both analytics and business operations.

Continuously Improve Analytics

Panelists highlighted the importance of continuously improving analytics programs to keep their capabilities current by revisiting and reassessing existing analytics tools and techniques. To measure improvement, panelists suggested that entities establish a baseline against which progress might be measured. Panelists stressed that entities should improve analytics models through an iterative process, building on existing knowledge, historical patterns, and feedback from customers. For example, an established data-analytics program may revisit its data-collection methods, its process of entering data, or the time frame required to deliver results. The sidebar on the next page describes an example in which a state Supplemental Nutrition Assistance Program (SNAP) agency worked with various parties to refine its methodology for using data to inform its investigations of potentially illegal benefit trafficking.[12]

---

[12]The goal of SNAP, formerly known as the federal Food Stamp Program, is to help low-income individuals and households obtain a more nutritious diet and help alleviate their hunger. It does so by supplementing their income with benefits to purchase allowable food items. SNAP trafficking refers to the misuse of program benefits to obtain nonfood items.

Wisconsin's Supplemental Nutrition Assistance Program (SNAP) Recipient Fraud Prevention and Detection Project

In 2016, the state of Wisconsin improved its ability to address potential fraud within SNAP by refining its analytics methodology. The state previously analyzed electronic transaction data from SNAP purchases and combined these data with documentation from the U.S. Department of Agriculture's Food and Nutrition Service (FNS) to charge recipients with trafficking.

Some SNAP recipients charged with trafficking appealed to administrative law judges within Wisconsin. During these appeals, judges overturned the state's actions, citing that the state failed to meet the evidentiary standard required to prove the individual recipients engaged in trafficking. Due to this ruling, Wisconsin temporarily stopped charging suspicious recipients.

To overcome this legal hurdle, the state refined its analytics methodology and devised a predictive model intended to target suspicious recipients more effectively. In the new model, when FNS disqualifies a store, the state examines recipients who shopped at that store and targets only individuals with suspicious activity. Additionally, the state worked with administrative law judges to understand the evidence necessary to meet a clear and convincing evidentiary standard. This collaboration helped inform the investigators how to change their investigation process to ensure they provide the necessary evidence to substantiate a case. Furthermore, the data team developed stronger models that incorporated the new investigation process and evidentiary standards into the analysis in order to identify potential fraud that most closely fit those models.

Ultimately, Wisconsin received approval from the state Office of Judicial Counsel to present cases using the new analytics methodology. This change has helped the state to remove a 2-year trafficking case roadblock and legally disqualify SNAP recipients found guilty of trafficking.

Source: GAO analysis of SNAP data. | GAO-17-339SP

Panelists also encouraged entities to refresh and add data sources when they revisit current data-analytics tools and techniques, including from nontraditional or external sources. For example, according to one panelist in a follow-up interview following the forum, every 6 months CMS uses information from various nontraditional or external sources to update its analytical models. Such sources include information obtained from meetings with investigative managers from the OIG and federal and state law-enforcement agencies, as well as information from private companies or other federal agencies that have models for data analytics to detect improper payments in the health-care industry. According to the panelist, some of these ideas led directly to the development of better methods for detecting improper payments.

As another example, one panelist mentioned a recent academic project where students used a "web-scraping" methodology to search Twitter for keywords relating to fraud.[13] Similarly, another panelist mentioned that the U.S. Department of Agriculture's (USDA) Food and Nutrition Service (FNS) is currently experimenting with text-mining approaches for social-media data. For example, the panelist explained that FNS is looking at various social-media monitoring tools to address potential fraud, including to detect potential evidence about individuals' efforts to sell or purchase benefits or commit eligibility fraud on websites for classified advertisements. However, the panelist noted that there have been significant challenges in developing an appropriate methodology that yields a full return on investment.

---

[13]Web scraping may also be referred to as data extraction or web data mining.

Employ Individuals Who Understand Both Analytics and Business Operations

Cost of Compound Drugs in Workers' Compensation

The United States Postal Service (USPS) Office of Inspector General (OIG) developed analytics models to identify medical providers with suspiciously large increases in workers' compensation billing over a short amount of time. Specifically, analysts looked at Postal Service data on workers' compensation costs and used heuristics, a type of practical, exploratory analysis, to identify a $68.6 million increase in certain workers' compensation drug costs from 2014 to 2015.

Although analysts initially anticipated doctors and hospitals would have the highest increases in billing, after analyzing the data the analysts found that pharmacies had the highest increases. Further, according to a panelist, the experienced analysts recognized that the high billing increase seemed unusual and developed multiple iterations of the analytics model to dig deeper into the data. Ultimately, the analysts identified a specific type of drug—compound drugs—as the main contributor to these large billing increases. Compound drugs are created when licensed pharmacists combine, mix, or alter ingredients of drugs to tailor them to individual patients. While the number of paid workers' compensation medical cases stayed relatively stable from 2011 to 2015, compound-drug costs increased 1,924 percent, growing from about $5 million to about $100 million during the same period.

The Postal Service OIG is continuing to investigate the increase in billing for compound drugs and has coordinated with the Department of Labor on implementing best practices for controlling compound-drug costs.

Source: GAO analysis of USPS OIG data; USPS OIG (image). | GAO-17-339SP

Panelists noted that entities need knowledgeable subject-matter experts who understand both analytics and the business of the entity to successfully combine data and technology with public policy. For example, experienced data analysts may have the knowledge to explore unexpected findings and identify unknown improper payments and fraud schemes in an entity's business operations, as described in the sidebar.

When an entity does not have individual staff members with knowledge of both the subject matter of its business operations and technical data-analytics skills, one panelist suggested pairing individuals with each of these skillsets. For example, in a follow-up discussion after the forum, one panelist explained that FNS employs both technical experts who perform data analytics and subject-matter experts in business operations. The panelist emphasized that, to successfully use data to identify improper payments, data analysts need to work with subject-matter experts to understand (1) any underlying assumptions of the data analysis and (2) limitations to the data analysis that arise from an understanding of business operations, such as how the data are used in the application process or where the data are stored during various parts of the business process.

During the forum, panelists also noted that new academic programs train data analysts in both public policy and data analytics. Further, these analysts may have experience in new analytics programming languages. For example, one panelist noted that these academic programs may teach and use open-source analytics tools rather than the more established proprietary analytics or legacy tools that many government entities may be familiar with. This panelist encouraged entities to educate themselves on new analytics programming languages to help find new talent that is proficient in both public policy and data analytics.

Opportunities for Collaboration

During the forum, panelists identified many opportunities for collaboration among analytics programs in the government, private sector, public–private partnerships, and academia. Collaboration may allow entities to share resources and knowledge to address the government-wide challenge of improper payments. In doing so, panelists identified considerations such as (1) defining objectives to guide collaboration, (2) building trust, (3) overcoming common barriers to collaboration, and (4) utilizing existing partnerships to guide future collaborations.

Defined Objectives May Guide Collaboration

Any collaboration should start by identifying the defined objectives and desired outcomes, according to panelists. For example, entities might focus on reducing fraud, increasing targeted audits, or promoting greater collaboration between investigations and audits. Collaboration may even have a high-level goal, such as seeking to improve a program. According to one panelist, identifying defined objectives may serve as a starting point to inspire collaboration because potential partners can more clearly recognize the potential benefits to their mission or objectives.

Building Trust Is Essential for Collaboration

Source: GAO (data); PhotoDisc (image). | GAO-17-339SP

Panelists stated that entities should work to build trust with each other and demonstrate a willingness to collaborate. Although building trust may take time, panelists suggested finding common ground as a starting point. For an example of building trust, see the sidebar.

Building trust is particularly important when an entity seeks to share data among entities, according to panelists. Initially, entities may be reluctant to share data due to concerns about data security, but entities should look for opportunities to share nonsensitive information as a way to build trust. For example, according to one panelist, the Central Intelligence Agency (CIA) expressed interest in collaborating with the Institute for Advanced Analytics at North Carolina State University, but had privacy and security concerns about sharing its data outside of the agency. To test this potential collaboration, the agency initially provided a completely anonymized, fabricated data set to the institute to evaluate its security controls and the benefits of the analytics service. After the students worked with the data set and solved the proposed problem, the CIA determined it was satisfied with the solution. It subsequently shared additional, more sensitive data, with the appropriate safeguards, and established an ongoing collaboration with the institute, according to this panelist.
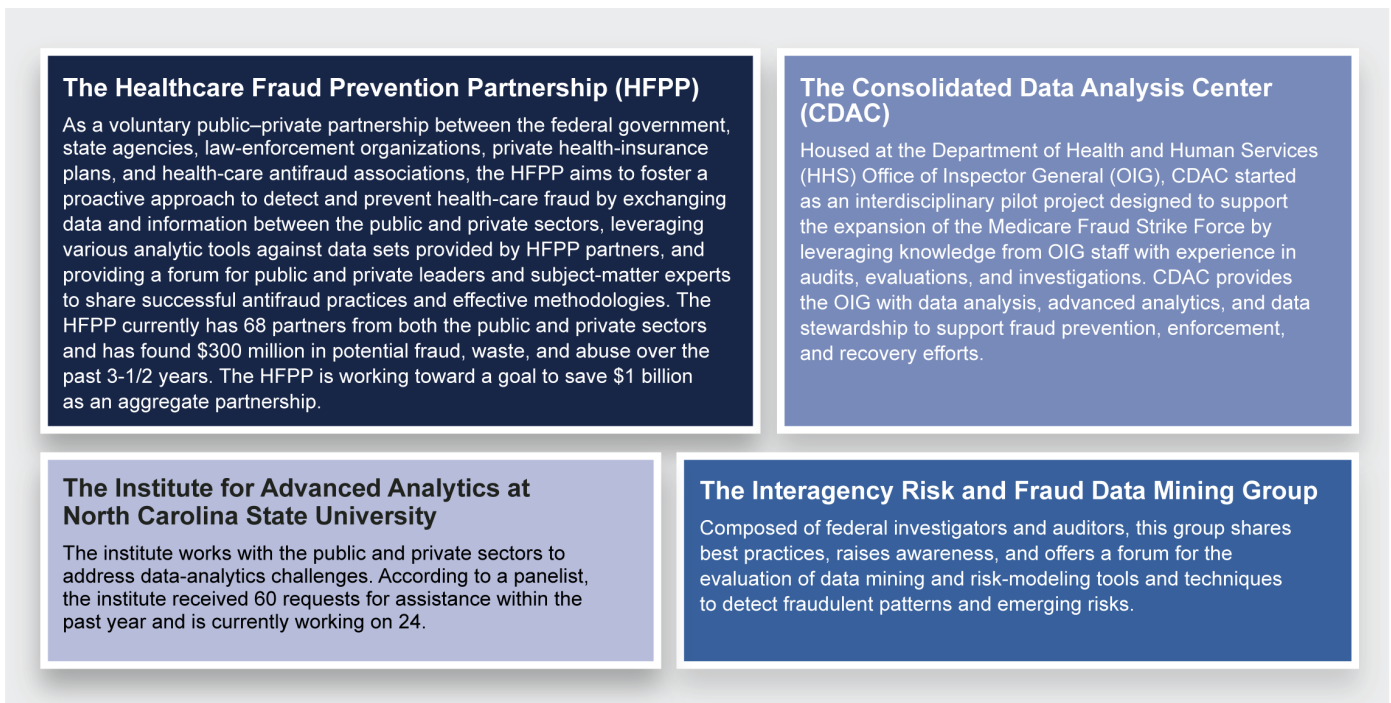
Overcoming Common Barriers to Collaboration

Panelists stated that many commonly cited barriers to collaboration can be overcome, including perceived legal and data-sharing barriers. To address legal and data-sharing challenges, panelists suggested asking probing questions to understand where the boundaries to collaboration may lie. For example, when an entity cites perceived legal barriers that may prevent data sharing, panelists suggested asking for specific information about the legal requirements to better understand the problem. Panelists noted that by asking questions and clearly communicating with potential partners, an entity may be able to overcome data-sharing barriers. For example, one panelist stated that ineffective communication about a potential partner's needs and concerns is one of the biggest barriers in collaboration and is something people are sometimes reluctant to confront. Although not all barriers can be overcome, panelists recommended that entities consider whether sharing a portion of a data set may provide enough information to meet business objectives without violating legal requirements.

Existing Partnerships May Provide Guidance on Future Collaboration

Forum participants suggested that conferences, forums, professional organizations, and some existing partnerships and resources may serve as examples for future collaboration. See figure 5 for additional examples provided by the panelists.

Figure 5: Examples of Partnerships and Resources Provided by the Panelists

The Healthcare Fraud Prevention Partnership (HFPP)

As a voluntary public–private partnership between the federal government, state agencies, law-enforcement organizations, private health-insurance plans, and health-care antifraud associations, the HFPP aims to foster a proactive approach to detect and prevent health-care fraud by exchanging data and information between the public and private sectors, leveraging various analytic tools against data sets provided by HFPP partners, and providing a forum for public and private leaders and subject-matter experts to share successful antifraud practices and effective methodologies. The HFPP currently has 68 partners from both the public and private sectors and has found $300 million in potential fraud, waste, and abuse over the past 3-1/2 years. The HFPP is working toward a goal to save $1 billion as an aggregate partnership.

The Consolidated Data Analysis Center (CDAC)

Housed at the Department of Health and Human Services (HHS) Office of Inspector General (OIG), CDAC started as an interdisciplinary pilot project designed to support the expansion of the Medicare Fraud Strike Force by leveraging knowledge from OIG staff with experience in audits, evaluations, and investigations. CDAC provides the OIG with data analysis, advanced analytics, and data stewardship to support fraud prevention, enforcement, and recovery efforts.

The Institute for Advanced Analytics at North Carolina State University

The institute works with the public and private sectors to address data-analytics challenges. According to a panelist, the institute received 60 requests for assistance within the past year and is currently working on 24.

The Interagency Risk and Fraud Data Mining Group

Composed of federal investigators and auditors, this group shares best practices, raises awareness, and offers a forum for the evaluation of data mining and risk-modeling tools and techniques to detect fraudulent patterns and emerging risks.

Source: GAO. | GAO-17-339SP

Although collaboration can be helpful, one panelist cautioned that entities should not assume that an analytics program can be directly transferred from one entity to another. This panelist encouraged entities to collaborate and learn from each other while acknowledging that every analytics program has a unique business problem to address. Consequently, there is no one-size-fits-all solution to using data analytics to address improper payments.

Appendix I: Objectives, Scope, and Methodology

This report summarizes discussions from the Comptroller General Forum on Data Analytics to Address Fraud and Improper Payments that was held on September 21, 2016, at GAO headquarters. For purposes of the forum, an improper payment is defined by statute as any payment that should not have been made or that was made in an incorrect amount (including overpayments and underpayments) under legally applicable requirements. Among other things, an improper payment includes payment to an ineligible recipient, payment for an ineligible good or service, and any duplicate payment. An improper payment also includes payments for goods or services not received (except for such payments where authorized by law) and any payment that does not account for applicable discounts.[1] In addition, the Office of Management and Budget's (OMB) guidance instructs agencies to report as improper payments any payments for which insufficient or no documentation was found.

The forum focused on three main topics: (1) key components and suggested practices for establishing and refining data-analytics programs; (2) case-study examples of analytics tools, techniques, and methods, including data visualization, dashboards, geographic mapping, social-network analysis, link analysis, text mining, and predictive analysis, among others; and (3) opportunities for collaboration and future partnerships.

To inform our understanding of data analytics to address improper payments, including fraud, we conducted a literature review for studies that discussed

- fraud and improper-payment prevention, and

- detection and government performance management.

To identify existing studies from peer-reviewed journals, we searched various databases, such as ProQuest, SCOPUS, and SciSearch, for

---

[1]While all federal payments made fraudulently are considered improper payments, not all improper payments are fraud. Fraud involves obtaining something of value through willful misrepresentation (see GAO, *Standards for Internal Control in the Federal Government*, GAO-14-704G [Washington, D.C.: Sept. 10, 2014], 8.02, for discussion on types of fraud). Whether an act is in fact fraud is a determination to be made through the judicial or other adjudicative system and is beyond management's professional responsibility for assessing risk. For the purposes of the forum and this report, unless noted otherwise, we generally use the term "fraud" to include potential fraud for which a determination has not been made through the judicial or other adjudicative system.

studies published from 2011 through February 2016. From these sources, we identified 14 studies that were relevant to our research objectives. Our criteria for choosing these studies included their relevance to the public sector and their use of quantitative methods related to fraud or improper payments. We performed these searches in February 2016. Additionally, we reviewed literature recommended by GAO colleagues on data analytics to address improper payments, including fraud. We used the background information gathered during this literature review to develop questions to interview data-analysis experts.

To identify potential forum panelists and gather additional information on data analytics to address improper payments, including fraud, we interviewed data-analysis experts. To identify these experts, we developed criteria including relevant work experience in data analytics, improper payments, or fraud; recent presentations or publications related to data analytics to address improper payments, including fraud; and recommendations from GAO colleagues, among other considerations. To select panelists for the forum, we met with 43 individuals from 21 organizations across government, the private sector, public–private partnerships, and academia. During these interviews, we gathered additional information on data-analysis topics and about the experience of the experts to inform our forum panelist selection. From these interviews, we selected 13 forum panelists from the government, private sector, public–private partnerships, and academia based on their demonstrated ability to provide examples of their work and discuss topics related to using data analytics to address improper payments, including fraud.[2]

Additionally, we used the data-analysis expert interviews to develop the focus of the forum to include a panel on initial analysis development and continuous refinement of analysis programs, as well as a panel on opportunities for continued collaboration. To discuss real-world applications of data analytics, we organized two case-study panels highlighting specific tools and techniques being used to address improper payments, including fraud. Since each panelist could speak on a variety of forum topics, we allocated each panelist to a topic based on his or her area of expertise. Three panelists had demonstrated experience in establishing and refining an analytics program, and four panelists were able to highlight examples of analytics collaboration for the detection and

---

[2]Although we originally selected 13 panelists for the forum, 1 panelist was unable to attend at the last minute. A total of 12 panelists spoke at the forum.

prevention of improper payments.[3] On the basis of our interviews, we selected six panelists to present case-study examples highlighting tools and techniques being used to address improper payments. Prior to the forum, we prepared background materials that framed the issues to be discussed during each session.

Following the forum, a team of GAO analysts reviewed the transcript of the panel discussions to identify key considerations identified by panelists for this report. We developed three categories for various challenges identified by the participants:

- considerations related to establishing an analytics program,
- considerations related to refining an analytics program, and
- opportunities for collaboration.

We identified the considerations and opportunities discussed by the panelists and organized them into the relevant category. Additionally, we contacted some panelists after the forum to request additional detail or clarification on the forum discussion on an as-needed basis.

This report summarizes the discussion by forum panelists and guests, highlighting key ideas that emerged during the discussion sessions. However, this report is not intended to represent an exhaustive list of all ideas and themes that emerged during the forum. Further, the information presented in this summary does not necessarily represent the views of all participants or the views of their organizations, including GAO. We structured the forum so that all panelists could openly comment on issues, although not all panelists commented on all topics. Throughout the report, we refer to "panelists" in order to indicate views expressed by more than one panelist.

We provided an opportunity for the panelists to comment on a draft of this report and incorporated their comments into the final report, as appropriate.

We conducted our work from December 2015 to March 2017 as part of our ongoing body of work on improper payments, including fraud, and in accordance with all sections of GAO's Quality Assurance Framework that

---

[3]The panelist who was unable to attend the forum was originally scheduled to discuss opportunities for collaboration.

are relevant to our objectives. The framework requires that we plan and perform the engagement to obtain sufficient, appropriate evidence to meet our stated objectives and discuss any limitations in our work. We believe that the information and data obtained, and the analysis conducted, provide a reasonable basis for any findings and conclusions in this product.

Appendix II: Forum Panelists

Dr. Brett M. Baker
Deputy Inspector General for Audit
U.S. Department of Defense Inspector General

Johan Bos-Beijer
Director, Analytics & Services
Office of the Commissioner
Technology Transformation Service
General Services Administration

Kelly Jin
Policy Advisor to the U.S. Chief Technology Officer
The White House Office of Science and Technology Policy

Bryan Jones
Director and Program Manager
Elder Research, Inc.

Edwin Krafsur
Manager, Data Analytics
United States Postal Service Office of Inspector General

Aric LaBarr
Assistant Professor of Analytics
Institute for Advanced Analytics at North Carolina State University

Alanna M. Lavelle
Program Director / Senior Advisor
MITRE Corporation

The Honorable David A. Montoya
Inspector General
U.S. Department of Housing and Urban Development Office of Inspector General

Steven G. Shandy
Supervisory Program Manager
U.S. Department of Health and Human Services Office of Inspector General

Ronald Ward
Director, Program Accountability and Administration Division
Food and Nutrition Service, Supplemental Nutrition Assistance Program (SNAP)
U.S. Department of Agriculture

Raymond Wedgeworth
Director, Data Analytics and Systems Group
Center for Program Integrity
Centers for Medicare & Medicaid Services

Christopher R. Westphal
Subject-Matter Expert
analysis365.com

COMPTROLLER GENERAL FORUM ON

Data Analytics to Address Fraud and Improper Payments

Wednesday, September 21, 2016 | Agenda

Time	Session	Speakers
9:00am - 9:15am	**Opening Session: Welcome**	**The Honorable Gene L. Dodaro**, Comptroller General of the United States **Vijay D'Souza,** Director, Center for Enhanced Analytics, Government Accountability Office
9:15am - 10:30am	**Initial Analytics Development and Continuous Refinement** What are key components of establishing and maintaining data-analytics tools and techniques to detect and prevent fraud and improper payments?	**MODERATED PANELIST DISCUSSION:** **Steve Shandy,** Office of the Inspector General, U.S. Department of Health and Human Services **Bryan Jones,** Elder Research, Inc. **Kelly Jin,** The White House Office of Science and Technology Policy
10:30am - 10:45am	**15-MINUTE BREAK**	
10:45am - 12:00noon	**Methods and Techniques: Case Studies I** Case studies will highlight specific tools and techniques being used to detect and prevent fraud and improper payments, including data visualization, dashboards, and geographic mapping software.	**CASE-STUDY PRESENTATIONS:** **Chris Westphal,** analysis365.com **David Montoya,** Office of the Inspector General, U.S. Department of Housing and Urban Development **Brett Baker,** Office of the Inspector General, U.S. Department of Defense **DISCUSSION WITH MODERATOR**
12:00noon - 1:15pm	**BREAK FOR LUNCH**	
1:15pm - 2:30pm	**Methods and Techniques: Case Studies II** Case studies will highlight specific tools and techniques being used to detect and prevent fraud and improper payments, including text mining and the future use of predictive techniques.	**CASE-STUDY PRESENTATIONS:** **Ron Ward,** Food and Nutrition Service, U.S. Department of Agriculture **Edwin Krafsur,** Office of the Inspector General, U.S. Postal Service **Ray Wedgeworth,** Centers for Medicare & Medicaid Services **DISCUSSION WITH MODERATOR**
2:30pm - 2:45pm	**15-MINUTE BREAK**	
2:45pm - 4:00pm	**Opportunities for Continued Collaboration** Is there a role for data analytics as a shared service? Could partnerships (private, CIGIE, joint agency operations, academic) help government leverage resources and knowledge?	**MODERATED PANELIST DISCUSSION:** **Alanna Lavelle,** MITRE Corporation **Caryl Brzymialkiewicz,** Office of the Inspector General, U.S. Health and Human Services **Johan Bos-Bejier,** Office of the Commissioner, Technology Transformation Service, General Services Administration **Aric LaBarr,** Institute for Advanced Analytics at NC State University
4:00pm - 4:15pm	**Wrap-Up**	

Appendix IV: Forum Issue Brief

Data Analytics to Address Fraud and Improper Payments

In fiscal year 2015, federal agencies estimated making almost $137 billion in improper payments, a number that has increased considerably in recent years.[1] Improper payments, including fraud, are a significant and pervasive government-wide issue and have presented a continuing challenge to the federal government.[2] Reducing improper payments is critical to safeguarding federal funds, and could help achieve cost savings and improve the government's fiscal position.

GAO's previous work has shown that implementing preventive and detective controls, including data-analytics tools and techniques, is one strategy to address fraud and improper payments. At the Comptroller General Forum on Data Analytics to Address Fraud and Improper Payments, GAO will convene data-analysis experts to outline specific analytics techniques. The experts will offer case-study examples and discuss the challenges associated with developing and maintaining a data-analytics program. Given GAO's abilities to work on crosscutting issues that affect dozens of agencies in multiple branches of government, the agency is well-positioned to host this event and produce a dialogue that stimulates new partnerships and identifies solutions to leverage data, skills, and resources across the public sector.

To advance the intergovernmental dialogue on implementing data analytics to prevent and detect fraud and improper payments, this forum will highlight three main topics in panel discussions:

Panel 1:	Key components and suggested practices for establishing and refining data-analytics programs
Panels 2 and 3:	Case-study examples of analytics tools, techniques, and methods including: data visualization, dashboards, geographic mapping, social-network analysis, link analysis, text mining, and predictive analysis, among others
Panel 4:	Opportunities for collaboration and future partnerships

Additionally, this forum will seek to address the following questions, among others:

How have agencies used a risk-based approach to data analytics, considering the benefits and costs of investing in specific data-analytics tools and techniques and focusing data analytics on the program's highest risks?	What tools/techniques may be considered leading practices for fraud and improper payment prevention and detection?	What type of partnerships (i.e., private, CIGIE, joint agency operations, academic) may help government leverage resources and knowledge?	Is there a role for data analytics as a shared service? Is there a risk of overlap and duplication as each agency develops independent data-analytics programs?

---

[1] An improper payment is defined as any payment that should not have been made or that was made in an incorrect amount (including overpayments and underpayments) under statutory, contractual, administrative, or other legally applicable requirements.

[2] It is important to note that while all fraud involving a federal payment is considered an improper payment, not all improper payments are fraud.

GAO U.S. GOVERNMENT ACCOUNTABILITY OFFICE

Appendix V: GAO Contacts and Staff Acknowledgments

GAO Contacts

Vijay D'Souza, (202) 512-2700 or dsouzav@gao.gov

Seto Bagdoyan, (202) 512-6722 or bagdoyans@gao.gov

Staff Acknowledgments

In addition to the contacts above, Johana Ayers, Bruce Cain, Timothy Carr, Beryl Davis, Colin Fallon, James M. Healy, Venugopal Katta, Lauren Kirkpatrick, Barbara Lewis, Phillip McIntyre, Linda Miller, Jonathon Oldmixon, David Plocher, Kiran Sreepada, Walter Vance, and Elizabeth Wood made important contributions to organizing the forum and producing this report.

Related GAO Products

Improper Payments: Strategy and Additional Actions Needed to Help Ensure Agencies Use the Do Not Pay Working System as Intended. GAO-17-15. Washington, D.C.: October 14, 2016.

Highlights of a Forum: Data and Analytics Innovation: Emerging Opportunities and Challenges. GAO-16-659SP. Washington, D.C.: September 20, 2016.

A Framework for Managing Fraud Risks in Federal Programs. GAO-15-593SP. Washington, D.C.: July 2015.

Highlights of a Forum: Data Analytics for Oversight and Law Enforcement. GAO-13-680SP. Washington, D.C.: July 15, 2013.

Assessing the Reliability of Computer-Processed Data. GAO-09-680G. Washington, D.C.: July 2009.