NHS
Counter Fraud Authority

Machine learning/AI opportunities in the identification of NHS fraud

7th November 2023

Matthew Jordan-Boyd
David Dixon

What We See – Fraud Reporting

At least 1% of NHS funding is vulnerable to fraud

However

- 69% of organisations recorded no fraud identified from reactive investigation
- 76% of organisations recorded no fraud prevented
- 82% of organisations recorded no funds recovered from reactive investigations
- 77% of organisations recorded no sanctions of any type

Opportunity - Data and the NHS

- Electronic data captured at the point of almost every interaction with the NHS (patient, staff, supplier).
- Covid taught the world digitisation is paramount, with many manual practices becoming digital to accommodate the pandemic and remaining in effect.

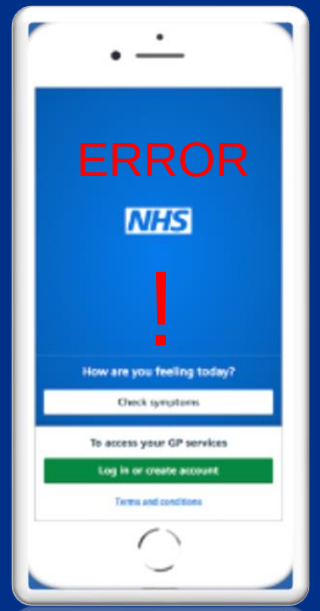
e.g. GP appointment can now be conducted by video; every element of which is data (Images can be turned into numbers)

- Estimates suggest NHS data is worth £10bn a year through operational savings, improved patient outcomes and wider benefits to the economy

…but the value to research, e.g. bespoke patient treatment plans that include genomic data from cradle to grave - surely priceless.

Our shared fraud challenge is ensuring as much money as possible goes towards ensuring patients get the best standard of care.

Challenge - Data and the NHS



- Only 20% of NHS organisation are digitally mature. With 86% having a form of electronic patient record in place. But the majority of NHS services yet to have a digital foundation in place (UK Gov't, June 2023).

- UK Gov't estimates that digitally mature trusts operate with approximately 10% improved efficiency compared with their less digitally mature peers.

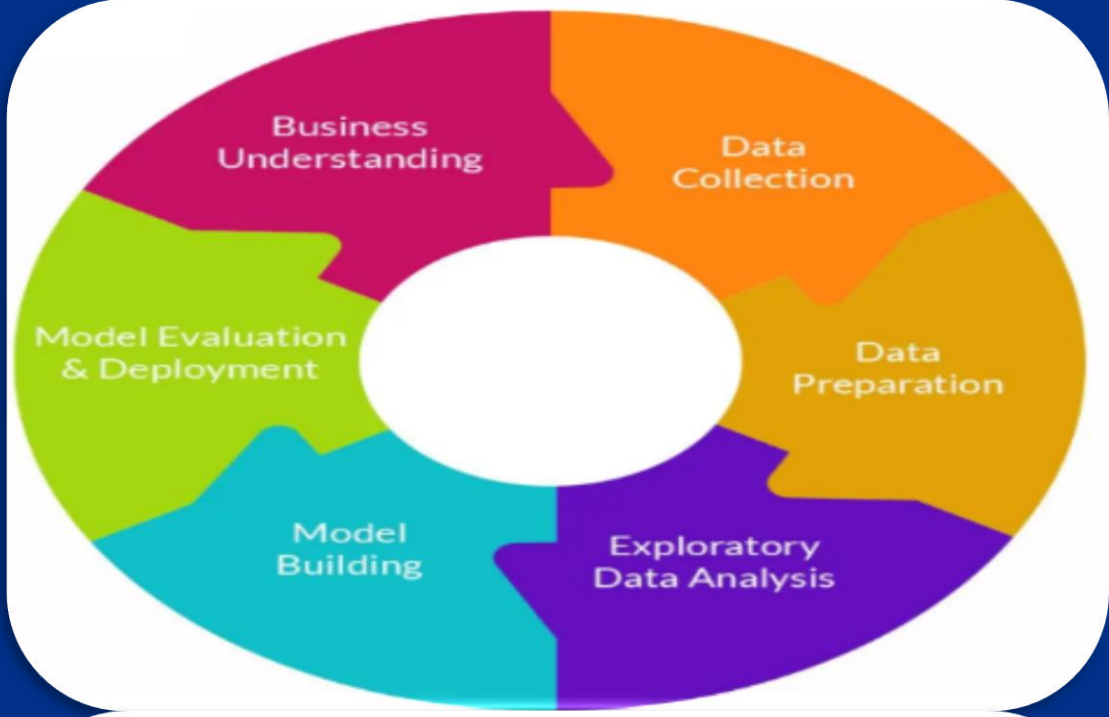
- This extends to mitigation and prevention of fraud through oversight and deterrence
  e.g. the ability to detect and prevent fraud through scanned copies of paper-based invoices as opposed to structured datasets that can be analysed.

The NHS data ecosystem is made up of multiple fragmented and geographically dispersed data silos often even scattered across several sites.

Consistent and centrally held datasets are few and far between, even where the same systems are in use.

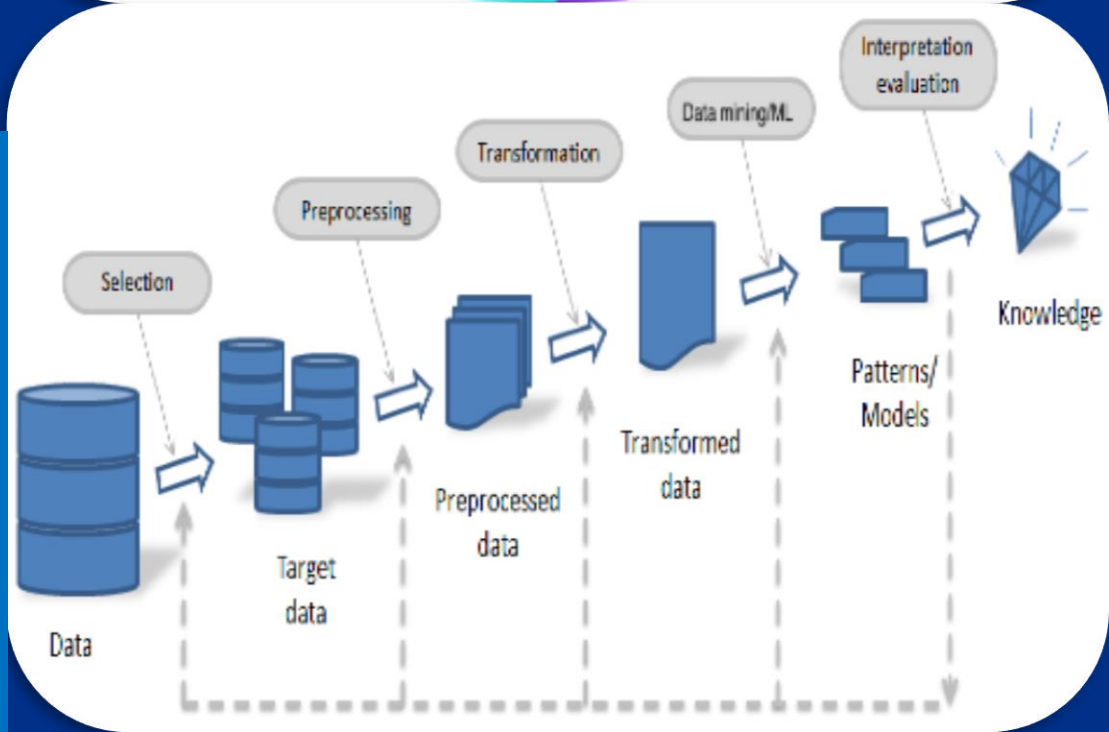
The NHSCFA Approach

- Structured approach to analytical enquiry through the application of data.
- The core of our approach always starts with “**THE PROBLEM**” and then focussing on answering the problem presented.
- This approach formed part of the design of the Cabinet Office Analytical Standards.



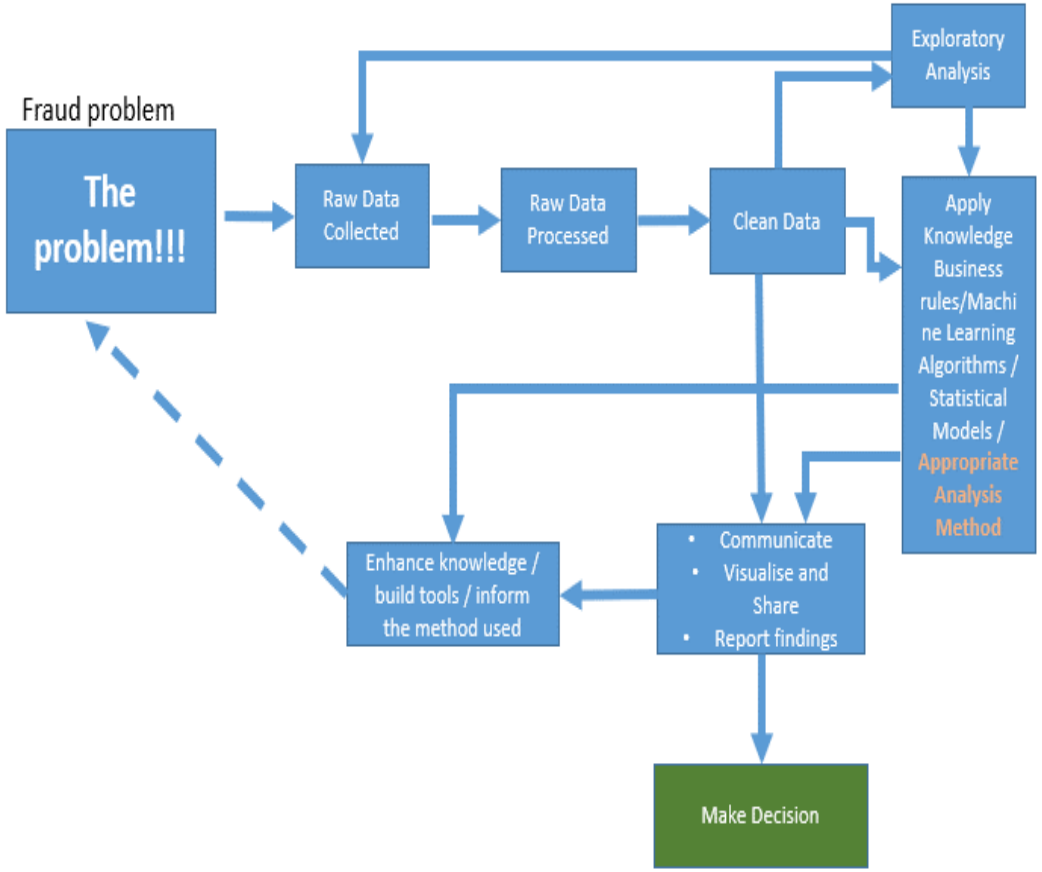
ONLY THEN CAN YOU:

1. Answer the business questions/ requirements presented at the beginning of a project.
2. Illustrate knowledge, information, and insights derived from the analysed data.
3. Tell ‘the story’ in way that contains valuable and relevant result-focussed recommendations or anomalies



Opportunity – A Developed CF Analytical Framework

- **Explore** where / if the data exists and how it can be accessed
- **Process** raw data (is it in the right format / structure, joined and filtered from multiple systems)
- **Clean** the data, Is this always necessary? Some data need to be “dirty” to find the outlier
- **Contextual recognition**– Is there anything else in the data that has emerged? What is the data telling us?
- **Application** of appropriate analytical approach (for NHSCFA this is currently business rules and application of knowledge)
- **Report** findings to inform decision making / action and use the knowledge gained to improve & monitor



THIS DOESN’T OCCUR IN ISOLATION!

- Happens multiple times in one project (we call them iterations)
- Projects can grow and expand

NHS
Counter Fraud Authority

Investigation

- Utilise a formalised tasking process to transform high assurance intelligence packages into counter fraud investigations

Successful outcome – sanctions & redress (£)

Problem-centric approach and domain expertise

Information Analytics

- Undertake proactive data projects to derive fraud loss through outlier detection

Successful outcome – Outliers indicative of fraud

Data from external sources (other NHS bodies and wider sources)

Intelligence packages (potential investigations)

Problem-centric approach and domain expertise

Outlier development into intelligence packages

Findings to inform fraud prevention activity

Information from external parties (the public, wider NHS, Police etc.)

Intelligence

- Turn information received from internal /external stakeholders into intelligence packages for potential investigation
- Extrapolate fraud risks to estimate financial loss

Successful outcome – inference of fraud and extrapolated estimations of fraud loss (£)

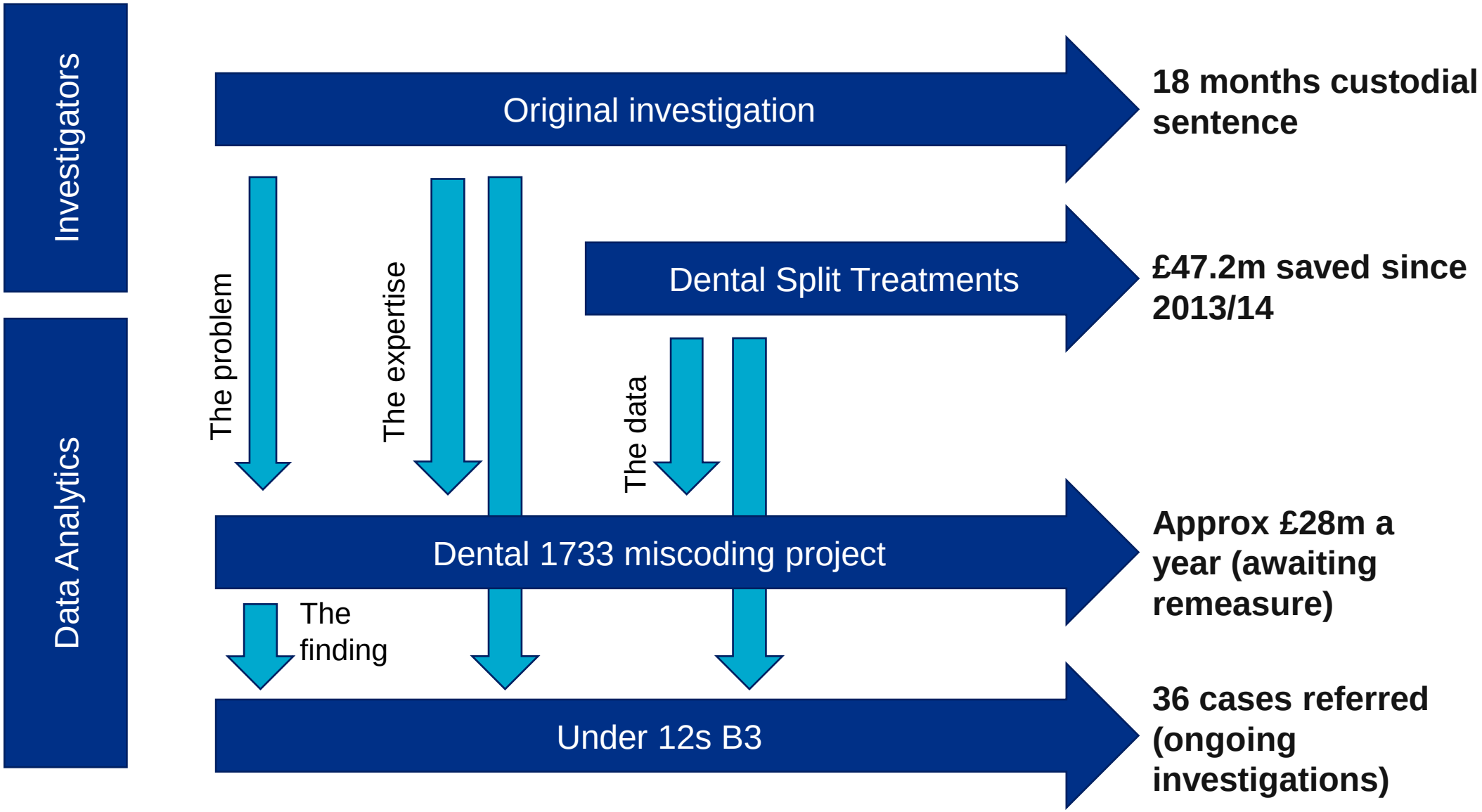
Identified fraud risks (system weaknesses)

Fraud Prevention

- Develop fraud prevention activity through policy change, new guidance, education etc.

Successful outcome – changed behaviour (i.e. fraud prevented) (£)

In practice: Dental Case Study



Challenge: Advanced Tools & Methods

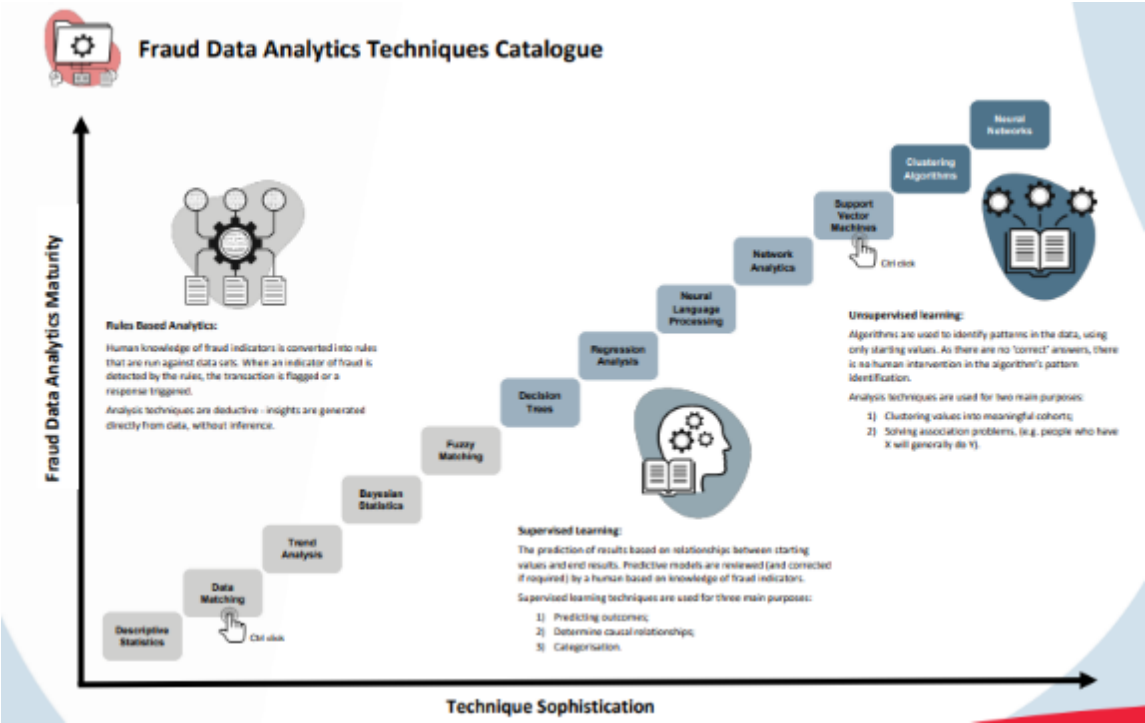
Not one tool fits all - depends on several factors

- Data (qualitative / quantitative)
- Problem
- Desired Outcome

Once you have all these factors then a specific algorithm or approach can be identified and applied.

More than one approach/method/algorithm in most instance must be used. For example, rule based with clustering to produce a risk base output

- Rules (if you can see the pattern, we can build)
- Statistics – exploration / sampling
- ML – supervised / unsupervised / semi-supervised



Machine Learning and Counter Fraud

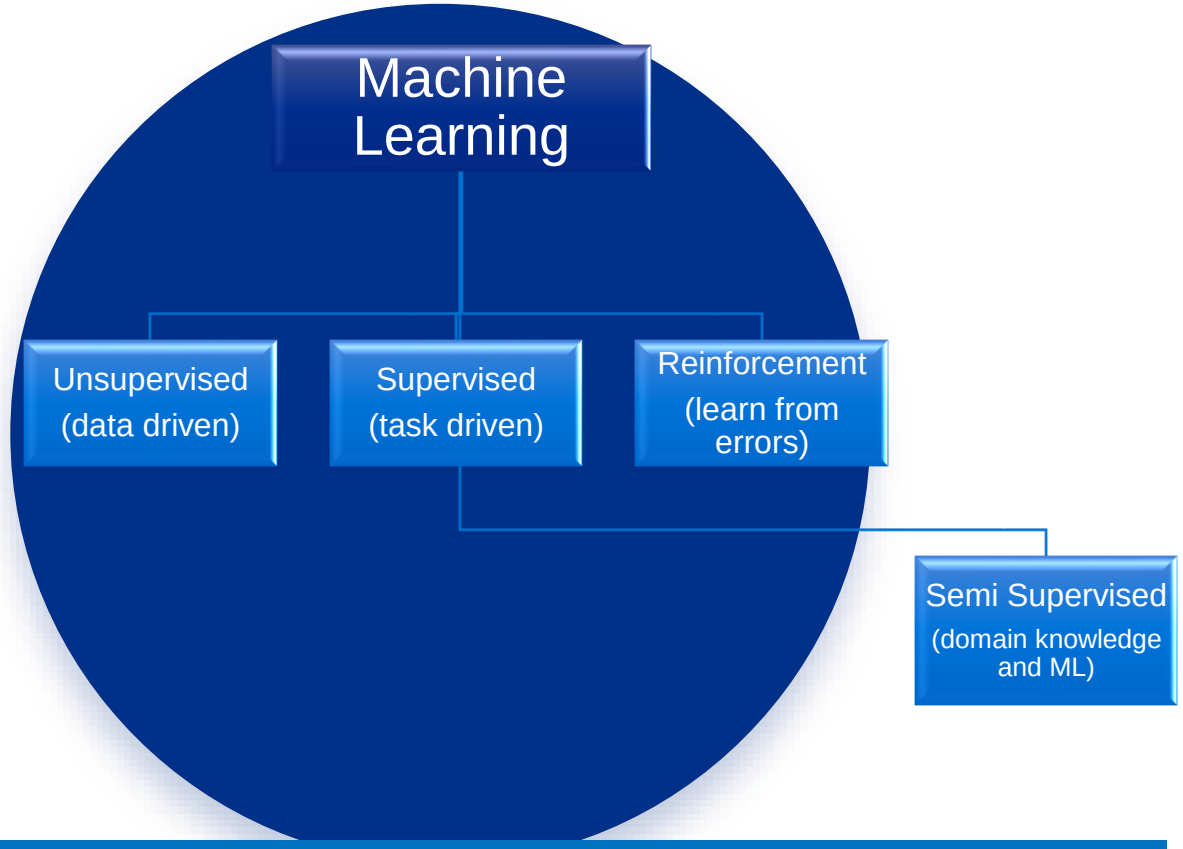
Three main types of machine learning, dependant on problem to be answered and data available (This will determine the method and algorithm(s) used)

Most methods use **Supervised**: a fraud flag (**classified/labelled**) is applied to the individual record e.g. **credit card fraud**

Second method is **Unsupervised.** N**o flag/no label** - but very limited to what methods can used.

- No Fraud Flag captured within NHS fraud cases and applied to individual records. But this is prevalent within the fraud domain.

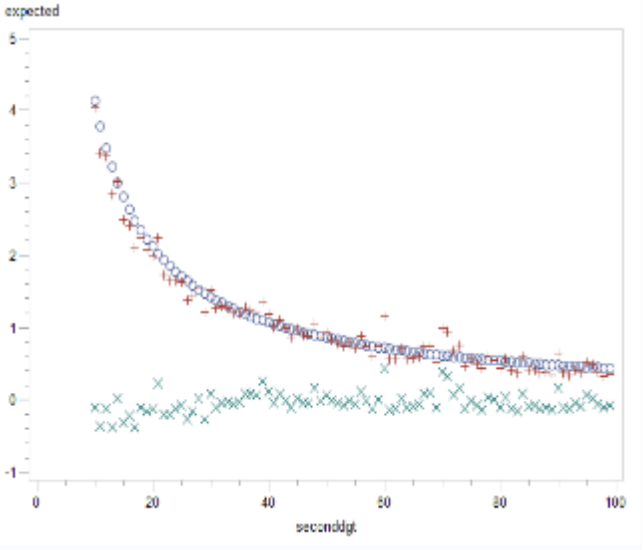
- **Semi supervised** is a way to use domain knowledge and create the needed flag required to use supervised ML algorithms (weaker statistically)

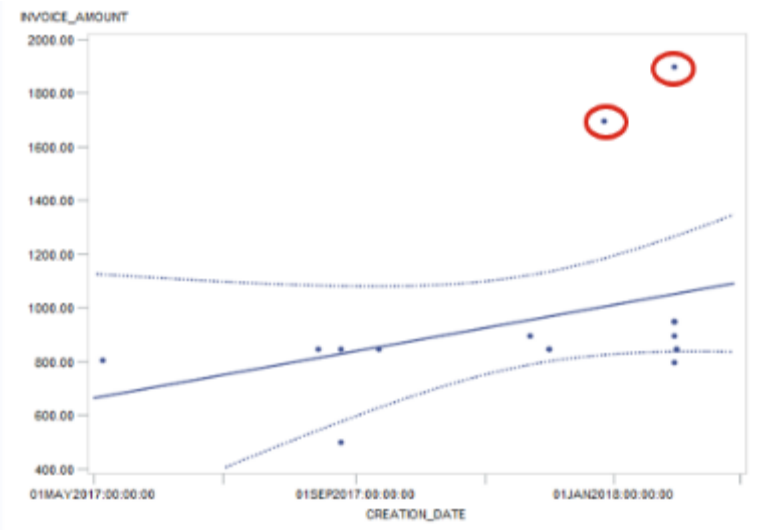


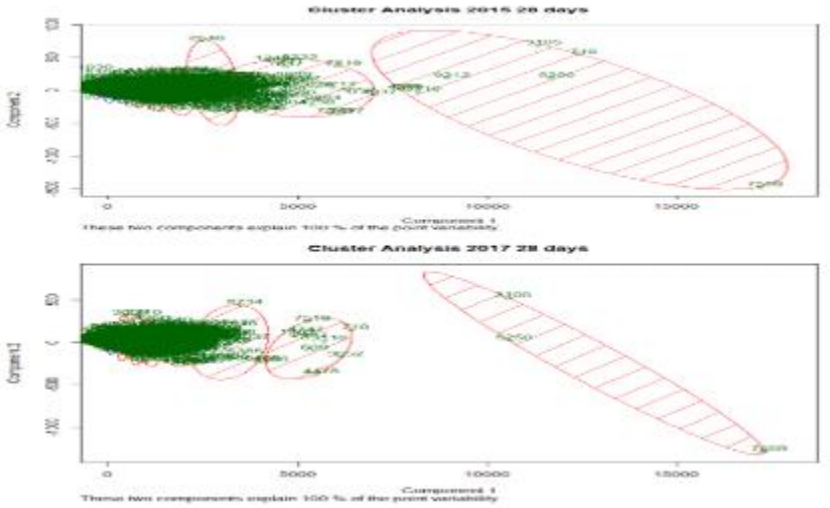
Models improve over time with more examples of fraudulent behaviour

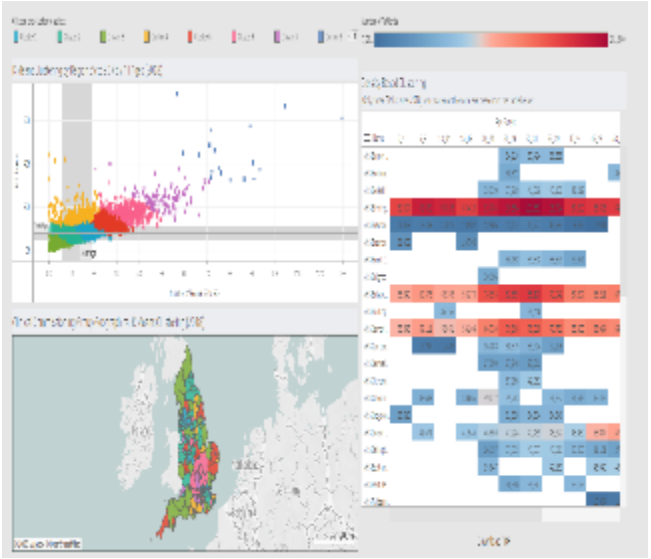
Many outliers are statistically insignificant (0.001%)

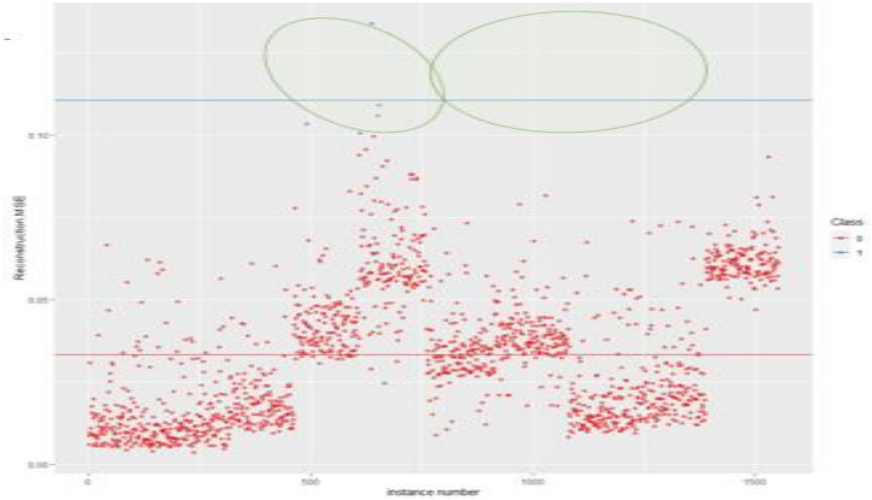
Machine Learning Examples in NHSCFA







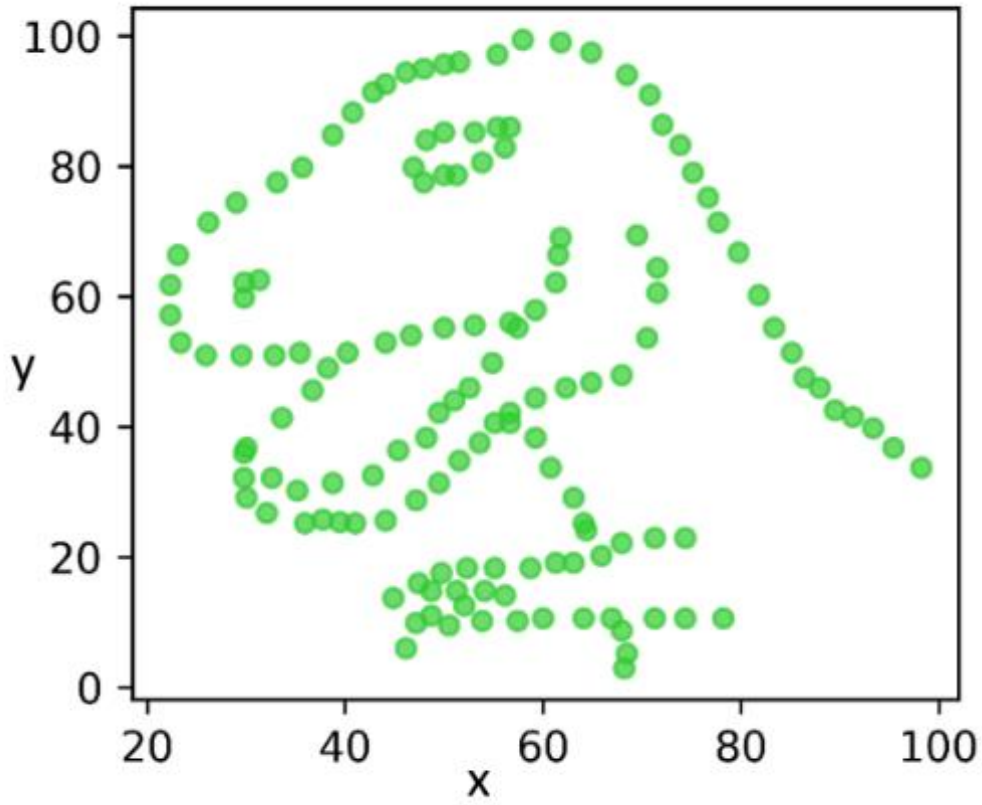




What is this data telling us?

- All data has a story, but understanding it is more complicated.
- Even when applying a methodology data science framework, often a table of data offers no insight into the problem.
- Behaviour and patterns continue to be hidden from view
- For example: what does this data tell you?

	ID	NHSN	NINO
1	NHS_ID	55.3846	97.1795
2	NHS_ID	51.5385	96.0256
3	NHS_ID	46.1538	94.4872
4	NHS_ID	42.8205	91.4103
5	NHS_ID	40.7692	88.3333
6	NHS_ID	38.7179	84.8718
7	NHS_ID	35.6410	79.8718
8	NHS_ID	33.0769	77.5641
9	NHS_ID	28.9744	74.4872
10	NHS_ID	26.1538	71.4103
11	NHS_ID	23.0769	66.4103
12	NHS_ID	22.3077	61.7949
13	NHS_ID	22.3077	57.1795
14	NHS_ID	23.3333	52.9487
15	NHS_ID	25.8974	51.0256
16	NHS_ID	29.4872	51.0256
17	NHS_ID	32.8205	51.0256
18	NHS_ID	35.3846	51.4103
19	NHS_ID	40.2564	51.4103
20	NHS_ID	44.1026	52.9487
21	NHS_ID	46.6667	54.1026



The lesson?

- More than one method must be used (rule based and scientific)
- Nothing ever 100%, only confirmed by using a domain expert (who knows what a dinosaur looks like)
- Tools give us the ability to break the data down, but the story of the data is often more than face value

Challenge: Access to data

NHS fraud is within NHS activity – which inherently concerns patients, staff and stakeholders.

No data = no analytics = no outcomes

NHSCFA owns very little bulk data – most we must access from wider stakeholders (e.g. NHSBSA) through formalised data shares.

--

The investigative elements within counter fraud naturally utilise very specific data requests which are well established through criminal legislation

…as a more recent development for CF activity, the legal gateways for bulk data are far less developed

…The DPA and GDPR caused a very risk adverse environment for data sharing, even where there is a strong public interest.

And of course, the public MUST have confidence in how we access and use data

Opportunities: Access to data

This historically has provided challenge: data access issues can impede, impair and even prevent projects being undertaken.

NHSCFA has developed a mature approach, but no option is perfect (or without risk or delay):

Anonymise data	Pseudonymise data	Access and use personal data
• **Ensures swifter/smoother data shares** • **No need for legal gateways – MOU's, ISAs, NHSCFA mandate is sufficient** • **Limits what can be achieved** • **Necessitates follow-up requests WITH personal data to utilise findings**	• **Ensures swifter/smoother data shares** • **The pseudonymisation tool must be developed, tested and used** • **Distinct pseudonymisation critical** • **Prevents using personal data (i,e, surname, NI no.) to join to separate datasets**	• **Allows use for intended purpose** • **No developmental steps to remove/edit the dataset** • **Necessitates development of DPIA's, MOU's etc within the framework of a legal gateway** • **Legal frameworks can be open to interpretation and provide an easy excuse**

This is supported by:

- Sound stakeholder engagement / domain expertise
- The problem centric approach - no fishing expeditions!
- Legal basis
- Support and emphasis from senior management to overcome hurdles

Our shared challenge

Returning to the beginning - the scale of the problem (£1.19bn)

- Our interests are aligned, and we have a shared responsibility to safeguard our NHS.
- Naturally, directing resource is critical, particularly at a local level.
- Prevention is always the best solution.
- The value of assurance, regardless of findings.
- Recognising that AI, ML and advanced analytics takes a lot of time and specialist tools – an accessible solution is needed NOW

Risks of the wrong approach

- "Analysts, analyse!" – untargeted analysis
- The push to 'Go Big' immediately and the danger of drowning in data
- Forgetting that analytics is only part of a wider process – what matters is what you can do with it
- Forgetting a tool is just that. Data is the fuel.

So, how can a proactive data culture be instilled in localised counter fraud activity across the

Shared opportunities

1) Play to your advantages

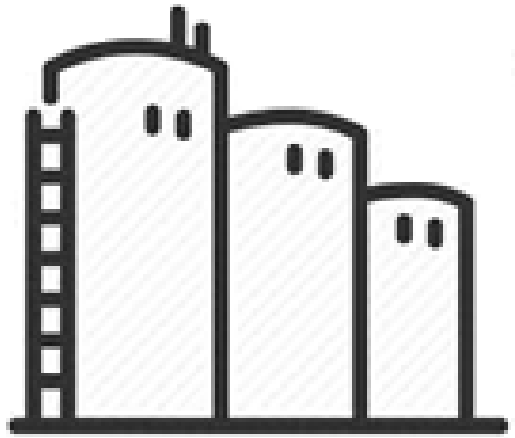
The domain knowledge is already present and available “in-house”

…as is the data itself, held in a single, consistent format

Development of a problem centric approach in house that is bespoke to your circumstances – with the opportunity to escalate

Start small and focussed – use a real case example

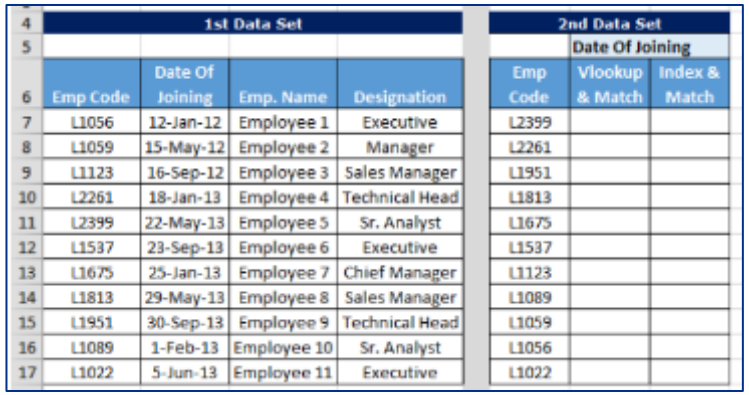
Develop a collaborative approach led by localised experts



Shared opportunities

2) Use what you have

The most fundamental and useful tool for proactive analysis is data matching, which can be achieved on Microsoft Excel.



	1st Data Set				2nd Data Set		
					Date Of Joining		
6	Emp Code	Date Of Joining	Emp. Name	Designation	Emp Code	Vlookup & Match	Index & Match
7	L1056	12-Jan-12	Employee 1	Executive	L2399		
8	L1059	15-May-12	Employee 2	Manager	L2261		
9	L1123	16-Sep-12	Employee 3	Sales Manager	L1951		
10	L2261	18-Jan-13	Employee 4	Technical Head	L1813		
11	L2399	22-May-13	Employee 5	Sr. Analyst	L1675		
12	L1537	23-Sep-13	Employee 6	Executive	L1537		
13	L1675	25-Jan-13	Employee 7	Chief Manager	L1123		
14	L1813	29-May-13	Employee 8	Sales Manager	L1089		
15	L1951	30-Sep-13	Employee 9	Technical Head	L1059		
16	L1089	1-Feb-13	Employee 10	Sr. Analyst	L1056		
17	L1022	5-Jun-13	Employee 11	Executive	L1022		

… as can rule based analysis, using “IF” statements

--

Utilise existing data matching –

e.g. the National Fraud Initiative (NFI)

NHSCFA Disseminations – Less than 2% of the disseminations shared by NHSCFA from 1st April to end of June 2023 resulted in an incident or investigation being recorded on the NHSCFA Case Management System

….Even if it doesn’t find an outlier, it may find a ‘problem’

Shared opportunities

3) Know your organisation

The counter fraud perspective
- strengths, weaknesses, assurance, ambiguity.

Policy meeting practice

The nature of your counter fraud provision – both proactive and reactive.

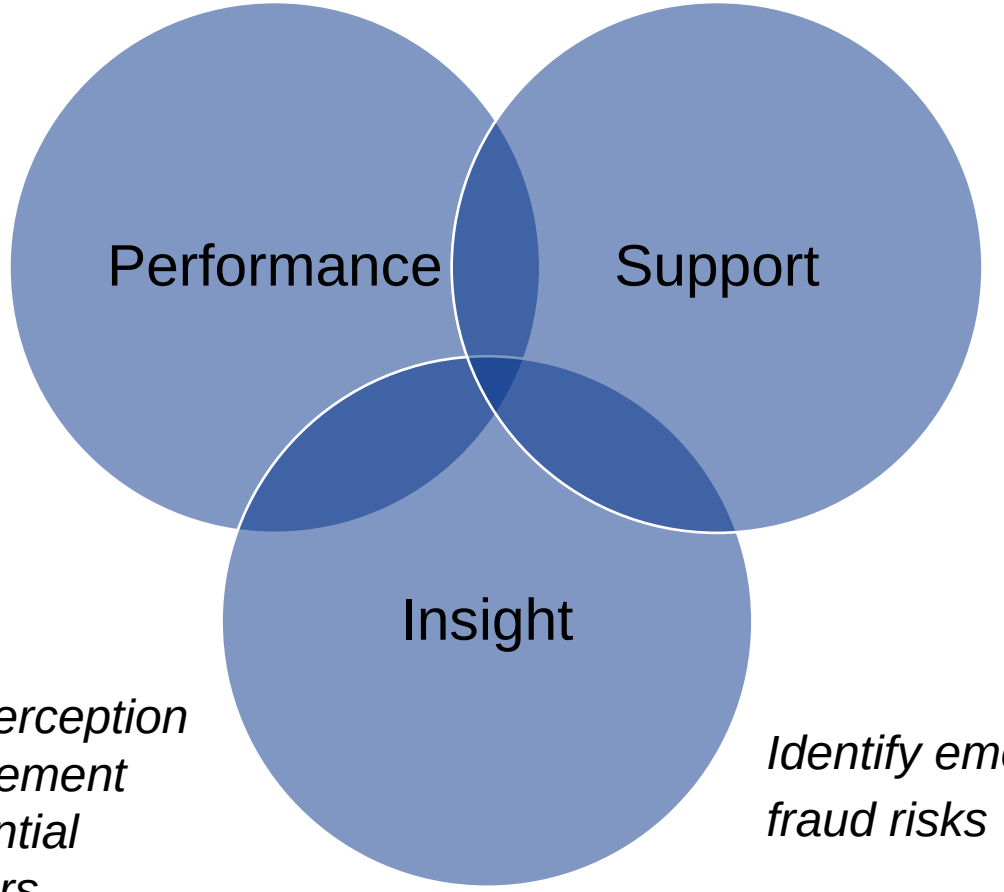
Context against a wider picture

- A tool for LCFS's, Fraud Champions, Directors of Finance and Audit Committee Chairs
- Launching in Late Nov / Early Dec
- A quarterly reporting tool, delivered directly to your inbox via URL
- Bespoke reporting on your health body, against the wider context

Purpose:

Return on investment for counter fraud provision

Highlight best practice in counter fraud activity



Reactive and proactive effectiveness

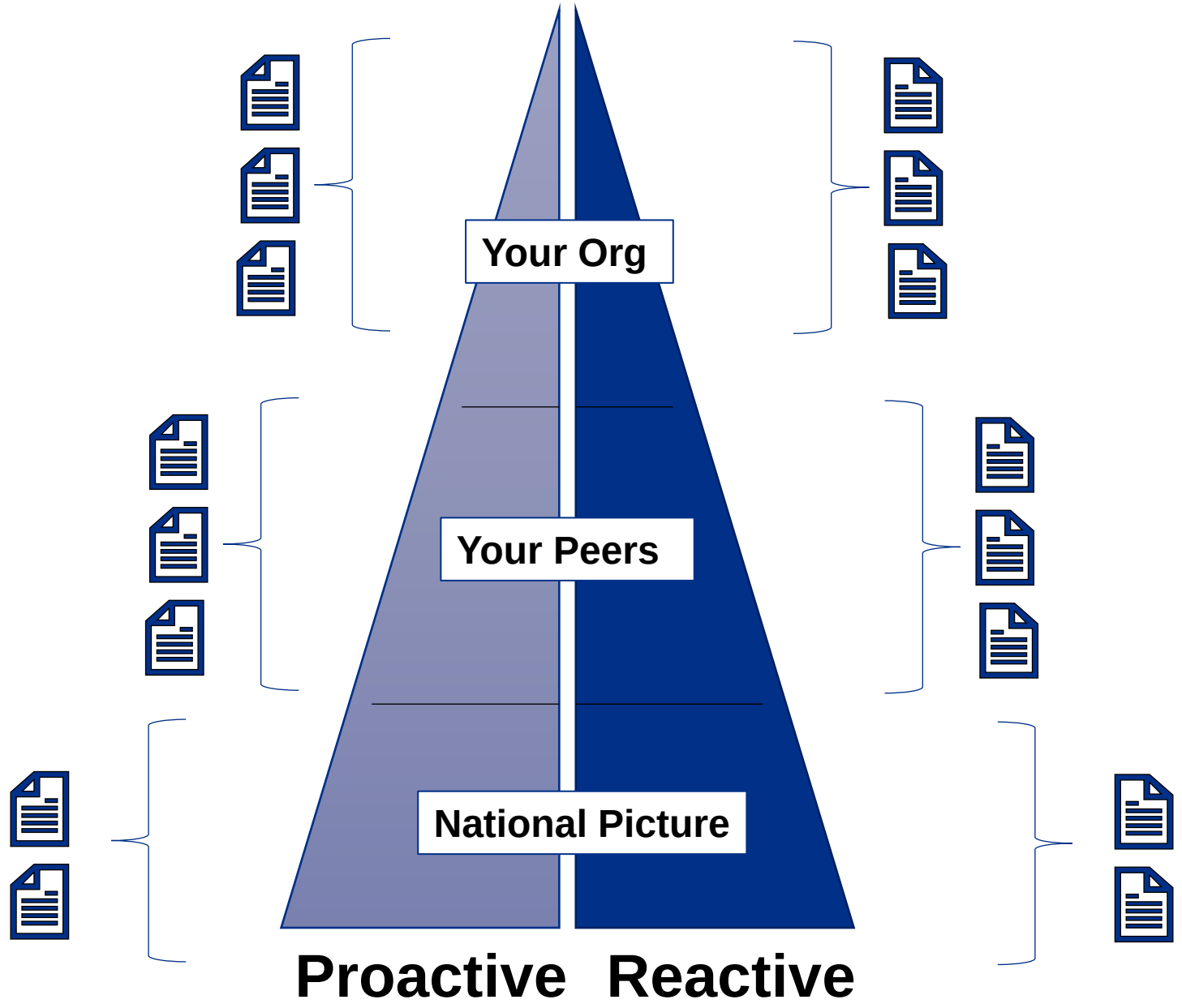
Encourage collaboration and wider activity

Increase perception and engagement from influential stakeholders

Identify emerging fraud risks

X Saying what is right or wrong, only what ‘is’ - and where it is comparable

Structure of reports



DoF / ACC only

- Return on Investment
- Cost / compliance matrix
- Sector/national averages

Organisation Name

All

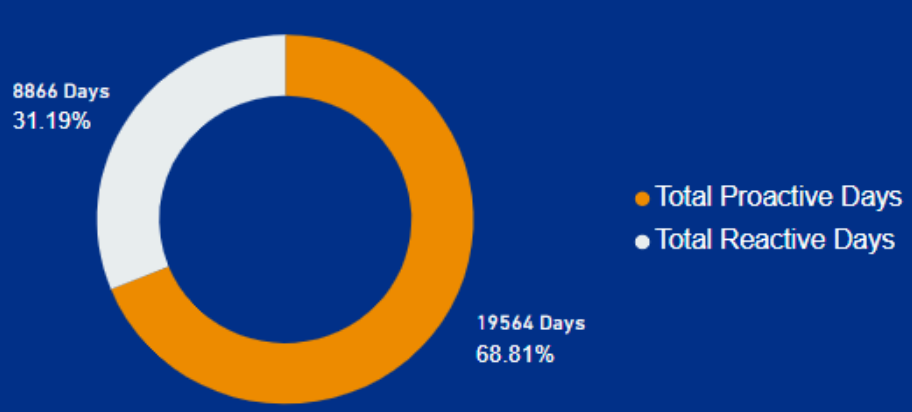
NHS
Counter Fraud Authority

Healthbody Return On Investment 2022-23

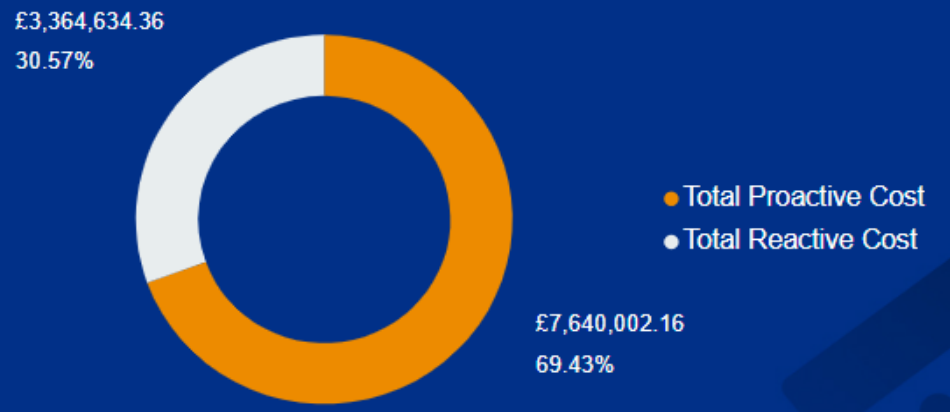
Total Fraud Identified	Total Fraud Prevented	Total Fraud Recovered
£1,332,305	£7,325,533	£469,047

Fraud Identified for Every Pound Spent	Fraud Prevented for Every Pound Spent	Fraud Recovered for Every Pound Spent
£0.12	£0.67	£0.04

Total Proactive and Reactive Days



Total Proactive and Reactive Cost



Screenshot utilise dummy data intended for demonstration purposes only.

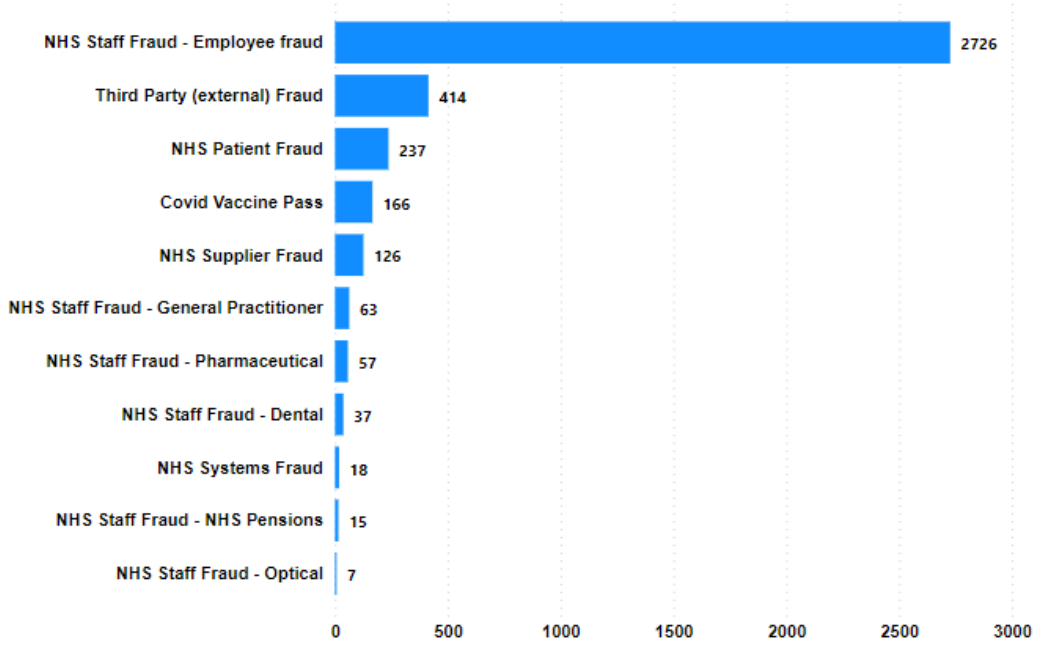
Healthbody Reactive Activity: Investigations

NHS Counter Fraud Authority

Organisation Name: All | Data Period: Year: All | Data Period: Quarter: All

Investigations Created Since Apr-21	Investigations Closed Since Apr-21	Investigations Currently Open
3866	2233	1633

Investigation Type



Investigation Subtypes

Subtype	Open Investigations	Closed Investigations
Timesheet/overtime fraud	354	392
Working Whilst Sick	303	541
Employee Fraud - Other	232	278
	190	218
Grant Funding	69	29
Recruitment process	64	108
Employee declaration	47	51
Payment Fraud - Mandate fraud	42	155
Travel/subsistence fraud	29	50
Identity Fraud	27	28
Dispensing fraud - dispensing irregularities	23	17
Post contract fraud - Invoicing fraud	20	25
Usage fraud - Plastic card	19	10
NHS Patients - Misuse of services	18	41
NHS Patients - Misuse of prescriptions	17	38
NHS Patients - NHS additional funding	17	18
Employee insider issues	16	20
Staff collusion - Bribery	14	25
Dental contract - dental activity	12	4
NHS Supplier Fraud	11	15
Payment Fraud - unsolicited requests	10	39
GP Practice - False claims	9	9
GP Practice staff - diversion of funds	9	6
Overseas Patients - Secondary Care	9	15
Dental contract - patient claims	7	5
Administration - Diversion/Theft of funds	6	1
Total	1633	2233

Screenshot utilise dummy data intended for demonstration purposes only.

Sector Reactive Activity: Investigations

NHS Counter Fraud Authority

Data Period: Year: All | Data Period: Quarter: All | Headcount: All

Investigations Created Since Apr-21

3866

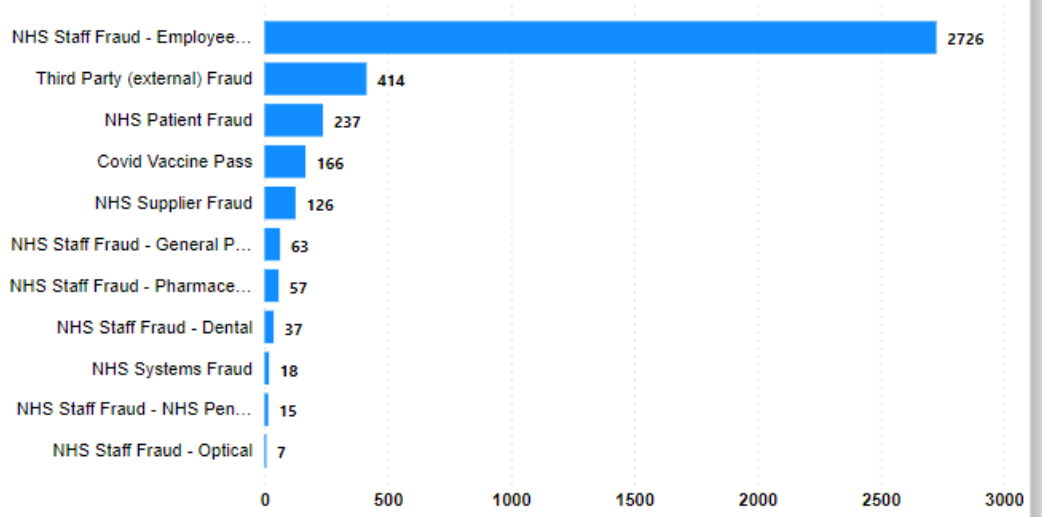
Investigations Closed Since Apr-21

2233

Investigations Currently Open

1633

Investigation Count by Type



Subtype	Open Investigations	Closed Investigations
Abuse of contract	1	1
Administration - Diversion/Theft of funds	6	1
Administration - Withholding funds	0	1
Contractor Collusion - Production/supply limitation	3	0
Dental contract - dental activity	12	4
Dental contract - patient claims	7	5
Dental Income - Double income fraud	4	1
Dispensing fraud - dispensing irregularities	23	17
Dispensing fraud - Double income	1	2
Dispensing fraud - Patient services	0	4
Dispensing fraud - reissued medicines	1	1
Domiciliary - False claims	1	0
Domiciliary - Testing/dispensing	0	1
Employee declaration	47	51
Employee Fraud - Other	232	278
Employee insider issues	16	20
GP Practice - Capitation figures inc Ghost patients	1	2
Total	1633	2233

Acute - Multi-service | Acute - Other | Acute - Specialist | Acute - Teaching | Ambulance Trust | Community and Mental Health Trust | Community Provider Trust | Integrated Care Board | Mental Health and Learning Disability | Private sector provider

Screenshot utilise dummy data intended for demonstration purposes only.

Sector Proactive Activity

NHS Counter Fraud Authority

Data Period: Year: All

Data Period: Quarter: All

Headcount: All

Total LPEs Created within data period

5088

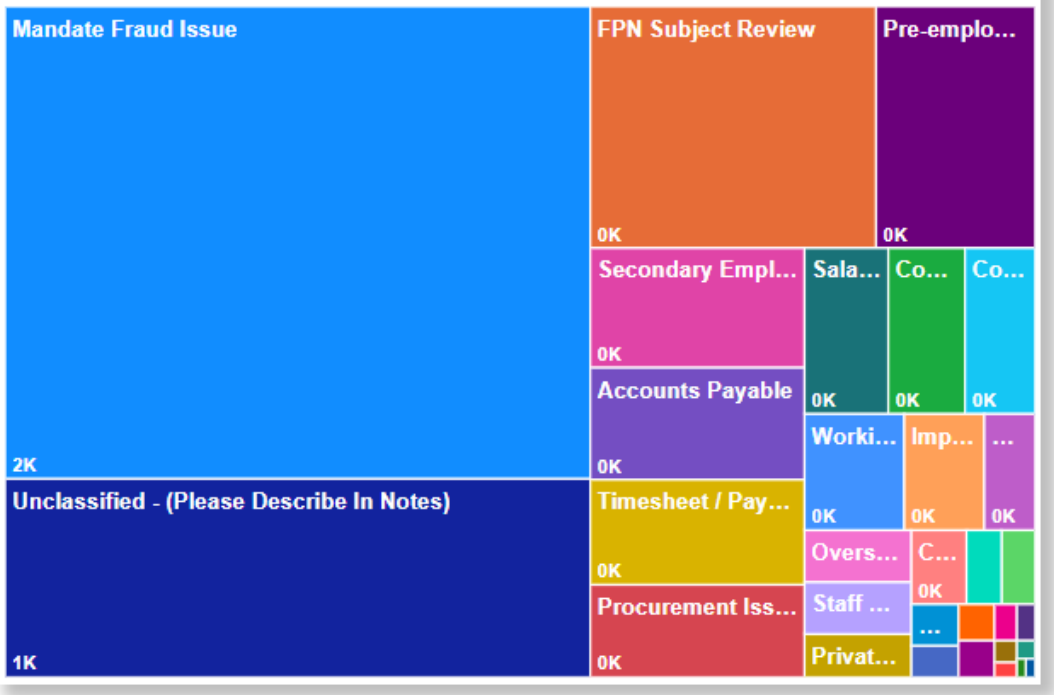
Total LPEs Currently Closed Since Apr-21

3870

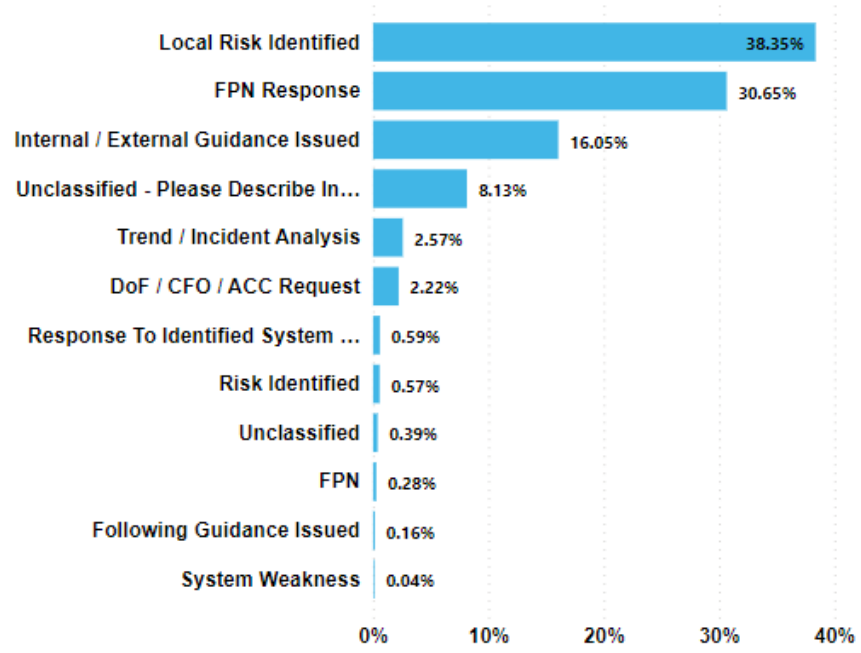
Total LPEs Currently Open Since Apr-21

1218

LPE Subject Area

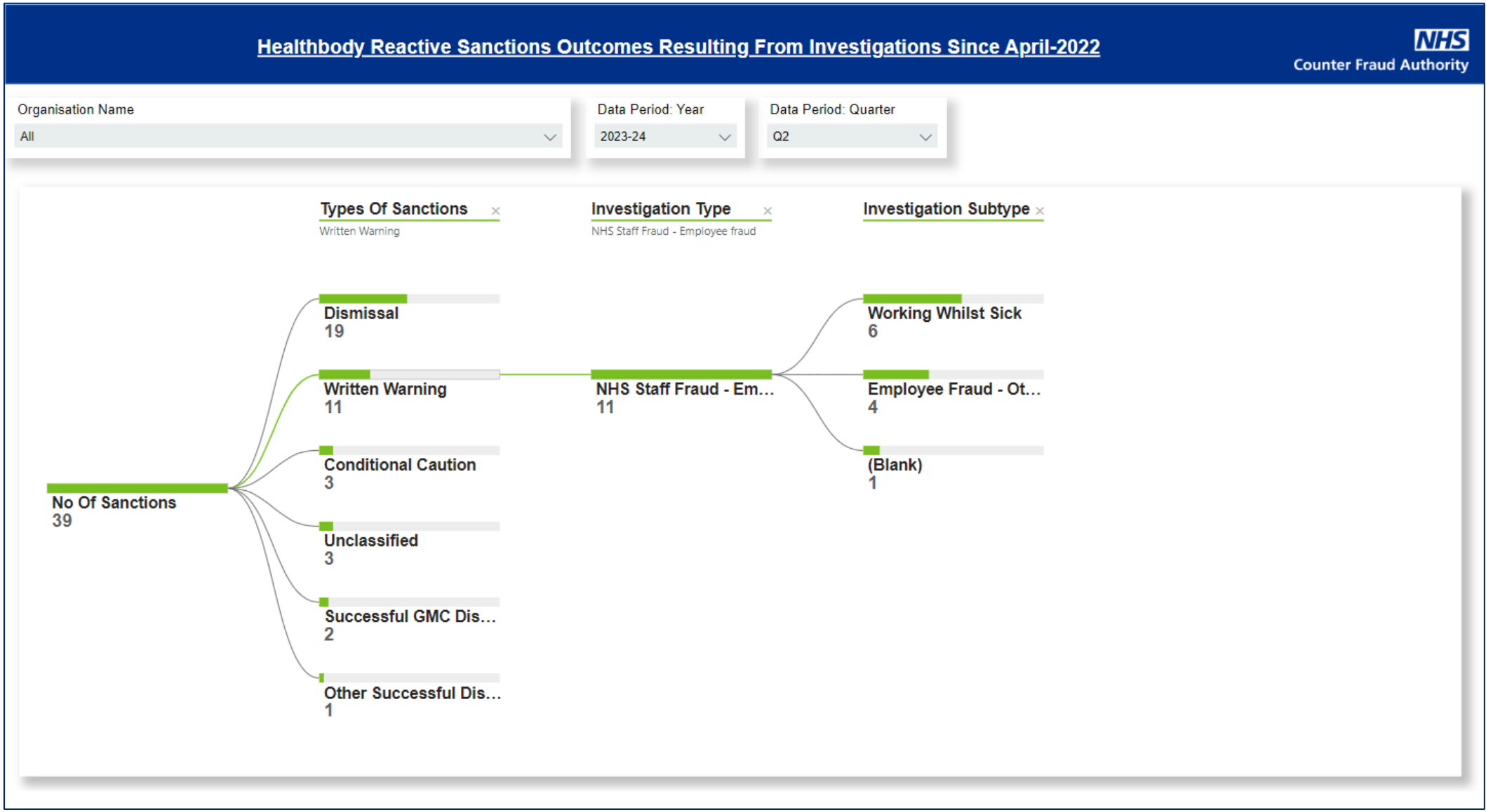


Risk Rationale

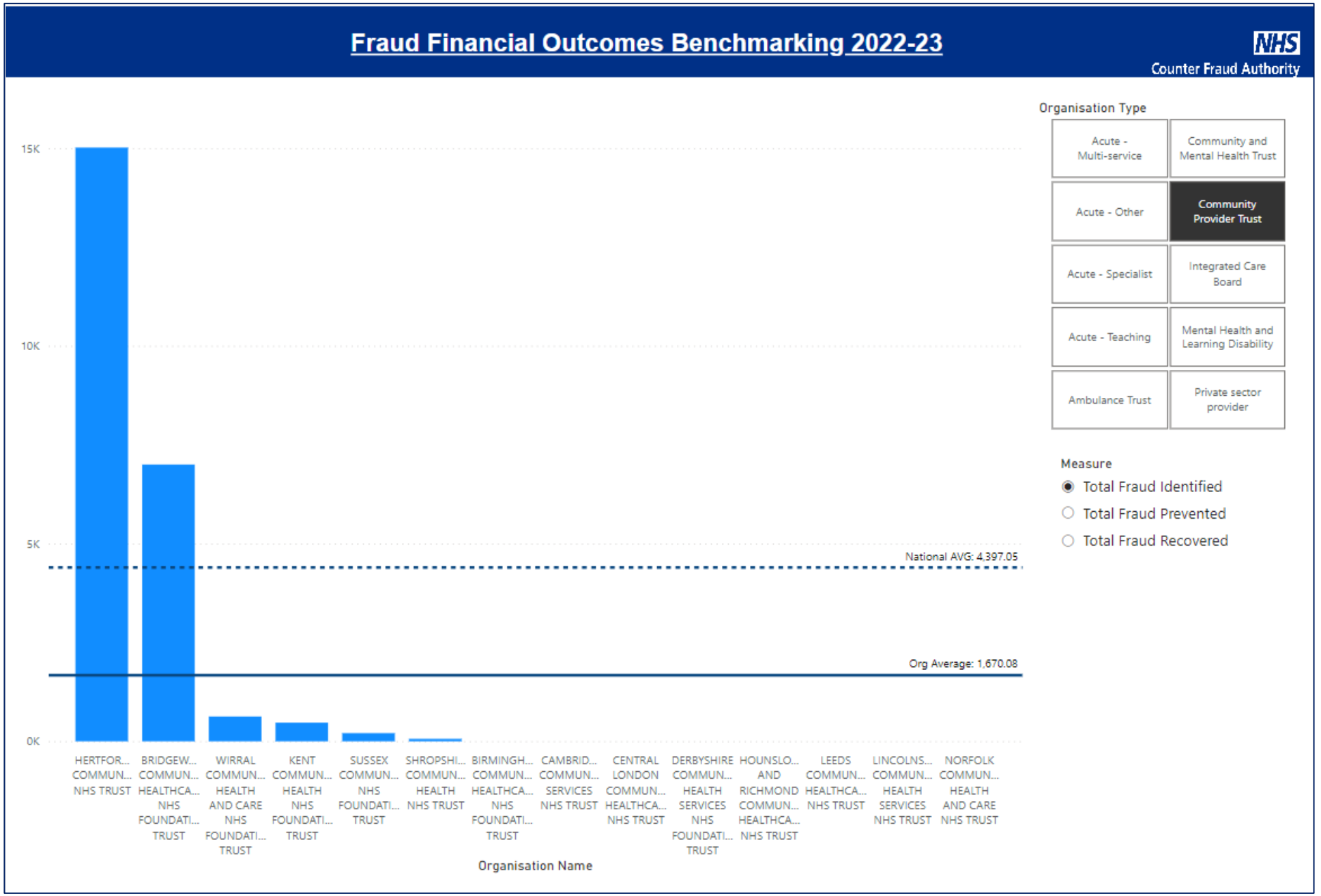


Acute - Multi-service	Acute - Other	Acute - Specialist	Acute - Teaching	Ambulance Trust	Community and Mental Health Trust	Community Provider Trust	Integrated Care Board	Mental Health and Learning Disability	Private sector provider

Screenshot utilise dummy data intended for demonstration purposes only.



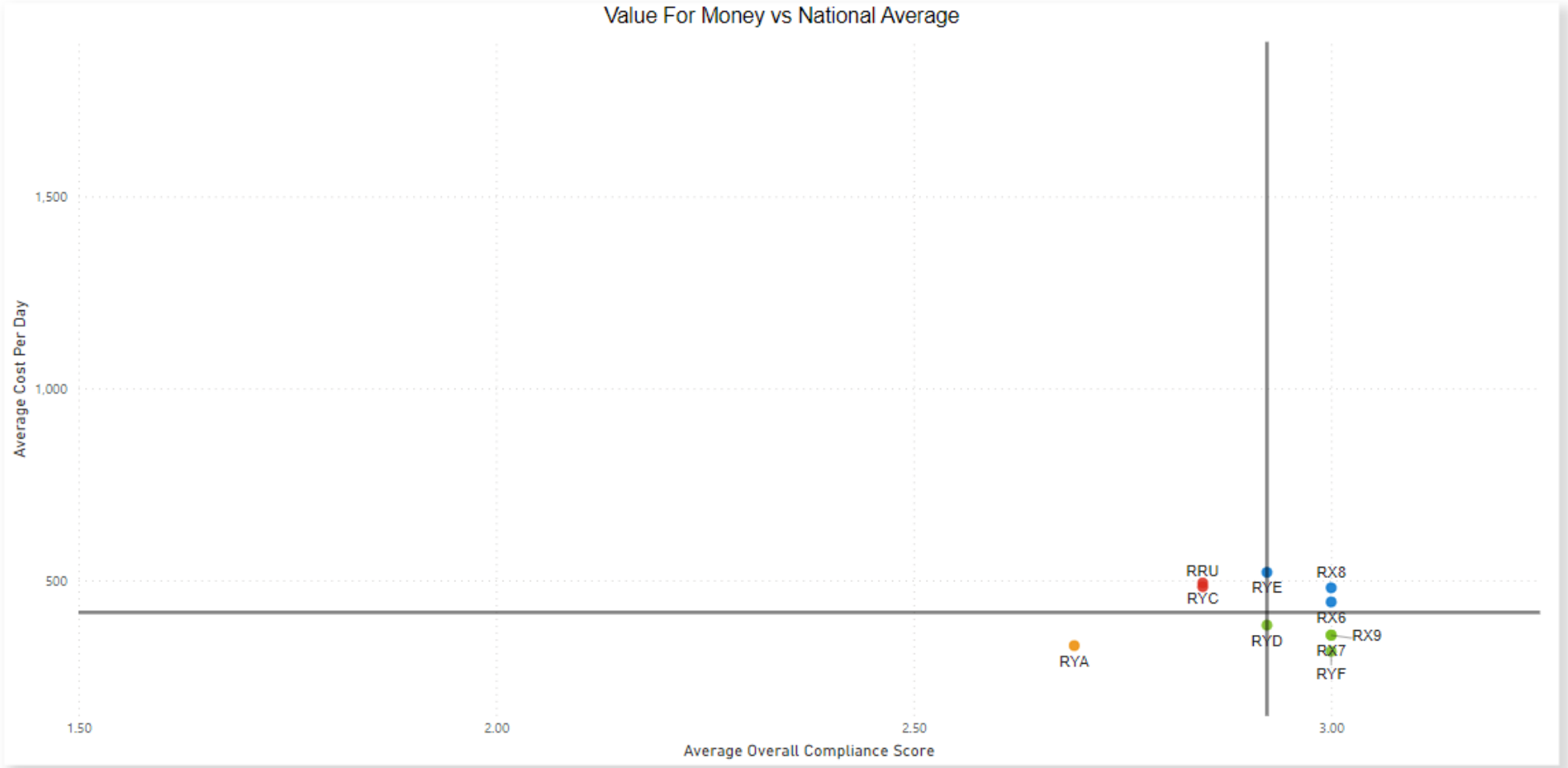
Screenshot utilise dummy data intended for demonstration purposes only.



Screenshot utilise dummy data intended for demonstration purposes only.

Value For Money 2022-23

NHS Counter Fraud Authority



Acute - Multi-service	Acute - Other	Acute - Specialist	Acute - Teaching	Ambulance Trust	Community and Mental Health Trust	Community Provider Trust	Integrated Care Board	Mental Health and Learning Disability	Private sector provider

Screenshot utilise dummy data intended for demonstration purposes only.

The Future

Data is now a valuable commodity but to fully realise its potential depends entirely on its application: treat data as an asset, but recognise its limitations alone.

A proactive, problem led approach is possible at all levels:

- Commence with the problem.
- Avoid the urge to ‘go big’ or complex immediately - start small and scale up
- Domain expertise is key
- Advanced AI and ML is a game changer, but the fundamentals are more important than anything.

The new External Reporting Tool : Watch this space! (launch and workshops to support use)

NHS
Counter Fraud Authority

Thank you