

# DATA ANALYTICS FOR IDENTIFYING IMPROPER PAYMENTS AND FRAUD

## **Beryl H. Davis**

CGFM, CPA, CIA, CGAP, CGMA, CCSA,  
Director, Financial Management and Assurance,  
U.S. Government Accountability Office (U.S. GAO)



## **Christopher R. Westphal**

Vice President, SME Analytics  
FORCEPOINT



# OVERVIEW OF IMPROPER PAYMENTS



# WHAT IS AN IMPROPER PAYMENT?

**An improper payment is any payment that should not have been made or was made in an incorrect amount - including overpayments and underpayments.**

- *Duplicate payments*
- *Payments to ineligible recipients*
- *Incorrect amounts paid*
- *Payments with insufficient or no documentation*

# RELEVANT LEGISLATION

- **Improper Payments Information Act of 2002 (IPIA)**
- **Improper Payments Elimination and Recovery Act of 2010 (IPERA)**
- **Improper Payments Elimination and Recovery Improvement Act of 2012 (IPERIA)**
- **Federal Improper Payments Coordination Act of 2015**
- **Government Performance Results Modernization Act (GPRA Modernization Act) of 2010**
- **Digital Accountability and Transparency Act (DATA Act) of 2014**

# FY 2015: IMPROPER PAYMENT ESTIMATES ARE ON THE RISE

## Government-wide Estimated Improper Payments (in billions)



- In FY 2015, **22 agencies** reported improper payment estimates for **121 programs**

# FY 2015 IMPROPER PAYMENT ESTIMATES: LARGEST 3 PROGRAMS

Program	Agency	Reported Improper Payment Estimates	
		Dollars (in billions)	Error rate (percent of outlays)
Medicare Fee-for-Service (Parts A and B)	HHS	\$43.3	12.1%
Medicaid	HHS	\$29.1	9.8%
Earned Income Tax Credit (EITC)	Treasury	\$15.6	23.8%

- Three programs were responsible for **\$88 billion** of the **\$136.7 billion** FY 2015 government-wide improper payment estimate
- With outlays for major programs, such as Medicare and Medicaid, expected to increase over the next few years, it is critical that actions are taken to reduce improper payments

# IMPROPER PAYMENT REPORTING OBSTACLES

**GAO's report on the Fiscal Year 2015 Financial Report of the United States Government continued to report a material weakness in internal control related to improper payments**

**The federal government is unable to determine the full extent to which improper payments occur and reasonably assure that appropriate actions are taken to reduce them**

# CFO ACT AGENCIES THAT DID NOT REPORT IMPROPER PAYMENTS IN FY 2015

- **CFO Act Agencies Not Reporting Estimates (No Programs Reported as Susceptible to Significant Improper Payments):**
  - U.S. Agency for International Development
  - Department of Energy
  - Department of State
  - National Science Foundation
  - Nuclear Regulatory Commission
- **CFO Act Agencies Reporting Only Hurricane Sandy-related Estimates:**
  - Department of Commerce
  - Department of the Interior
  - Department of Justice
  - NASA

# HOW AUDITORS CAN USE DATA



# HOW AUDITORS CAN USE DATA

Auditors can use data to:

- Inform risk assessment and planning process,
- Understand an agency's environment through the performance of substantive analytics,
- Improve audit quality, and
- Detect fraud.
  - Statistics
  - Machine Learning
  - Entity Analytics





# COMMON SENSE PATTERNS

- If it does not make sense...
- It is not normal...
- It seems unusual...
- Too coincidental...
- Too frequent...

There is no **right** answer

There is no **wrong** answer

Interpretation in context





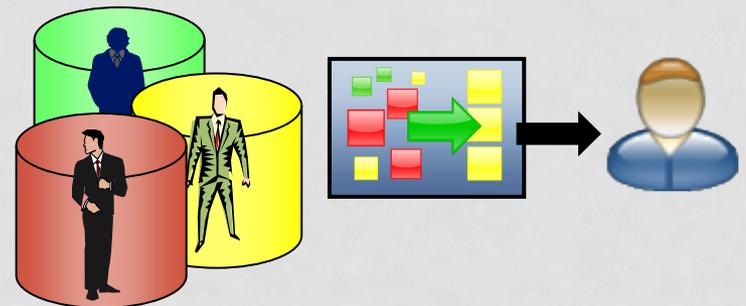
# CONTEXT IS IMPORTANT...





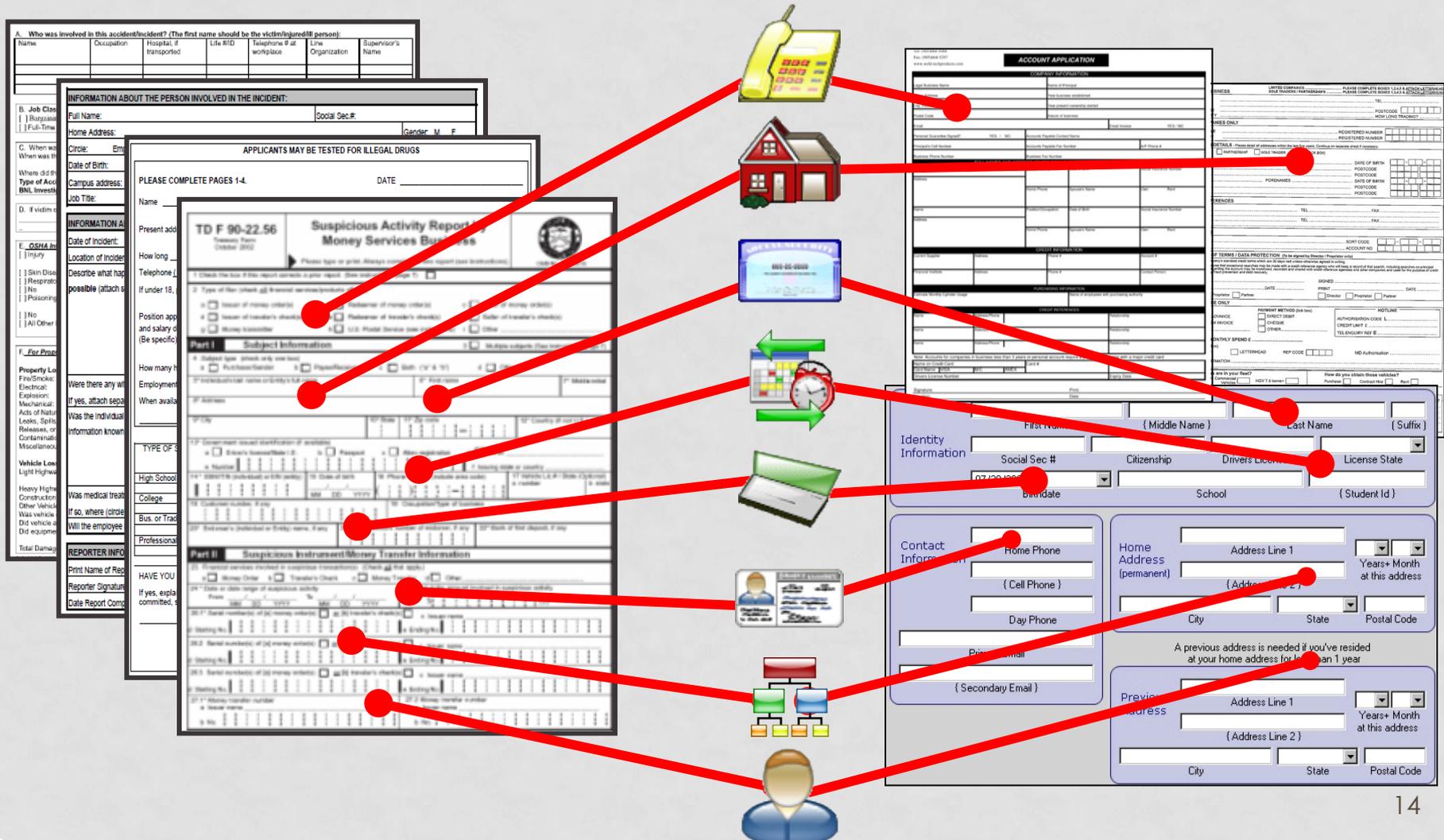
# SIMILAR DATA, DIFFERENT CONTENT

- SNAP – food stamps
- Supplemental Nutrition
- Welfare Benefits
- Energy Assistance
- Homestead Exemptions
- Unemployment Assistance
- Medicaid/Medicare
- Social Security Benefits
- Retirement/Pensions
- Workers Compensation
- Unregistered Businesses
- Underreported Income
- Prescription Drugs
- Tax Evasion
- ...etc. etc. etc. etc...





# EXTRACT DATA FROM...



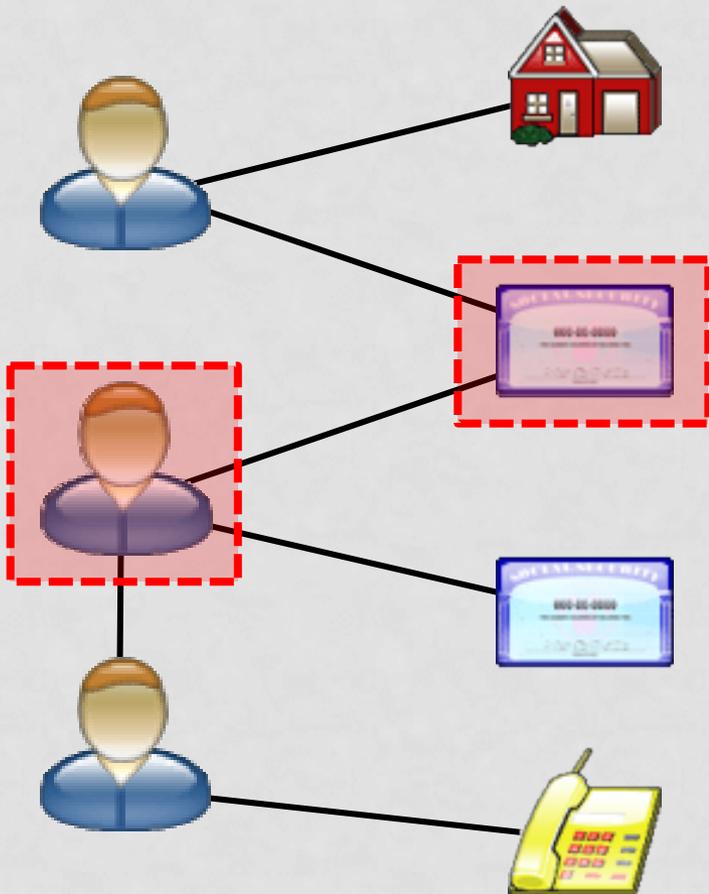


# CONNECT THE DOTS





# DISCOVER & VISUALIZE PATTERNS



- Certain “entities” should never be shared (e.g., SSNs)
- Data prone to typos and misspellings
- Possible misrepresentation and/or falsifying data on forms
- Appearance of avoidance by varying information

# CHALLENGES TO DATA ANALYSIS

- **Main challenges:**

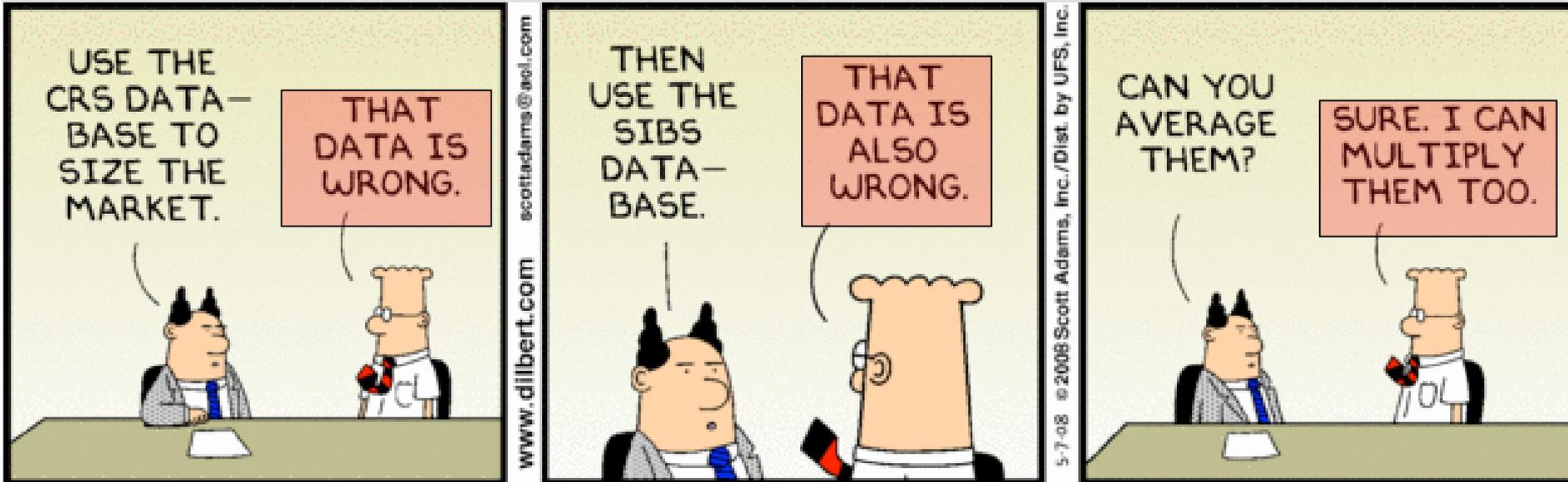
- reliability of data,
- accessibility of data,
- timeliness and age of data,
- diversity of data formats (extensive amount of data cleaning needed for types of data),
- different reporting requirements for separate data systems (i.e. Medicaid system vs. state systems),
- server speed and
- storage limitations





# DATA QUALITY AFFECTS RESULTS

<http://dilbert.com/strips/comic/2008-05-07/>



- Poor Collection Standards
- Information Often Not Verified
- Incomplete Data
- Errors in the Data
- Misrepresentation of Data
- Inconsistencies in the Data
  - Typos/Errors
  - Misspellings
  - Transpositions
  - Transliterations/Cultural
  - Qualifiers/Prefixes/Suffixes

# EXAMPLES AND USES OF DATA ANALYTICS



# GAO REPORT: MEDICAID FRAUD CONTROLS(GAO-15-313)

## What GAO found:

- **Thousands of Medicaid beneficiaries and hundreds of providers involved in potential improper or fraudulent payments during FY 2011**

**Table 1: Potential Improper-Payment Indicators Related to Medicaid Beneficiary Identity for Four Selected States during Fiscal Year 2011**

<b>Potential improper-payment indicator</b>	<b>Approximate number receiving benefits</b>	<b>Estimate of total Medicaid benefits paid (dollars in millions)</b>
Beneficiaries concurrently receiving benefits paid by two or more states	8,600	\$18.3
Deceased beneficiaries	200	9.6
Incarcerated beneficiaries	3,600	4.2

Source: GAO analysis of data provided by the Centers for Medicare & Medicaid Services (CMS); Arizona, Florida, Michigan, and New Jersey state Medicaid programs; the Social Security Administration (SSA); and Arizona, Florida, Michigan, and New Jersey state departments of corrections. | GAO-15-313

# GAO REPORT: MEDICAID FRAUD CONTROLS (GAO-15-313)

## How data analytics were used:

- **Analyzed data from**
  - Medicaid claims,
  - State Medicaid rolls, and
  - Exclusionary datasets.
- **Looked for indicators of improper or potentially fraudulent payments, including an analysis of specific exclusionary factors;**
  - beneficiaries were deceased,
  - incarcerated, or
  - currently receiving benefits from two or more states).



# MEDICARE CLAIM PIVOTS & RATIOS

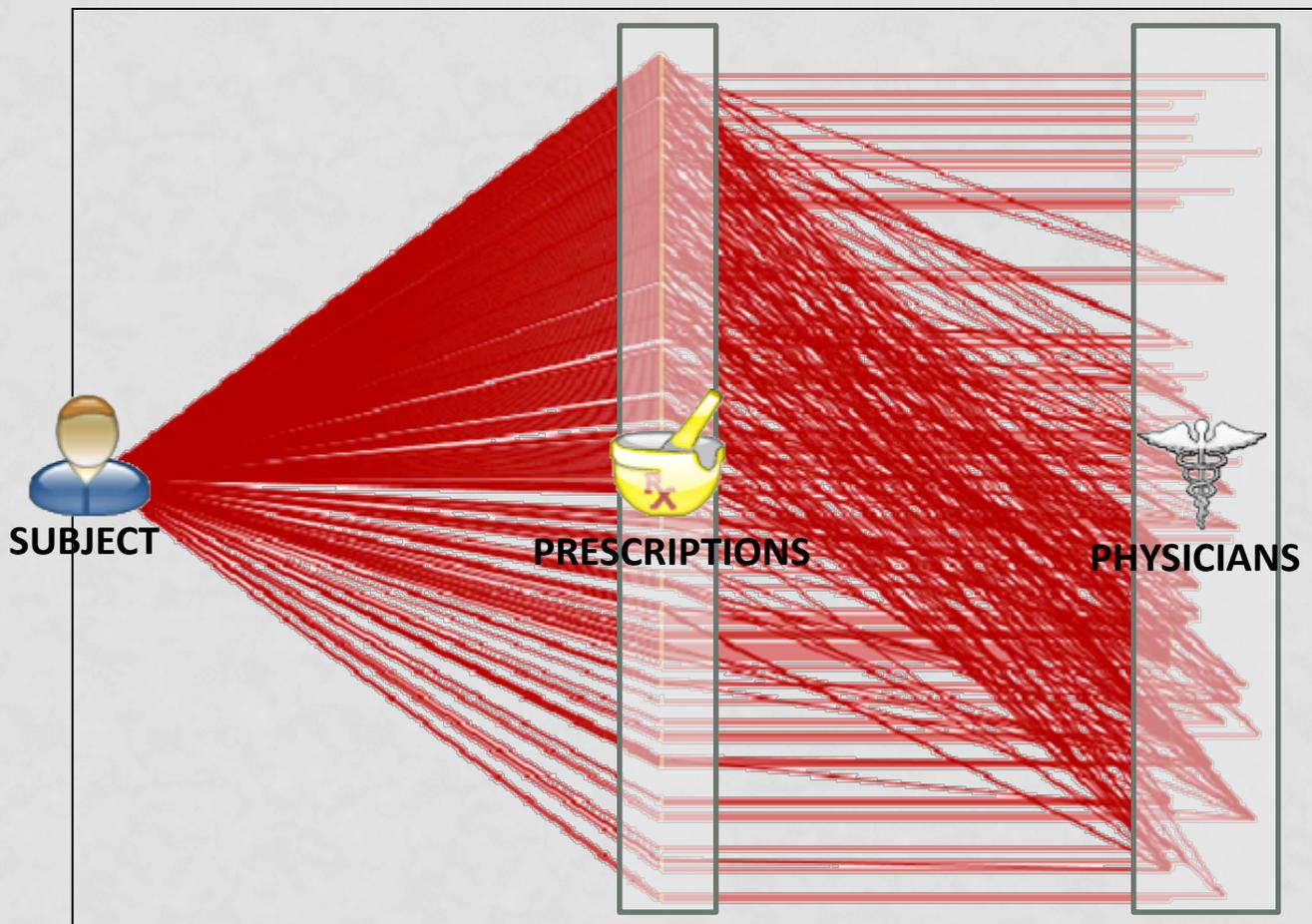
# CLAIMS	# MEMBER	ICD DESCRIPTION	PROVIDER	RATIO
825	75	LABORATORY EXAMINATION	PRV0042350	11:1
819	70	LABORATORY EXAMINATION	PRV0073294	12:1
680	62	LABORATORY EXAMINATION	PRV0067438	11:1
658	61	LABORATORY EXAMINATION	PRV0062230	11:1
624	58	LABORATORY EXAMINATION	PRV1754997	11:1
576	9	LABORATORY EXAMINATION	PRV3273933	<b>64:1</b>
534	71	LABORATORY EXAMINATION	PRV4120896	8:1
485	54	LABORATORY EXAMINATION	PRV0309492	9:1
459	51	LABORATORY EXAMINATION	PRV1251458	9:1
448	50	LABORATORY EXAMINATION	PRV2305050	9:1





# PRESCRIPTION DRUG (PDMP) FRAUD

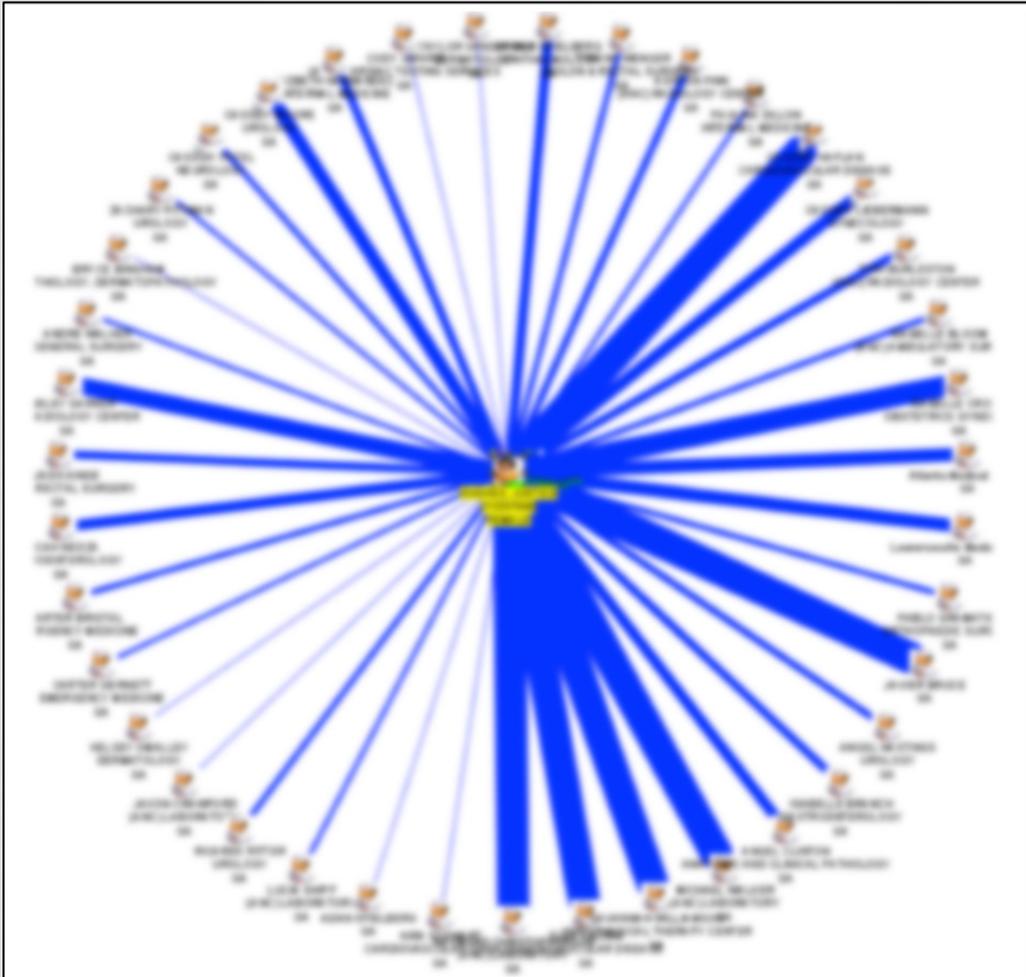
This view shows the relationship between a SUBJECT (left), to PRESCRIPTION (middle), and the related PHYSICIAN (right)



Oxycodone



# EXCESSIVE MEDICAL CLAIMS BY A SINGLE MEMBER



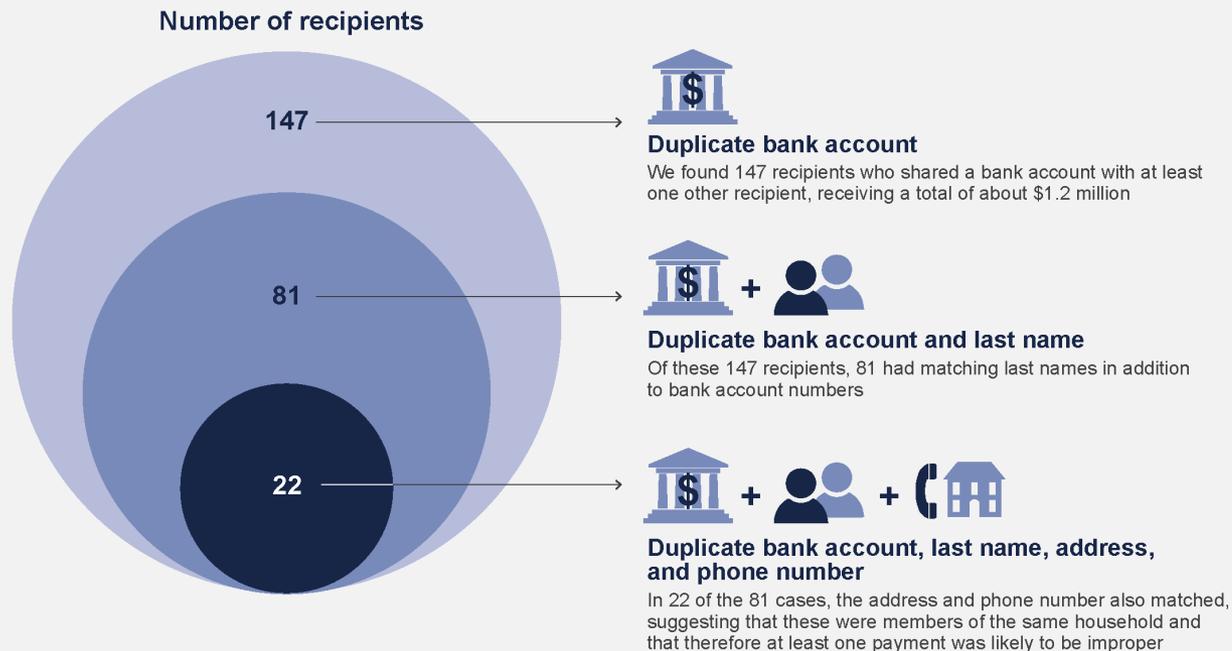
*[intentionally blurred]*

- Large utilization of provider services by a single member (~40)
- Link thickness represents repeated/frequent claims

# GAO REPORT: FEMA HURRICANE SANDY (GAO-15-15)

## What GAO Found:

- \$39 million of FEMA payments for Hurricane Sandy relief were at risk of being improper or fraudulent



- Over 1,000 addresses were identified at which multiple recipients received more than the maximum \$31,900 in benefits

# GAO REPORT: FEMA HURRICANE SANDY (GAO-15-15)

## How data analytics were used:

- **Data matching identified:**
  - duplicate social security numbers,
  - duplicate addresses,
  - duplicate bank account numbers, and other identifiers.
- Identified potentially ineligible recipients, such as those identified by other systems as having invalid SSNs.



# HURRICANE KATRINA\*

August 29, 2005  
1,833 people lost their lives  
1M people displaced  
\$120B spent post-Katrina  
\$75B for emergency relief  
\$8B distributed for...  
    needs assistance  
    housing assistance



Estimates \$1.4B lost to fraud

- *Using FEMA hotels while receiving housing/rental allowances*
- *People making up SSNs (they did not exist / match names)*
- *Listing addresses of cemeteries / vacant lots / PO Boxes*
- *Registrants for incarcerated state/federal prisoners*



# LOUISIANA DEPARTMENT OF LABOR

First Claim Submitted  
450 miles from NOLA  
Used own name

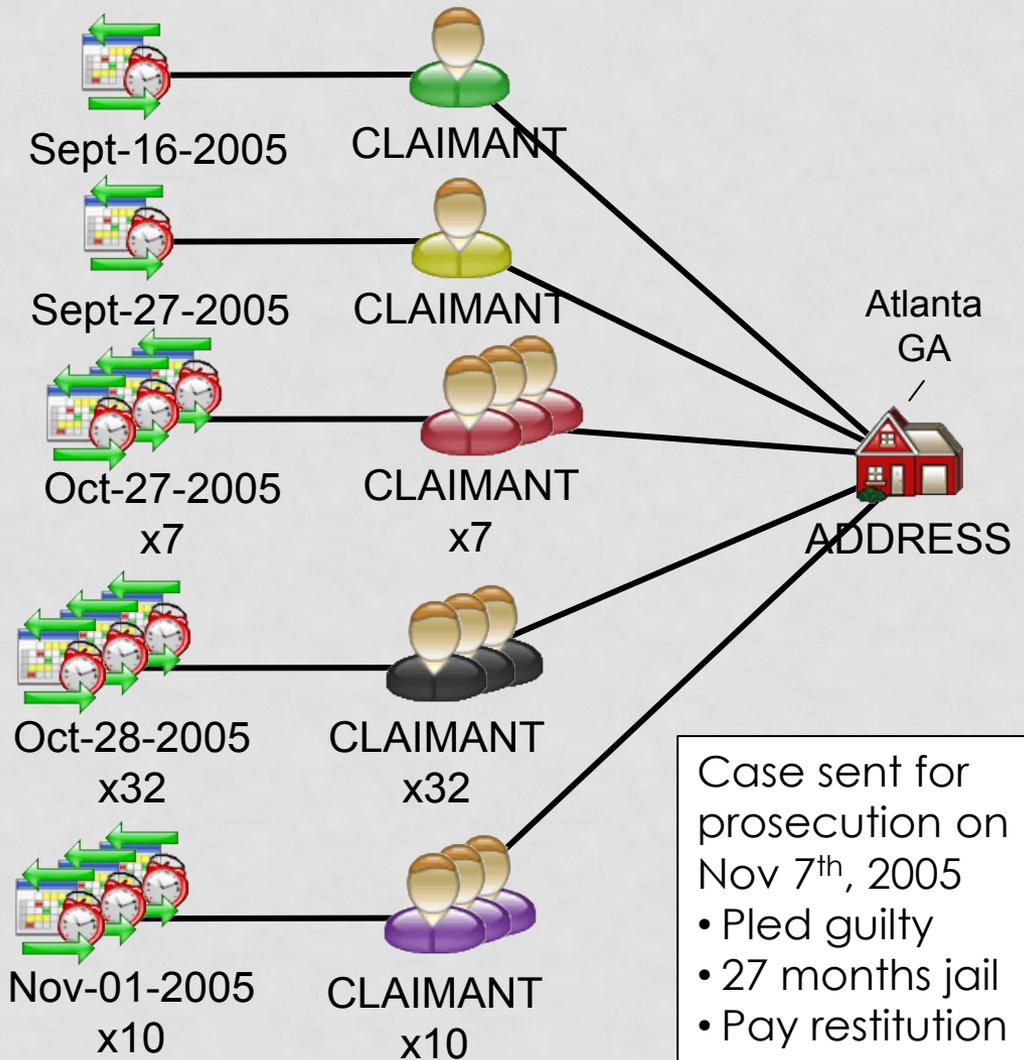
Second Claim Made  
Different name  
Same address

7 More Claims  
Same last name  
Different first name  
Same DOB  
Similar SSNs  
Same address

32 Additional Claims

10 More Claims

Total = 51 Debit Cards



Case sent for prosecution on Nov 7<sup>th</sup>, 2005

- Pled guilty
- 27 months jail
- Pay restitution



# GAO REPORT: SNAP FRAUD (GAO-14-641)

## **What GAO Found:**

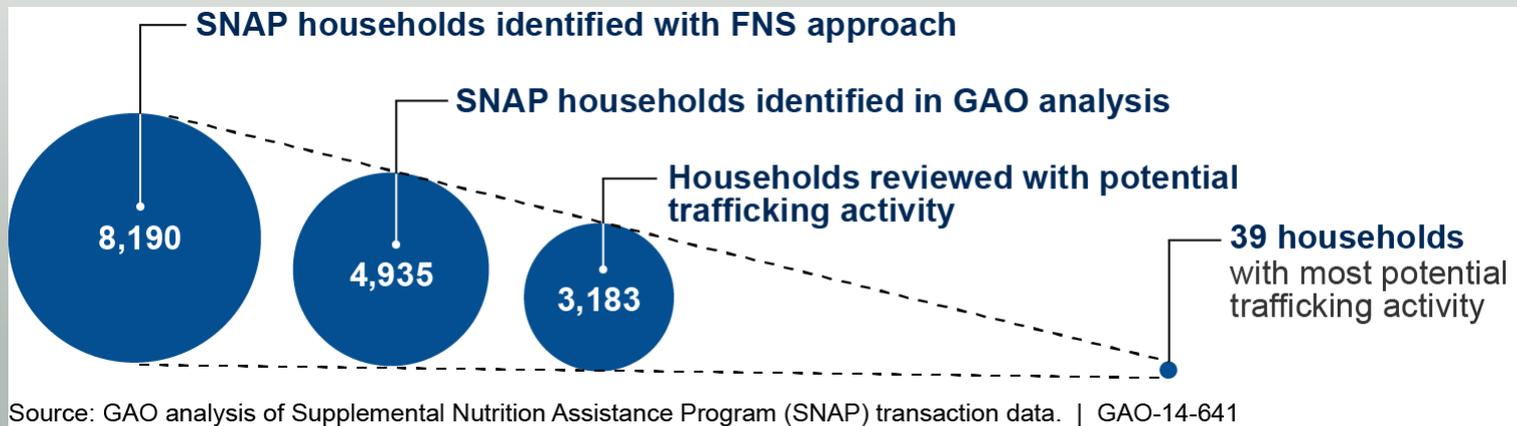
- **States used detection tools recommended by the USDA's Food and Nutrition Service (FNS) to combat potential fraud in the Supplemental Nutrition Assistance Program (SNAP)(formerly known as Food Stamps).**
- **However, the tools used by states reported limited success.**

# GAO REPORT: SNAP FRAUD (GAO-14-641)

## How data analytics were used:

- Replacement card data and suspicious transaction indicators were used to narrow down the number of households at high risk for potential trafficking in one state from over 8,000 to 39 households.

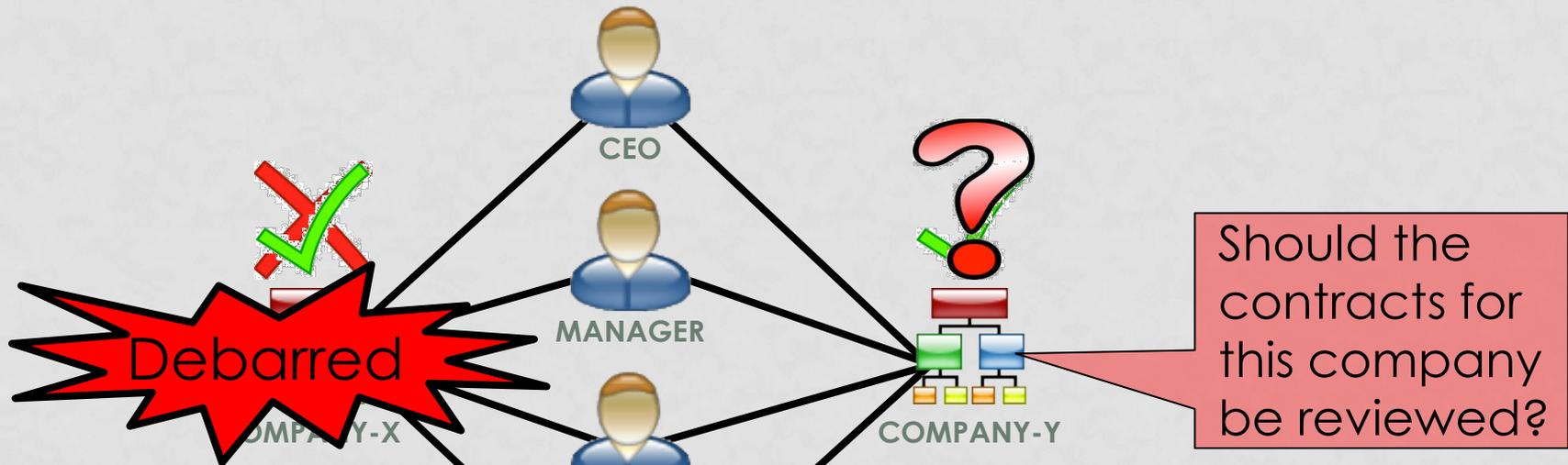
Figure 1: Using Replacement Cards to Target Trafficking in Michigan, Fiscal Year 2012



- Because states had reported limited resources to combat fraud in recent years, data analytics could assist the states in developing a more targeted approach to better manage SNAP risk.



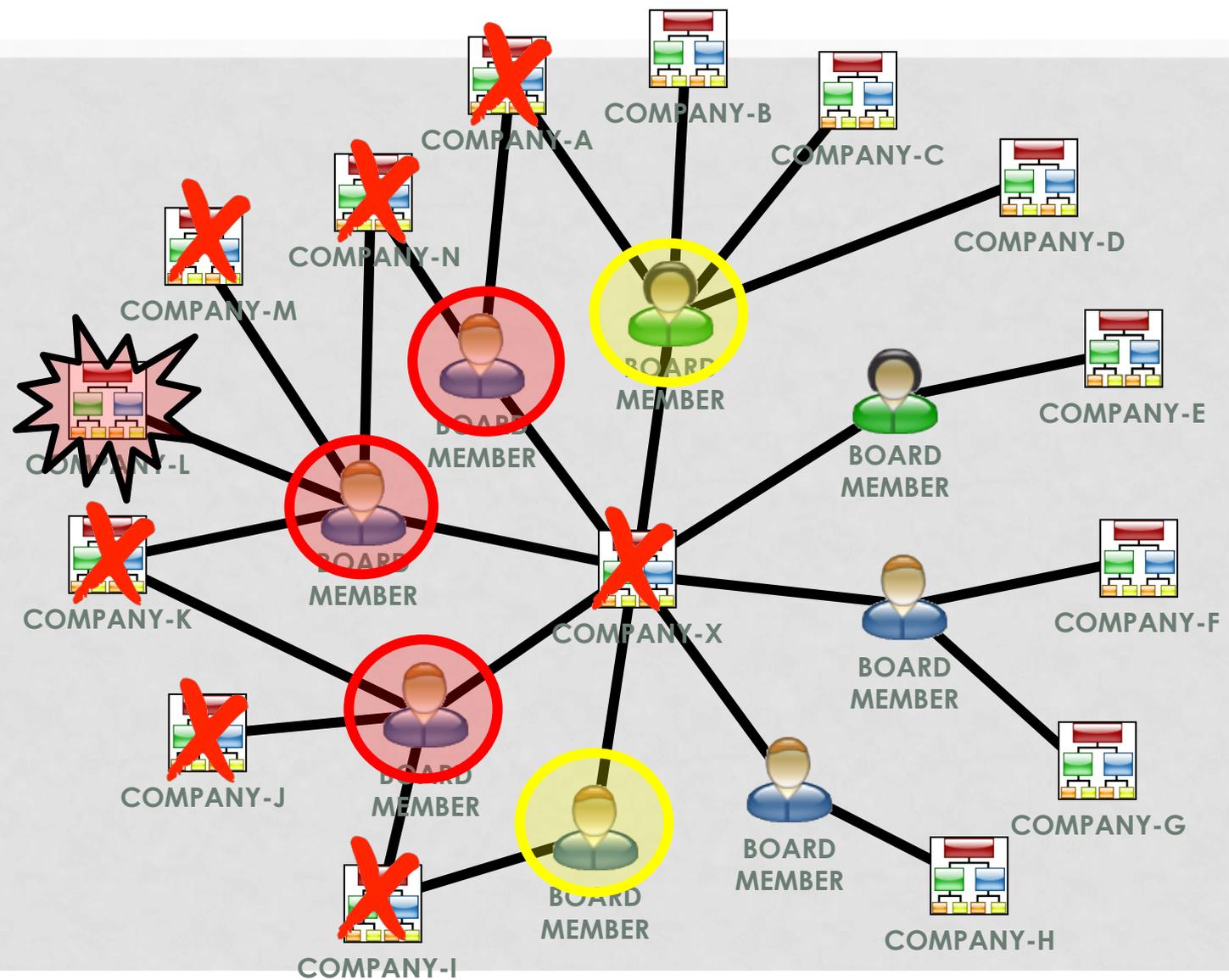
# DEBARRED PROVIDER - EXPAND



- Corporations/Contractors
- Health Care Providers
- Retail Programs
- Money Agents
- Merchants



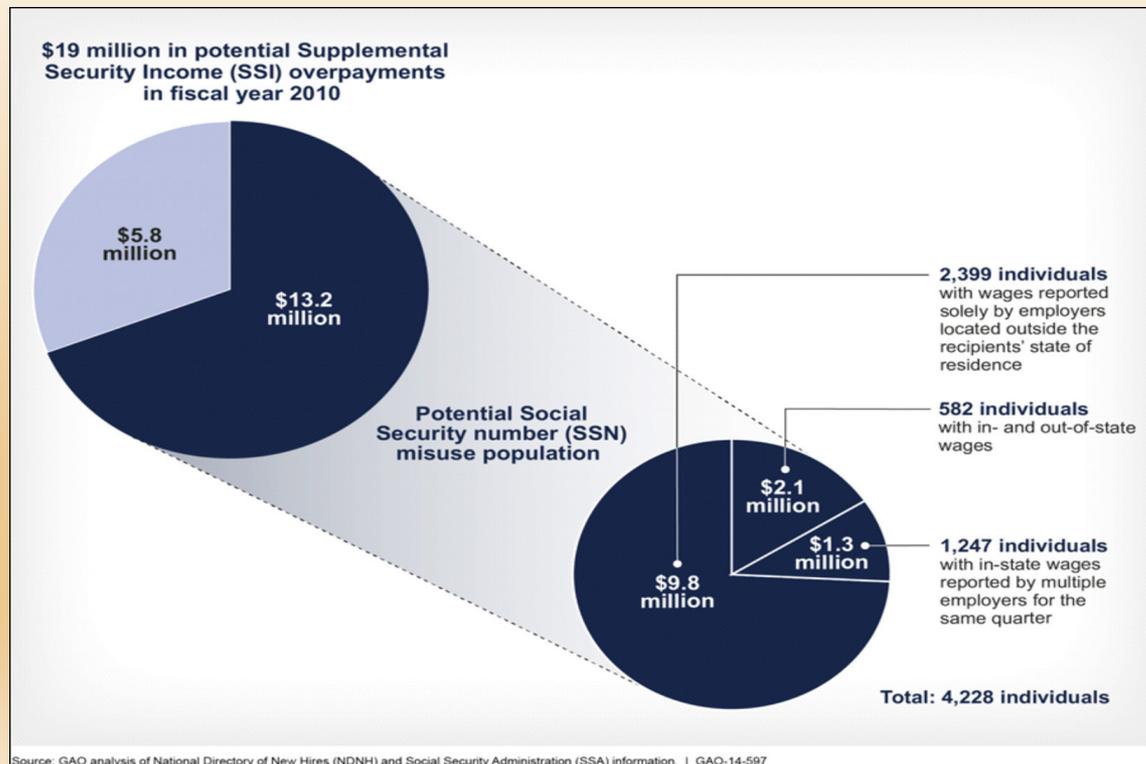
# INVESTIGATE CONNECTIONS – HOW FAR?



# GAO REPORT: SUPPLEMENTAL SECURITY INCOME (GAO-14-597)

## What GAO Found:

- Wages reported for SSI recipients showed indications of possible misuse of Social Security numbers, further indicating that SSA made \$19 million in potential overpayments to over 10,000 recipients.

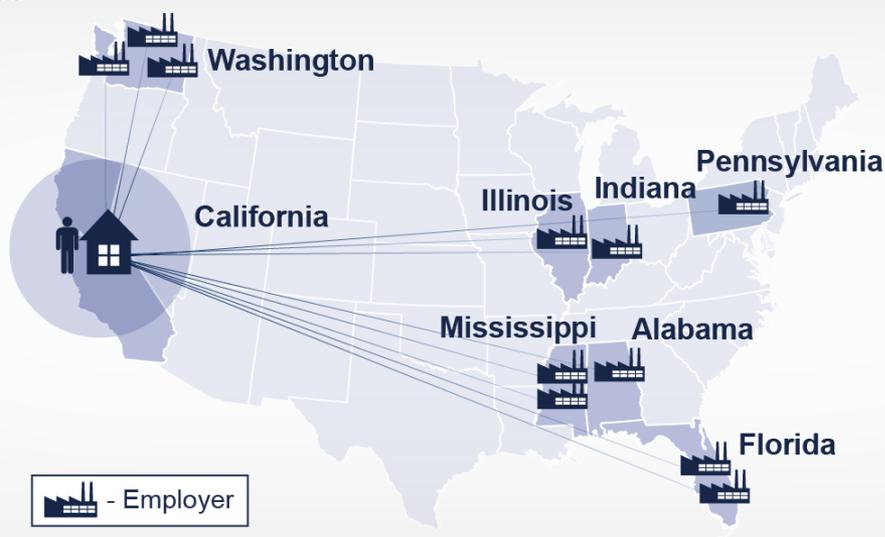


# GAO REPORT: SUPPLEMENTAL SECURITY INCOME (GAO-14-597)

## How data analytics were used:

- Potential overpayments were identified by comparing data from SSA and the National Directory of New Hires (NDNH).
- Data matching of employer identification numbers identified wages for over 2,300 SSI recipients that were reported solely by employers outside the recipient's state of residence

### Supplemental Security Income (SSI) Recipient with Reported Wages in Seven Different States



- One individual had wages reported from 11 different employers in 7 other states during the same quarter in 2010.



# META DATA FOR SOCIAL SECURITY #

454 million numbers assigned

Not Valid: (000-##-####, ###-00-####, ###-##-0000)

Will not start with a 666



Social Security Validation Check

Effective June 25, 2011, assigned randomly



Social Security Death Master Index (Date/Name)

Appx 95 Million



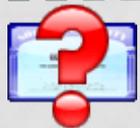
Individual Taxpayer Identification Number (ITIN)

9XX-70-XXXX

9XX-99-XXXX



Social Security Calculations (Age Proximity / Location)



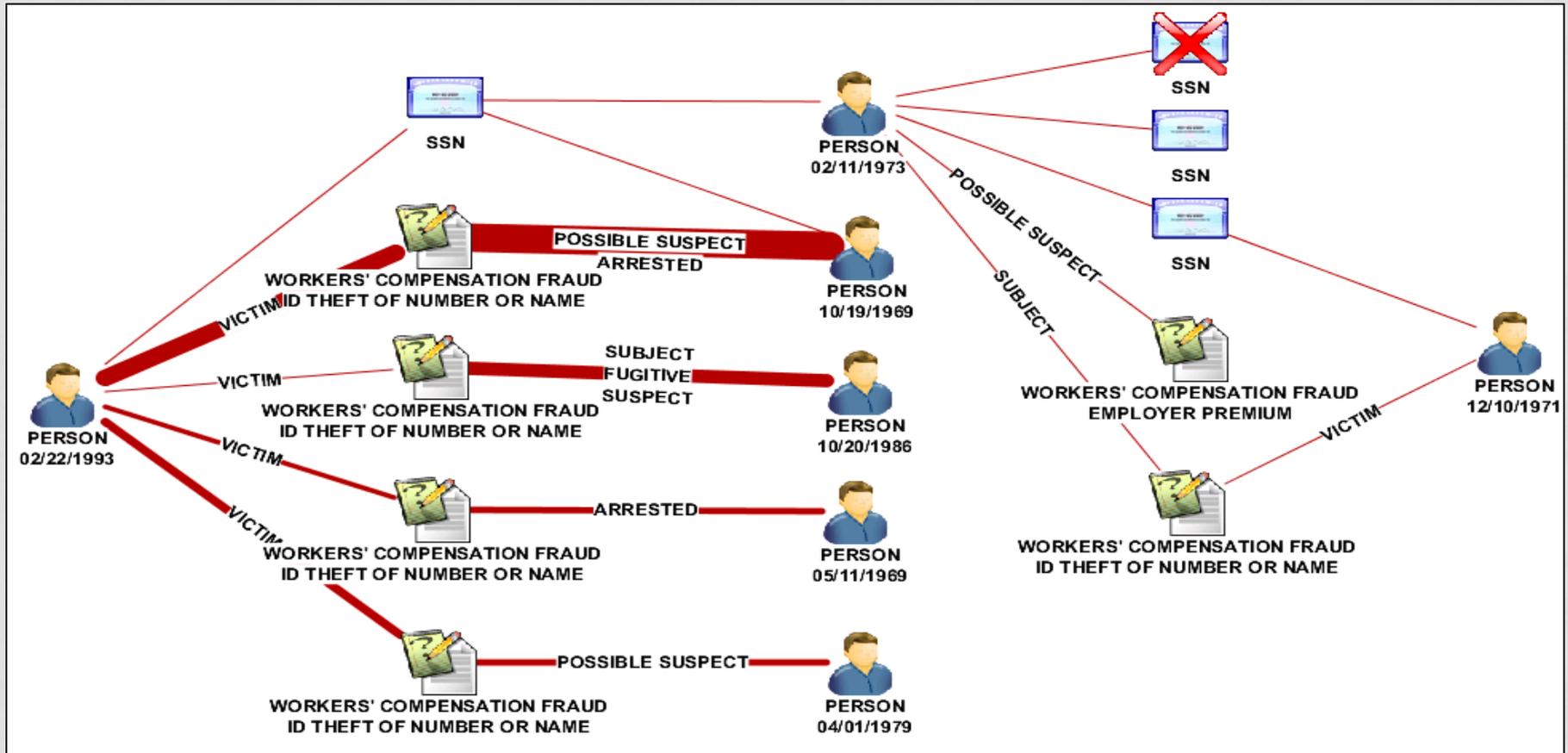
Number Verification Service



Previous Case Matches/Checks



# IDENTITY THEFT (WORKER COMP)



# FINAL THOUGHTS



# GAO'S ONGOING INITIATIVES

- **GAO's Center for Enhanced Analytics (CEA):** a new initiative that works to enhance access to data sources, deploy new technologies, and promote the use of novel analytic approaches in GAO's work.
- **Comptroller General Forums on Data Analytics:**
  - Purpose: bring together industry and public sector professionals to discuss challenges and identify opportunities for using data analytics for oversight in private and public sectors
  - Upcoming forum will focus on the use of analytics in government and document instances where analytics can be used to address fraud and improper payments.
- **A Framework for Managing Fraud Risks in Federal Programs (GAO-15-593SP):** directs agencies to evaluate data using a risk-based approach and adapt activities to improve fraud risk management.



# SOME LESSONS LEARNED

- Simple solutions can yield good/decent results
- Don't necessarily need complex/expensive software
- Combine data sources to show more/better patterns
  - Simple – weather (Mininno), SSDMI,
  - Complex – case management, core systems
- Metadata is a key feature for exposing patterns
  - Dates/times (behavior)
  - Lookup tables (watch lists, locations, SSDMI)
  - Custom heuristics and rules (DR #, DEA #, etc)
- Collaboration is important
  - Share results when possible
  - Patterns will be similar across systems
  - May find same people compromising multiple systems