Real-world outcomes and biomarker testing in cancer patients:

An exploration of a novel genetic database from routine clinical practice in England

Fiona C Ingleby, PhD (IQVIA)

Saskia P Hagenaars, PhD (IQVIA); Alexandrina Lambova, MSc (IQVIA); Mounika Parimi, MSc (IQVIA); Stephen Benson (IQVIA); Sophie Jose (Health Data Insight, NHS); Lora Frayling (Health Data Insight, NHS); Valeria Lascano, PhD (IQVIA)

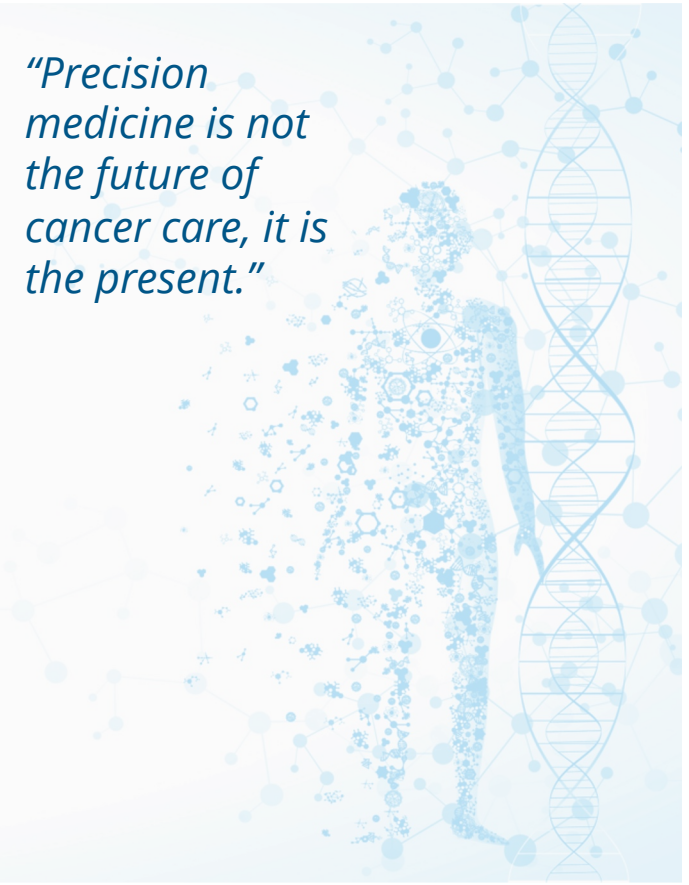
Contents

A brief introduction to CAS-MDx

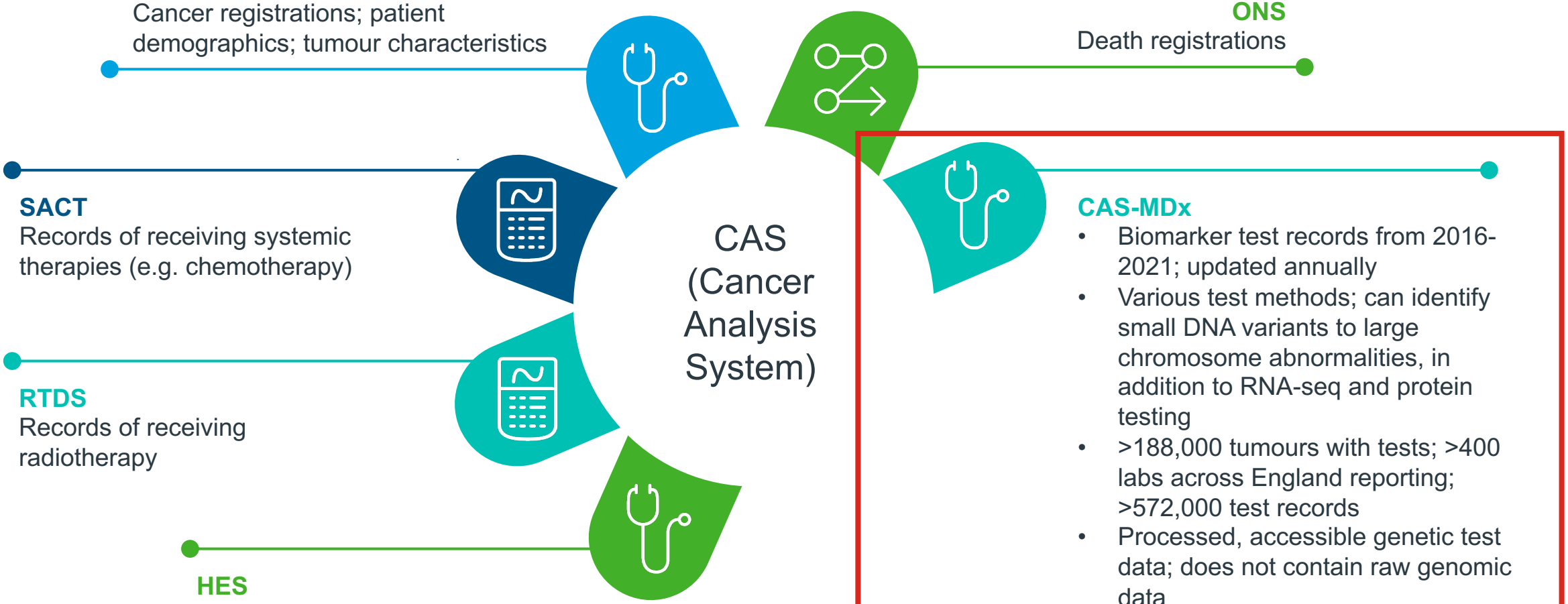
Targeted therapies for cancer are being developed increasingly quickly and enthusiasm for real-world biomarker data is matching this

- Clinical oncology traditionally makes treatment decisions based on tumour type or anatomical location, but is progressing to using **genetic biomarker data**
- An increasing number of oncologic agents are **tumour agnostic and instead targeted to a gene**
- Since 2020, ~90% of oncology drugs approved by the FDA were targeted therapies
- DNA sequencing & testing technologies are not new, but the pipeline to integrate these into routine care and to then have data available in real-world clinical databases is relatively slow
- Genetic data is mostly available for sampled cohorts, e.g. as part of specific studies/clinical trials, or in non-population-representative voluntary databases
- In England, **CAS** is an existing accessible real-world database based on cancer registry, and **CAS-MDx** is a new addition to CAS with genetic biomarker data

"Precision medicine is not the future of cancer care, it is the present."



CAS comprises of multiple linked databases, providing a rich and flexible repository of real-world cancer patient data



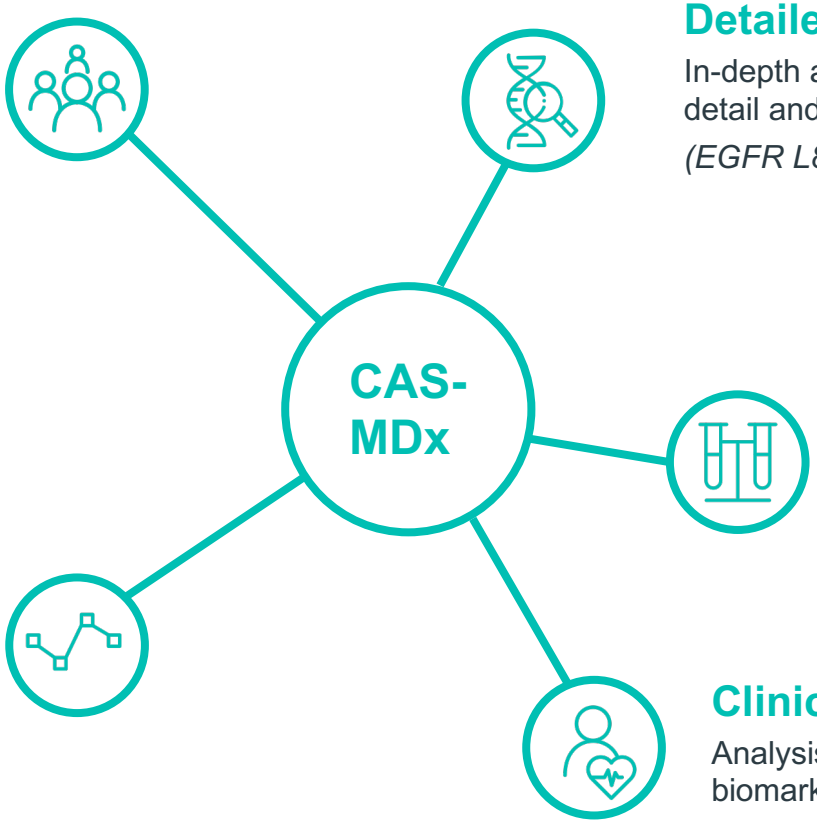
What can we do with CAS-MDx data?

CAS-MDx can generate real-world biomarker stratified outcomes analysis and insights into biomarker testing

Patient & tumour characteristics

Patient demographic and clinical characteristics of biomarker-tested cohorts to define sub-groups for analyses and examine mutation burden

(HR+ HER2+ status, TNM status, tumour grade, ethnicity, ECOG)



Detailed genetic data

In-depth analysis of sequence variant detail and clinical significance

(EGFR L858R, BRAFV600, IDH1 R132)

Genetic test results & testing patterns

Patient biomarker test results and specific types of genetic abnormalities can be described in detail for biomarkers of interest

(over-expressions, amplifications, mutations, etc.)

Patient pathways

Patient pathways, healthcare resource utilisation, and relative timing of biomarker testing, diagnosis and treatment can be analysed

Clinical outcomes

Analysis of key patient outcomes stratified by biomarker status, as well as comparisons/modelling of outcomes between patients with different biomarker status

The CAS data has been provided by patients and collected by the NHS as part of patient care and support. The data are collated, maintained and quality-assured by the National Disease Registration Service, which is part of NHS England. Access to this data was facilitated by the Simulacrum produced by Health Data Insight CIC with generous support from AstraZeneca and IQVIA.

The IQVIA-HDI collaborative pilot study with CAS-MDx set out to explore the new data

Pilot study overall aims:

- To describe biomarker testing in cohorts of cancer patients with commonly-occurring cancers in the UK (NSCLC, BrCa, CRC) with targeted therapies and explore patient demographics
- To explore available data in CAS-MDx to enable analysis of biomarker test dates, results, and stratified analysis of clinical outcomes by biomarker status
- To explore patient data linkage and investigate feasibility of key analysis methods and common pitfalls with the novel data

Examples of scenarios that were explored:

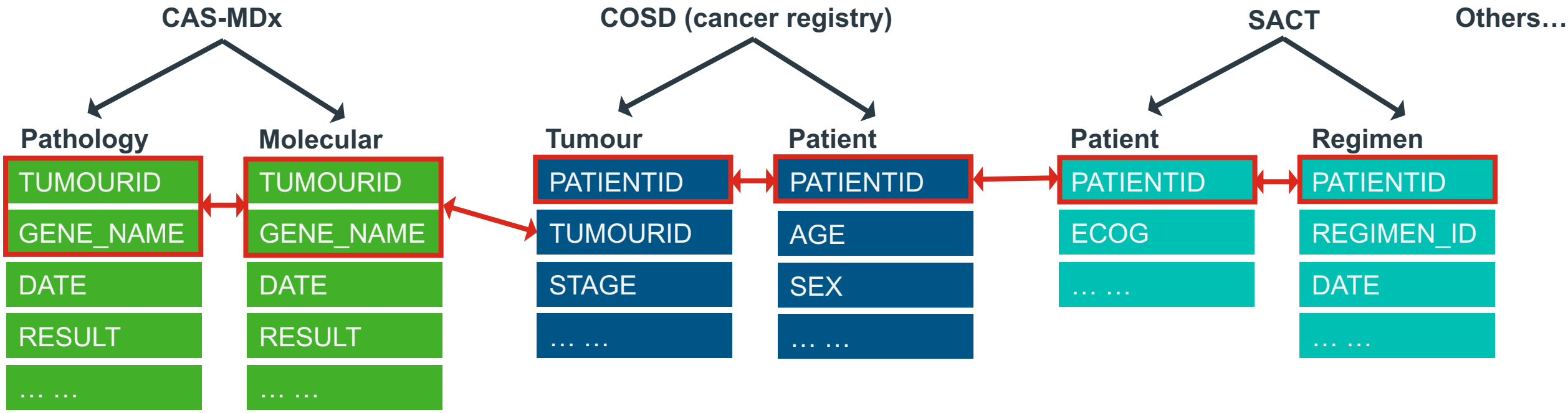
1) Identifying patient cohorts and accounting for multiple test records per patient
2) Preparing descriptive summary statistics of biomarker test data and biomarker tested patient cohorts
3) Stratified analysis of key clinical outcomes by patient biomarker status, to enable comparisons across groups

NSCLC = non-small cell lung cancer, BrCa = breast cancer, CRC = colorectal cancer

Scenario 1: identifying patient cohorts & test records

Pilot study aim: to identify patients & use linked data across multiple datatables within CAS, including the new CAS-MDx data

Linkage between CAS-MDx and other CAS datatables is via a tumour ID number



Recommendations:

- Patient cohort identification and records linkage is enabled in CAS via patient and tumour identifiers, but clear study design and patient criteria needs to be used to pre-process data for analysis
- Patient:tumour records are 1:many, and this needs accounted for
- Use of Simulacrum (simulated data to protect patient confidential data) is important to develop reliable analysis programs and check outputs, without directly accessing row-level patient data
- Knowledge of the data structure is essential

How is CAS-MDx data structured?

Data is recorded as overall lab results and aggregated to one record per tumour/gene combination

Tumour ID	Gene name	Overall test result date	Overall test result	Test method	Sequence variant	Fusions	CNV	Expression
1234	ERBB2 (HER2)	29-May-2018	Abnormal	Multiple	NA	NA	Abnormal	Normal
4242	BRCA1	16-Jan-2019	Abnormal	NGS	Abnormal	NA	NA	NA
4242	BRCA2	16-Jan-2019	Normal	NGS	Normal	NA	NA	NA
4242	ERBB2 (HER2)	24-Feb-2019	Normal	Multiple	NA	Failed	NA	Normal

Patient A has a NSCLC tumour (ID: 1234) and received three HER2 tests: first test for expression levels failed, second test showed 'normal' expression levels, and third test used a different method to show abnormal copy number variation

Patient B has a BrCa tumour (ID: 4242) which was tested for three biomarkers: BRCA1 (result = abnormal), BRCA2 (normal), and HER2 (tested twice via different methods; results were failed and normal)

Note that data and scenarios shown are hypothetical only, and not all variables in CAS-MDx are represented in the snapshot

How is CAS-MDx data structured?

Data is recorded as overall lab results and aggregated to one record per tumour/gene combination

Tumour ID	Gene name	Overall test result date	Overall test result	Test method	Sequence variant	Fusions	CNV	Expression
1234	ERBB2 (HER2)	29-May-2018	Abnormal	Multiple	NA	NA	Abnormal	Normal
4242	BRCA1	16-Jan-2019	Abnormal	NGS	Abnormal	NA	NA	NA
4242	BRCA2	16-Jan-2019	Normal	NGS	Normal	NA	NA	NA
4242	ERBB2 (HER2)	24-Feb-2019	Normal	Multiple	NA	Failed	NA	Normal

Considerations:

- The 'overall' status/date for a specific biomarker may not represent all the different types of mutations and may not be the first/last test for that patient*biomarker combination
- Some information may be lost in the aggregation step – e.g. specific details of multiple test methods, or repeated testing of the same biomarker using the same method
- However, some information is retained via additional variables in CAS-MDx, e.g. first/last test dates per biomarker

Note that data and scenarios shown are hypothetical only, and not all variables in CAS-MDx are represented in the snapshot

Scenario 2: describing biomarker test data

Pilot study aim: to use data across different parts of CAS to describe cancer patients in England with biomarker test records

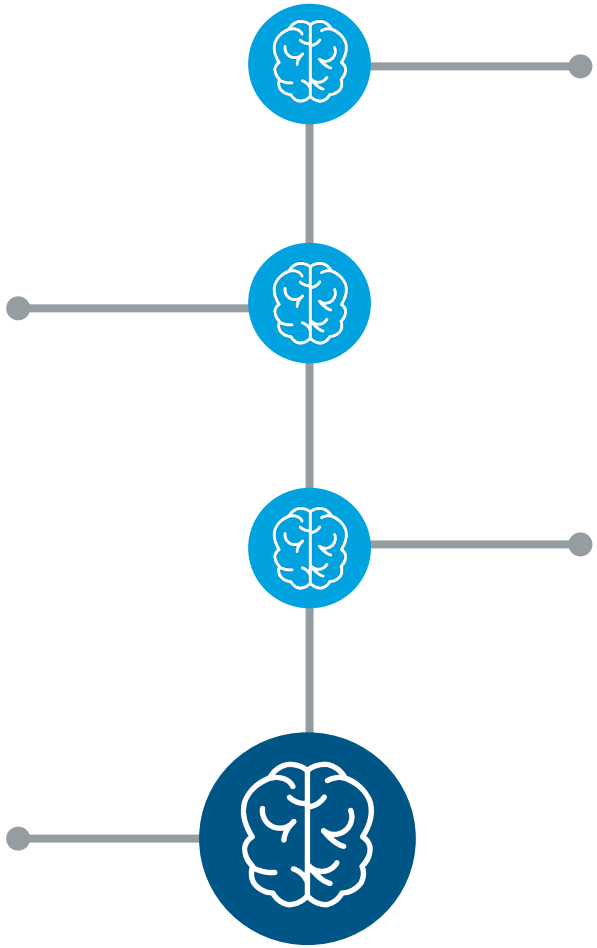
Variables describing the biomarker test are in CAS-MDx;the linked datasets contain variables to describe patients/outcomes

Linkage to treatments data

- SACT – drug names, dates of receipt, etc.
- Surgical procedures & dates
- Radiotherapy procedures & dates
- Treatment sequencing
- Treatment pathways integrated with biomarker tests

CAS-MDx

- Date of test
- Processed result of test (abnormal/normal/borderline)
- Method of test
- Type of genetic abnormality
- Hospital requesting test
- DNA sequence variants



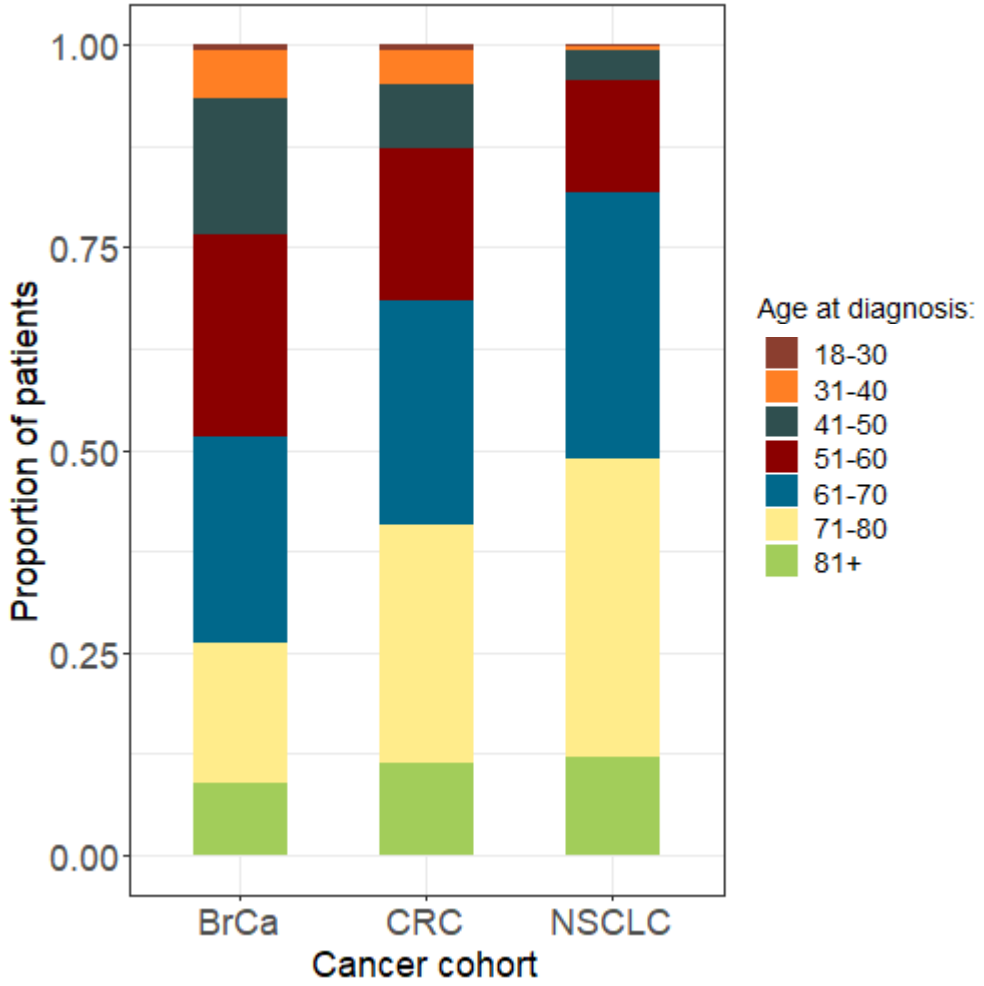
Linkage to secondary care and outcomes

- Death; date of death; cause of death
- HCRU event dates and types

Linkage to patient/tumour characteristics

- Age
- Sex
- Hospital of diagnosis
- Stage of tumour
- Tumour morphology
- Ethnicity
- Socioeconomic status
- ECOG

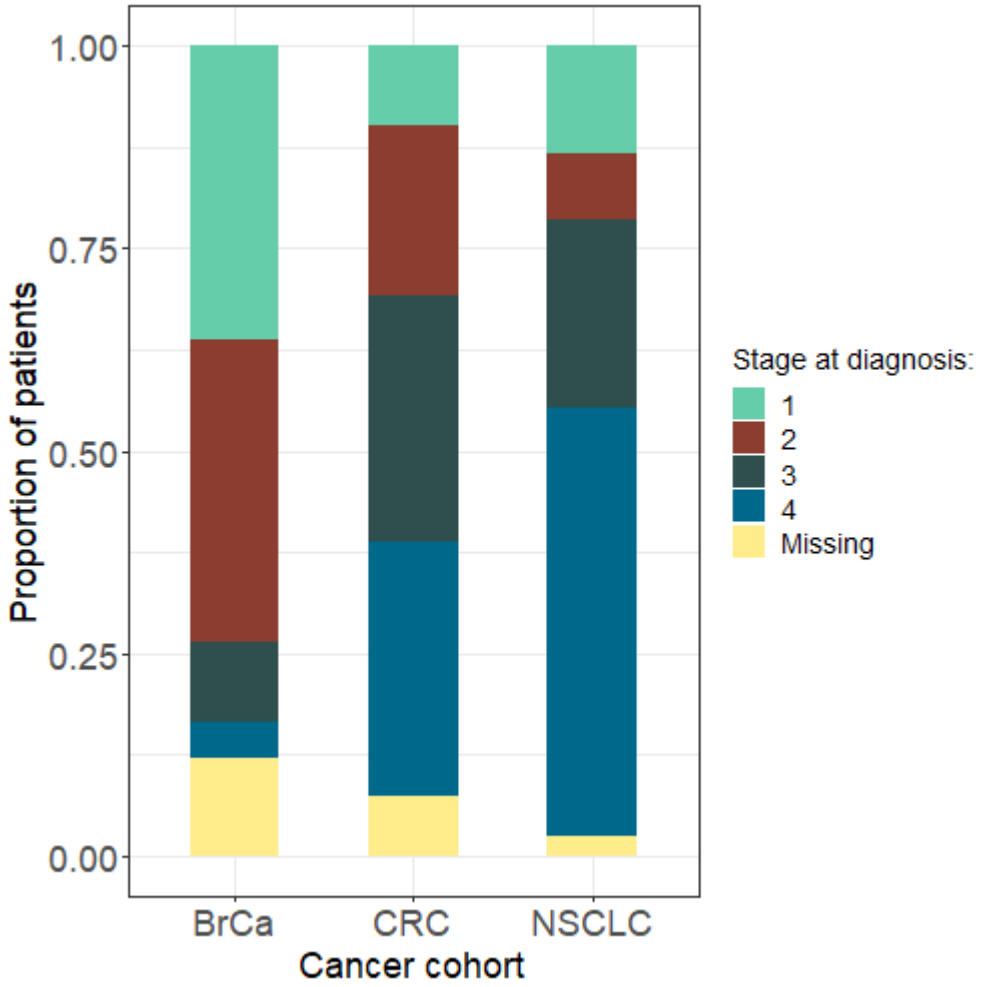
Distribution of ages at diagnosis in the pilot study CAS-MDx cohorts are broadly as expected for these cancers



- Biomarker-tested cancer patients were selected for each cohort
- Patient age at diagnosis was linked to the biomarker test record via unique anonymised identifiers in CAS
- Graph shows age distribution among each cancer patient cohort in the pilot study
- Most patients with biomarker test data are aged 51-80, in line with cancer incidence patterns
- Higher proportion of BrCa patients in the 41-50 age group compared to CRC/NSCLC, also in line with cancer incidence patterns
- Suggests representative sample of cancer patients with biomarker data available

NSCLC = non-small cell lung cancer, BrCa = breast cancer, CRC = colorectal cancer

Distribution of stage at diagnosis among the pilot study CAS-MDx cohorts varies a lot by cancer type



- Cancer stage at diagnosis was linked to the biomarker test record via unique anonymised identifiers in CAS
- Graph shows pilot study patient distribution across stages at diagnosis, with large variation between cancer types
- BrCa: ~75% of patients with a biomarker test are diagnosed at stage 1 or 2 (early stage)
- CRC: ~60% of patients with a biomarker test are diagnosed at stage 3 or 4 (advanced stage)
- NSCLC: ~80% of patients with a biomarker test are diagnosed at stage 3 or 4 (advanced stage)
- The patterns observed are aligned with what would be expected based on clinical care guidelines about biomarker testing in each cancer

NSCLC = non-small cell lung cancer, BrCa = breast cancer, CRC = colorectal cancer

The pilot study also carried out a deep-dive into mutational variant detail in CAS-MDx

Case study of N=30,110 NSCLC patients with EGFR biomarker test records

- >99% of patients with **abnormal EGFR** have **DNA sequence variants;** the remaining have copy number loss/gains
- Genetic locations and details of DNA sequence variants are available for N=2,340* patients
- We created mappings using ClinVar (a sequence variant online database) to report categories of 'Types of variant' and 'Clinical significance', as shown
- 50% of NSCLC patients with sequence variant data for EGFR had single nucleotide variants, with small numbers of other types of variant
- 50% of NSCLC patients with sequence variant data for EGFR had a variant associated with drug response

		N	%
	NSCLC cohort size	30110	100%
Total patients tested	N	30110	100%
Patients with sequence variant identified	N	3410	10%
Type of EGFR sequence variant	Single nucleotide variant	1660	50%
	Deletion	190	5%
	Duplication	30	0%
	Indel	60	0%
	Insertion	*	*
	Unmapped	260	10%
Clinical significance of EGFR sequence variant	Benign	*	*
	Pathogenic	340	10%
	Drug response	1680	50%
	Conflicting interpretations	*	*
	Uncertain significance	40	0%
	Unmapped	260	10%

* Notes: rounding applies to initial results outputs from CAS; some sequence variant data were not usable because of free text errors

Scenario 3: stratifying outcomes analysis by biomarker status

Pilot study aim: to explore methods for biomarker-stratified analyses of patient outcomes, to enable comparisons

A key part of many targeted therapy research studies will be analysis of clinical outcomes stratified by biomarker status

e.g. the pilot study aims to describe overall survival of cancer patients stratified by biomarker status, for biomarkers of known clinical significance per cancer type

- Overall survival was estimated from date of diagnosis until date of (all-cause) death or censoring, using a Kaplan-Meier analysis approach
- Key question: what is the best method to stratify these analyses appropriately?

If there was only one biomarker of interest:

Tumour ID	HER2 test result	Stratification group
5678	Normal	Normal
4242	Abnormal	Abnormal
1234	Failed	Unknown
0101	Borderline	Borderline

Note this applies equally to studies where one specific type of mutation, e.g. CNV, sequence variants, etc. is of interest

Analysis stratification will be highly specific to the research questions and patient cohort of interest

Tumour ID	HER2 test result	PIK3CA test result	PD-L1 test result	Any/none stratification	High-specificity stratification	Overlapping cohorts	
5678	Normal	Abnormal	Normal	Abnormal	Abnormal PIK3CA	Abnormal PIK3CA	
4242	Abnormal	-	-	Abnormal	Abnormal HER2	Abnormal HER2	
1234	Failed	Normal	-	Normal	Normal/Unknown	Normal/Unknown	
0101	Borderline	Abnormal	-	Abnormal	Abnormal PIK3CA	Abnormal PIK3CA	
4321	Normal	Normal	Borderline	Normal	Normal	Normal/Unknown	
8765	Normal	Failed	Normal	Normal	Normal/Unknown	Normal/Unknown	
9000	Normal	Abnormal	Abnormal	Abnormal	Multiple abnormalities	Abnormal PIK3CA	Abnormal PD-L1
1357	-	-	Failed	Unknown	Unknown	Unknown	
2468	Abnormal	-	Abnormal	Abnormal	Multiple abnormalities	Abnormal HER2	Abnormal PD-L1
1010	Borderline	-	Failed	Borderline	Borderline	Normal/Unknown	

Good degree of accuracy; non-mutually exclusive cohorts, so comparisons are limited

Applies a kind of hierarchy; gives precedence to available/definitive results; possibly oversimplifies by combining information across genes

Good degree of accuracy; complicated comparisons; potential small sample sizes in groups

Many biomarkers are tested in ‘panels’, and co-occurring abnormalities can occur

Within a limited set of biomarkers of interest per cancer type:

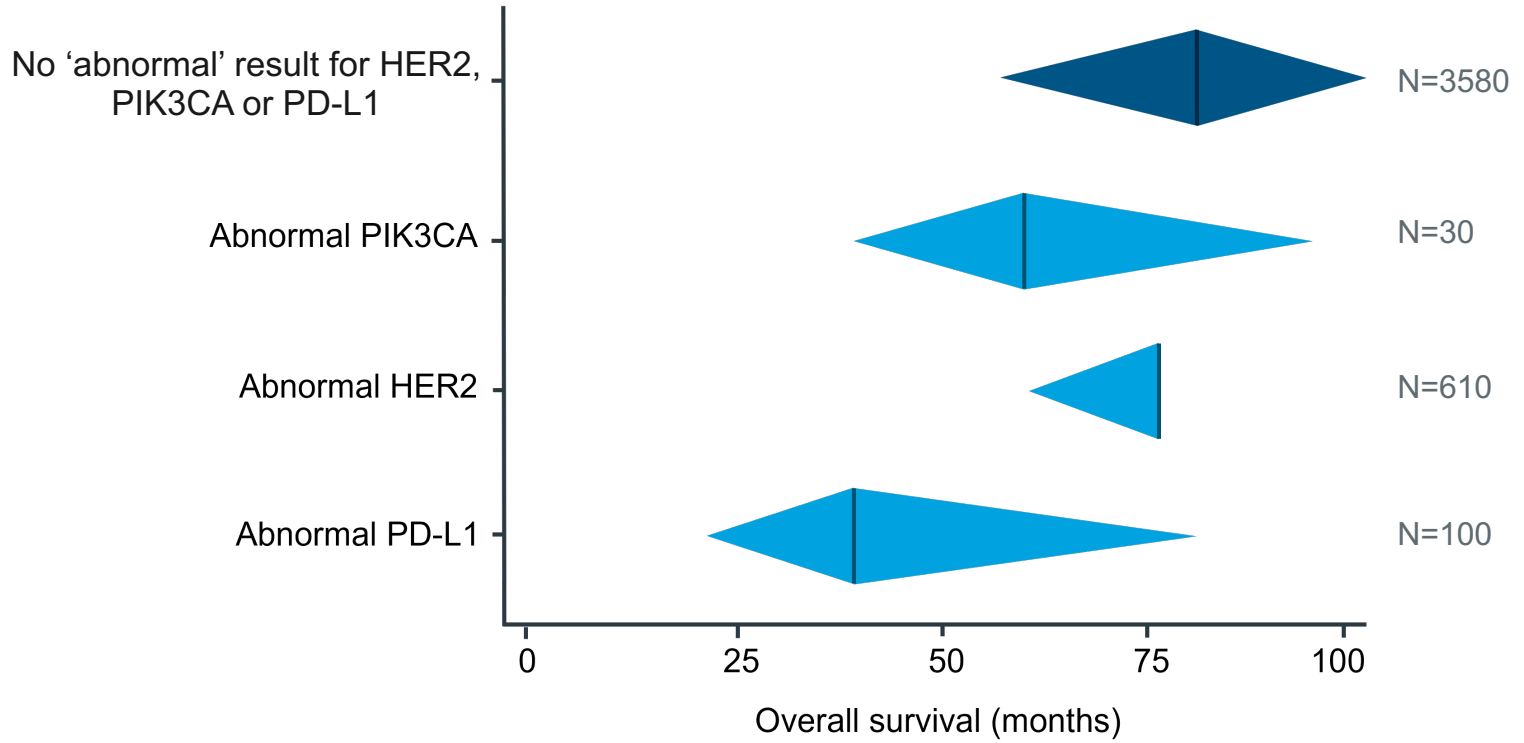
- **25% of BrCa patients, 60% of NSCLC patients** and **95% of colorectal cancer patients** in the pilot study had more than one biomarker (out of those included in the study) tested for their tumour
- Some biomarkers are tested in panels, and one tumour can tested multiple times, so this is not surprising

Co-occurring mutations:

- Approximately **10% of the colorectal cancer patients** had more than one biomarker (out of those included in the study) with an abnormal result
- Small proportions of both NSCLC and BrCa patients also had evidence of co-occurring mutations (out of those biomarkers included in the study)
- Multiple mutations can co-occur even in the same patient/tumour

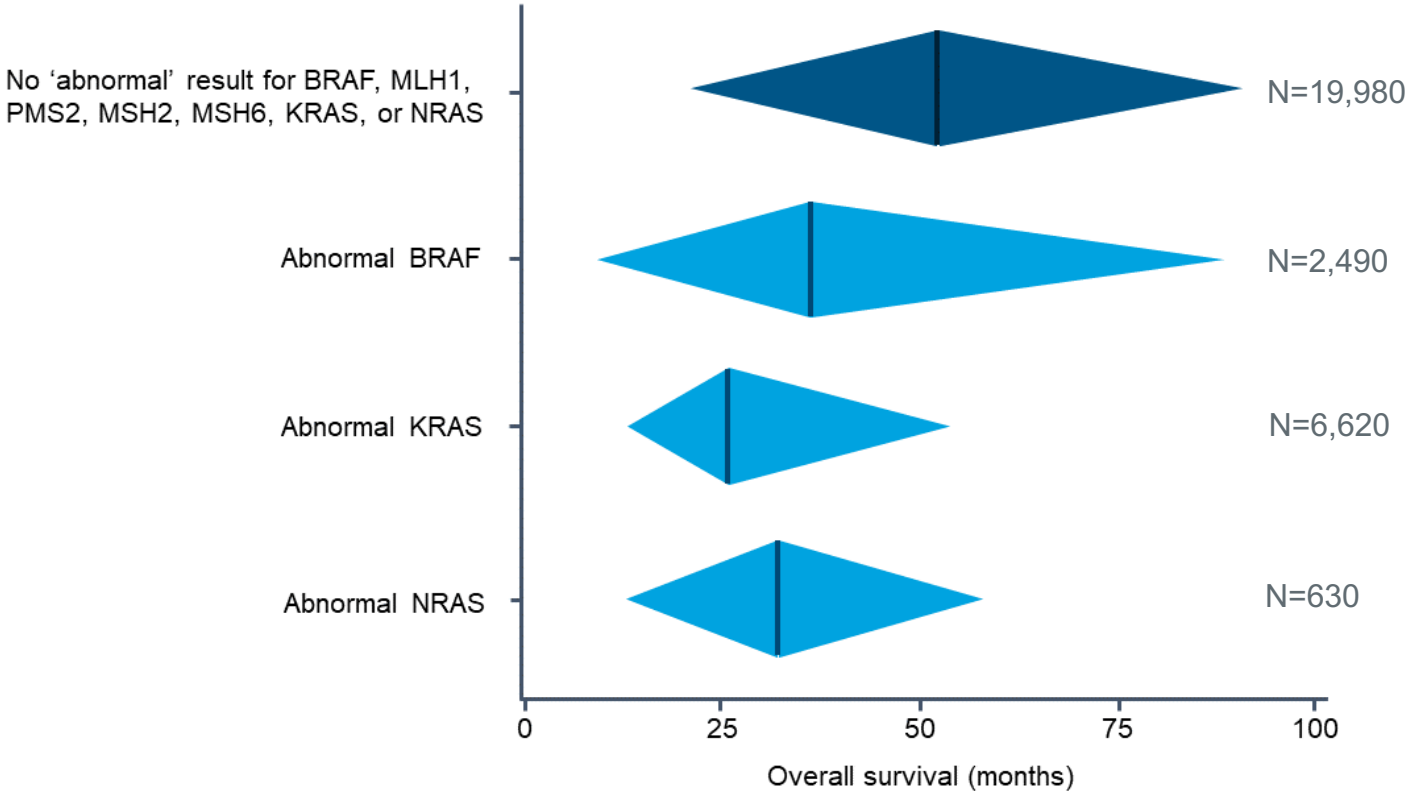
BrCa patients with abnormal PIK3CA, PD-L1, or HER2 tend to have shorter median survival than those without known abnormalities

- Box plot shows median unadjusted overall survival from diagnosis for biomarker-tested patients, with IQR (note upper quartile not reached for ‘abnormal HER2’ cohort)
- A patient can have abnormalities for multiple biomarkers, therefore can be included in more than one ‘abnormal’ subgroup
- BrCa patients with abnormal test results tend to have shorter median OS than patients with no known abnormalities



CRC patients with abnormal BRAF, KRAS, or NRAS tend to have shorter median survival than those without known abnormalities

- Box plot shows median unadjusted overall survival from diagnosis (and IQR)
- Note that median OS was not reached for patient subgroups with either abnormal MLH1, PMS2, MSH2 or MSH6, so these are not shown
- A patient can have abnormalities for multiple biomarkers, therefore can be included in more than one ‘abnormal’ subgroup
- CRC patients with abnormal test results for BRAF, KRAS or NRAS tend to have slightly shorter median OS than patients with no known abnormalities



Summary of lessons learned

The pilot study highlighted some key ‘lessons learned’ for working with CAS-MDx in future

1 Novel genetic RWD source

- Easily accessible genetic real-world data
- Genetic insights gain significance amid increasing availability of targeted therapies

2 Population-level genetic data

- CAS has ~99% of English cancer patients
- Pathology tests in CAS-MDx are population-representative
- Molecular lab tests cover large areas & coverage will increase in future data refreshes

3 Flexibility of linked data

- Reliable linkage between databases is achieved with unique patient identifiers
- Understanding of database structure is essential to reliably use patient-level data

4 Multiple test records per patient

- Multiple tumours per patient; and multiple tests
- Unique records per tumour ID*gene name combination
- Careful study design needed to select appropriate records

5 Structured genetic data

- Genetic test results in pre-processed structured format
- Bioinformatics skills not required, but genetics knowledge beneficial for data processing and interpretation

6 Multiple genetic abnormalities can co-occur

- Tumours can have multiple co-occurring mutations
- Stratified analyses require careful definition of mutually-exclusive groups
- Overlapping sub-cohorts need to be carefully analysed.

Thank you for listening!

Any questions?
fiona.ingleby@iqvia.com

The CAS data has been provided by patients and collected by the NHS as part of patient care and support. The data are collated, maintained and quality-assured by the National Disease Registration Service, which is part of NHS England. Access to this data was facilitated by the Simulacrum produced by Health Data Insight CIC with generous support from AstraZeneca and IQVIA.