

**Utilising the power of SNOMED CT
to evidence the impact of the pandemic
on cancer patients**

Monica Jones
Chief Data Officer – University of Leeds
Associate Director & National Strategic Lead - HDRUK



MISSION: to unite the UK's health data to enable discoveries that improve people's lives

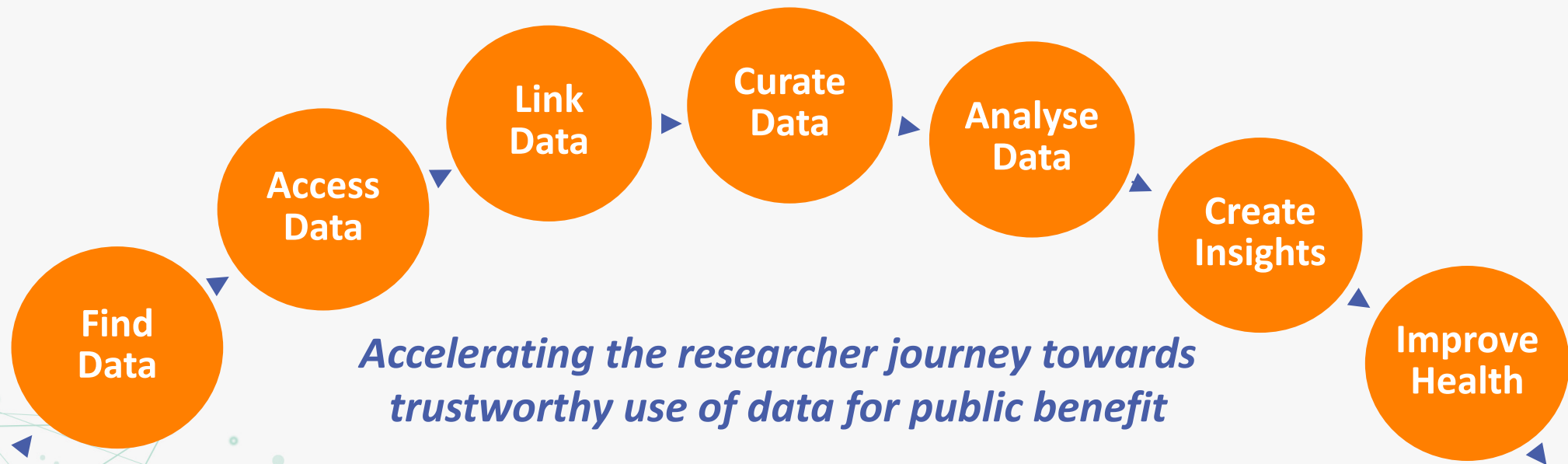
VISION: for large-scale data and advanced analytics to benefit every patient interaction, clinical trial and biomedical discovery, and to enhance public health

What are we about?

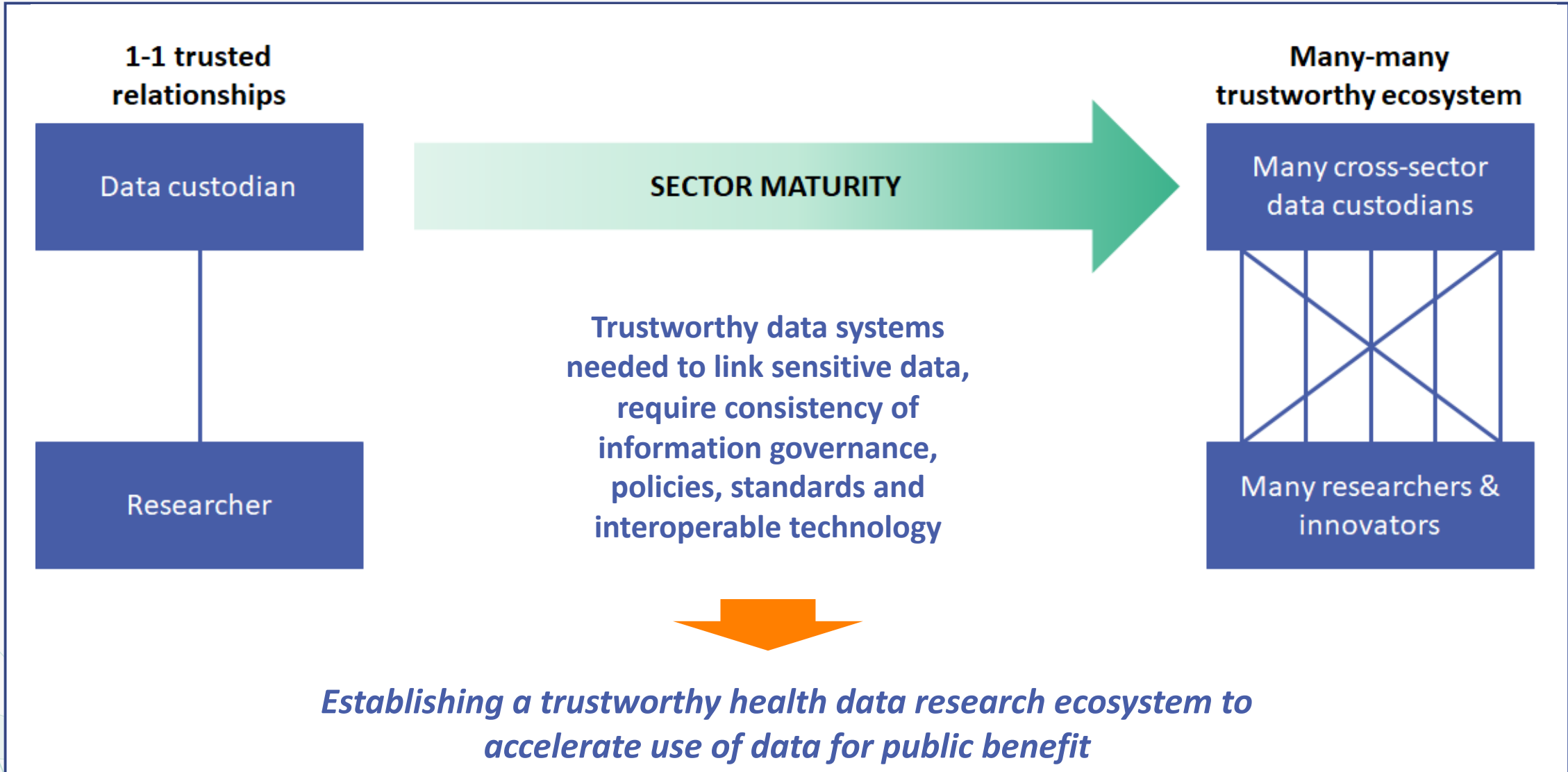
Improve health and boost UK science by making it easier for researchers to find, access and use diverse, high quality data

Provide leadership to fix difficult technical problems, by creating innovative solutions needed for researchers to use large-scale data safely and securely

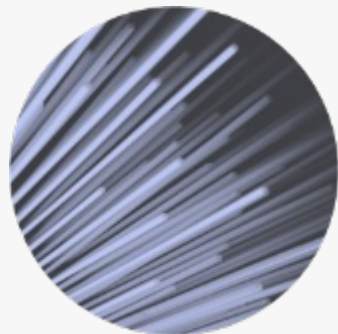
Accelerate & streamline health data science by developing open collaborations that connect data, people and organisations across the UK and internationally



The problem we are trying to solve



How we do this ?



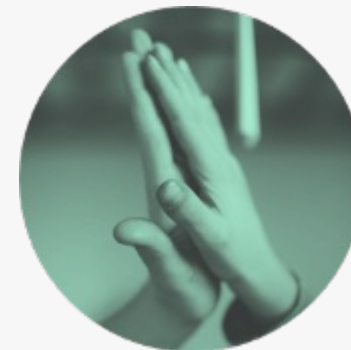
1. Accelerate Trustworthy Data Use

By implementing a national research data strategy and assembling infrastructure and services aligned to research and innovation needs



2. Empower Researchers

By valuing people with diverse perspectives & skills committed to open and team science to advance scientific discoveries and deliver patient and public benefit

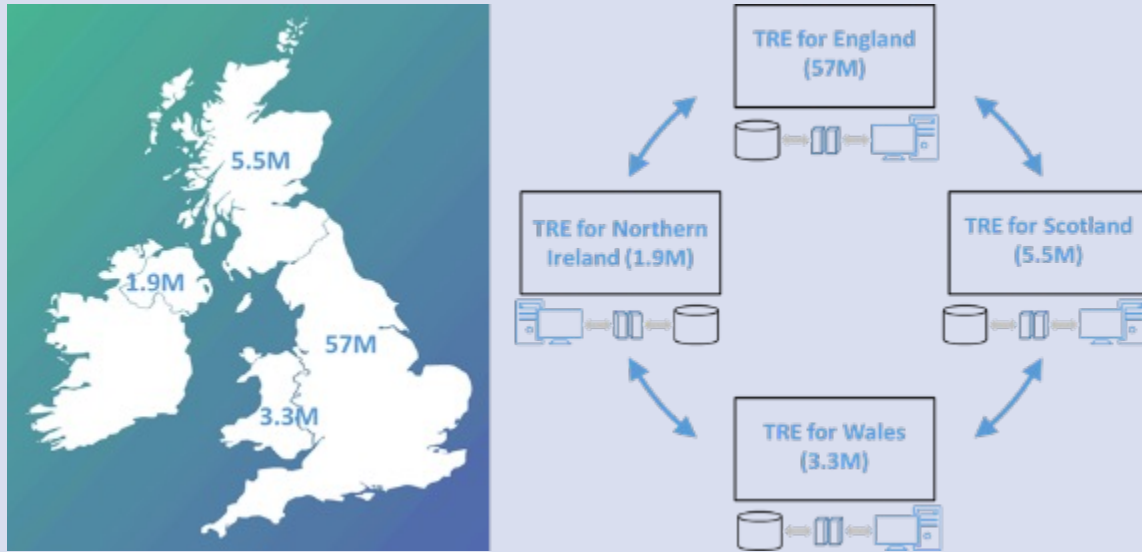


3. Promote Partnerships

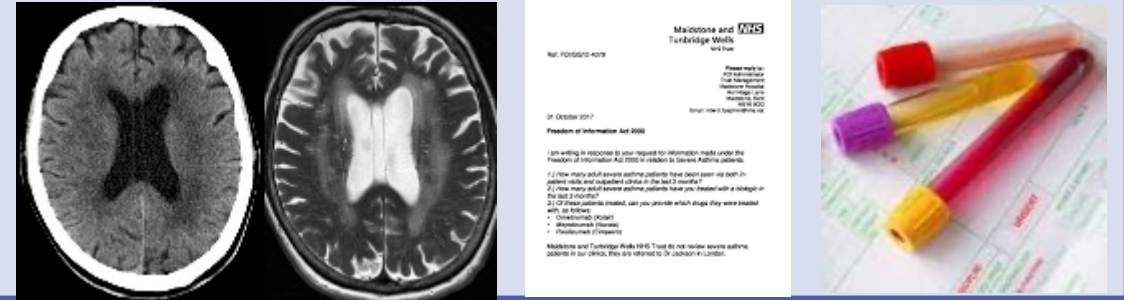
By building and maintaining critical partnerships, aligning incentives and reducing complexity across a fragmented landscape to streamline health data science

1. Accelerate Trustworthy Data Use - through a national data strategy

NHS: NATIONALLY COLLATED, POPULATION-WIDE DATA



NHS: REGIONAL DATA



NON-NHS: HEALTH-RELEVANT ADMINISTRATIVE DATA

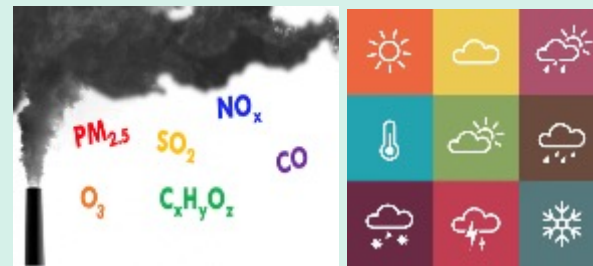
Census, education, disability, income....



PERSONAL MONITORING DATA



ENVIRONMENTAL DATA



RESEARCH-GENERATED DATA

Longitudinal cohorts & clinical trials, e.g.



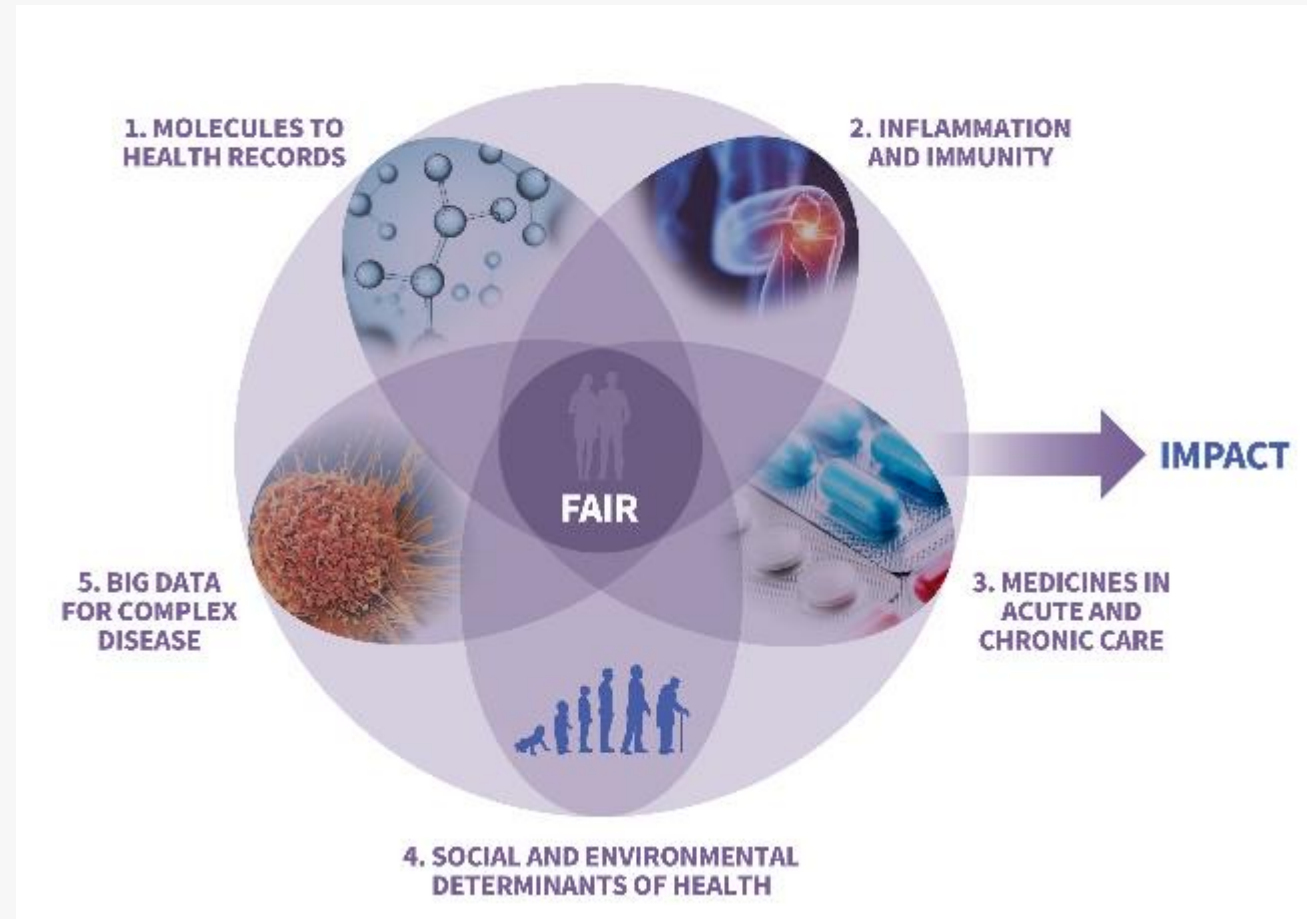


2. Empower Researchers

Valuing people with diverse perspectives committed to **open and team science** to **advance scientific discovery** and deliver **patient and public benefit**.

Driver Programmes

- Driving the design and delivery of a world-leading health data infrastructure
- Generate UK-wide 'research-ready' data foundations for wider re-use
- Deliver health & care impact





3. Promote Partnerships

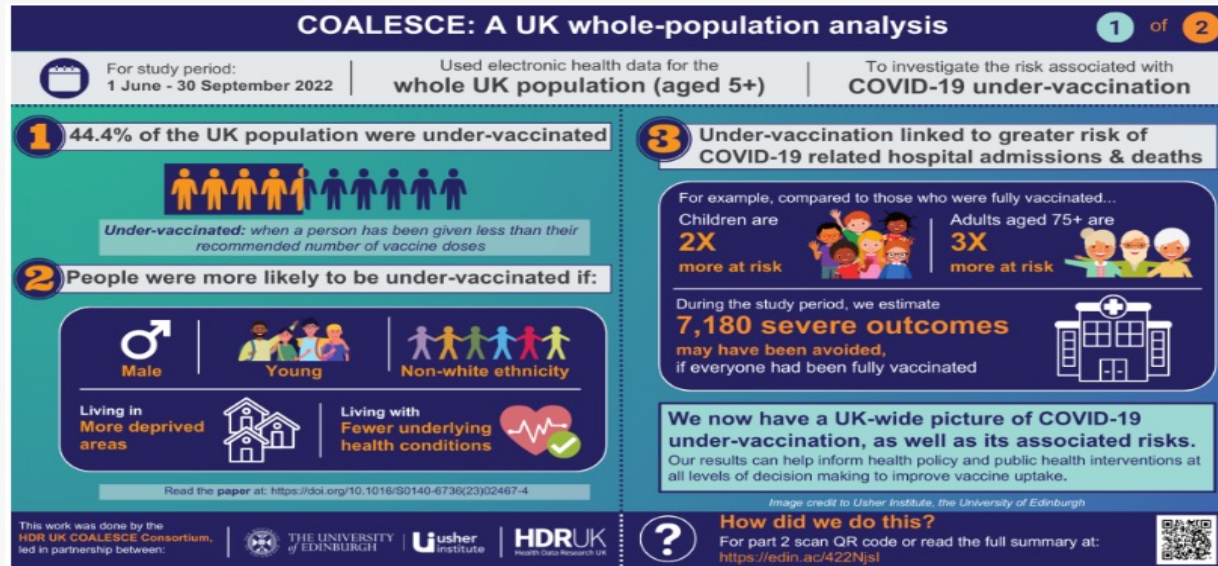
Aligning incentives and reducing complexity across a fragmented landscape to streamline health data science



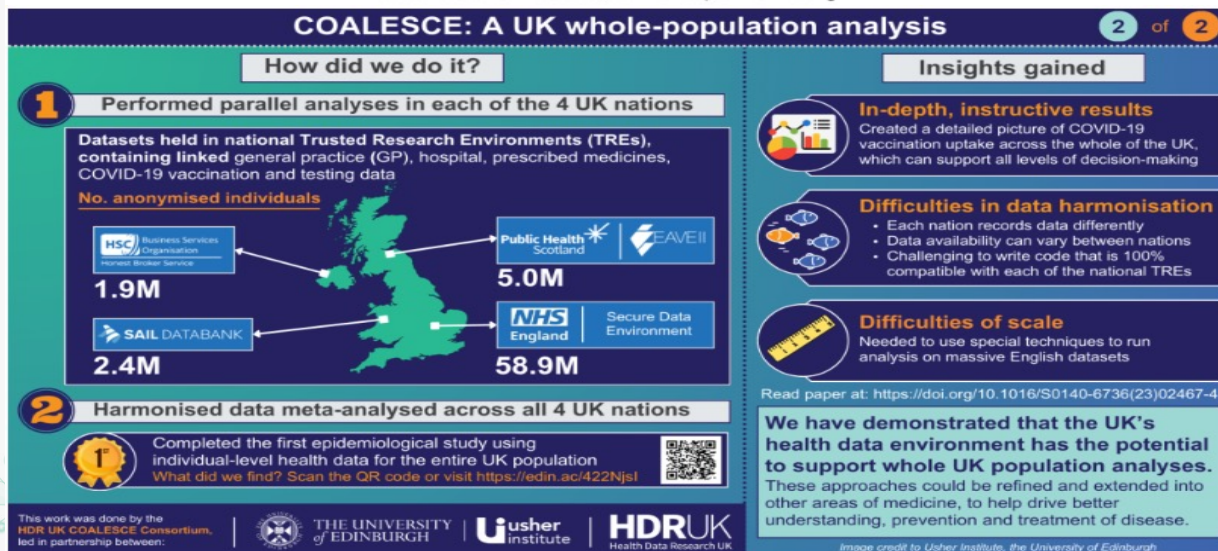
3. Promote Partnerships

Aligning incentives and reducing complexity across a fragmented landscape to streamline health data science

UK population study reveals impact of COVID-19 under-vaccination



Credit: Usher Institute, University of Edinburgh*



Published in [The Lancet](#), the [COALESCE study](#) suggests that more than 7,000 hospitalisations and deaths might have been averted in summer 2022 if the UK had had better vaccine coverage.

With COVID-19 cases on the rise and a new variant strain recently identified, this research provides a timely insight into vaccine uptake and hesitancy which could inform policy-makers.

This study marks a significant milestone in HDR UK's mission to **unite the UK's health data to enable discoveries that improve people's lives** and demonstrates the value and potential of population-wide health data studies.

Data Standards and Quality

The Data Improvement work supports researchers and custodians across multiple areas of the “FAIR” approach (Findable, Accessible, Interoperable, Reusable).

Across the institute, HDR UK has made great progress in ensuring data is **F**indable and **A**ccessible, the focus now in Data Improvement will be on make data **I**nteroperable and **R**eusable.

Standards and data quality improvements will support research taking place in a federated manner across data custodians, as there must be consistency in data standards and use of data elements for analysis across multiple datasets.

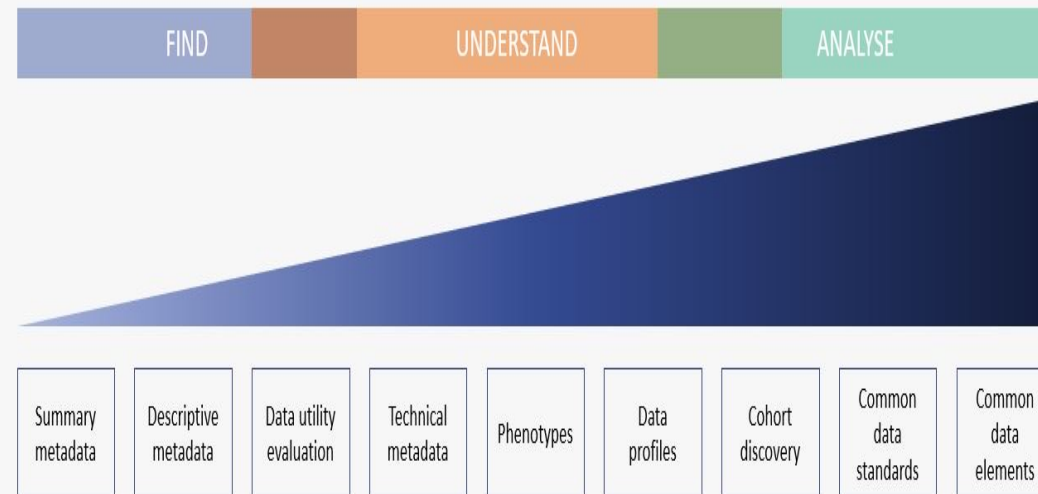


Fig 1. Metadata maturity gradient - information availability for a dataset.

Data Utility Framework - Development

Objectives The value of healthcare data is being increasingly recognised, including the need to improve health dataset utility. There was no established mechanism for evaluating healthcare dataset utility making it difficult to evaluate the effectiveness of activities improving the data.

To describe the method for generating and involving the user community in developing a proposed framework for evaluation and communication of healthcare dataset utility for given research areas.

Conclusion The process resulted in a user centred designed framework for objectively evaluating the likely utility of specific healthcare datasets, and therefore, of value both for potential users of health data, and for data custodians to identify the areas to provide the optimal value for data curation investment.

Category	Dimension	Definition	Bronze	Silver	Gold	Platinum
Data Documentation	Documentation Completeness	Proportion of metadata (as in the current metadata specification) which is available in the expected format	This element will be calculated automatically based on the level of metadata available on the Gateway, and values set for each category			
	Availability of additional documentation and support	Available documentation in addition to the data dictionary	Past journal articles demonstrate that knowledge of the data exists	Comprehensive ReadMe describing extracting and use of data, Dataset FACS available, Visual data model provided	As Silver, plus dataset publication was supported with a journal article explaining the dataset in detail, or dataset training materials	As Gold, plus support personnel available to answer questions
	Data Model	Availability of clear, documented data model	Known and accepted data model but some key field un-coded or free text	Key fields codified using a local standard	Key fields codified using a national or international standard	Data Model conforms to a national standard and key fields codified using a national / international standard
	Data Dictionary	Provided documented data dictionary and terminologies	Data definitions available	Definitions compiled into local data dictionary which is available online	Dictionary relates to national definitions	Dictionary is based on international standards and includes mapping
Technical Quality	Provenance	Clear description of source and history of the dataset, providing a "transparent data pipeline"	Source of the dataset is documented	Source of the dataset and any transformations, rules and exclusions documented	All original data items listed, including all transformations and exclusion listed and impact of these	Ability to view earlier versions, including transformations before any transformations have been applied data (in line with deidentification and information governance approval) and review the impact of each stage of data cleaning
	Data Quality Management Process	The level of maturity of the data quality management processes	A documented data management plan covering collection, auditing, and management is available for the dataset	Evidence that the data management plan has been implemented is available		Externally verified compliance with the data management plan, e.g. by International Organization for Standardisation (ISO), Care Quality Commission (CQC), Information Commissioner's Office (ICO) or other body
Coverage	Data Management Association (DAMA) Quality Dimensions	Technical data quality dimensions: Completeness, Uniqueness, Accuracy, Validity, Timeliness and Consistency	These elements will be calculated with data profiling tools, and the category breakdown evaluated following further data collection			
	Pathway coverage	Representation of multi-disciplinary healthcare data	Contains data from a single speciality or area	Contains data from multiple specialities or services within a single tier of care	Contains multimodal data or data that is linked across two tiers (e.g. primary and secondary care)	Contains data across more than two tiers
Access & Provision	Length of follow up	Average timeframe in which a patient appears in a dataset (follow up period)	Between 1 - 6 months	Between 6 - 12 months	Between 1 - 10 years	More than 10 years
	Allowable uses	Allowable dataset usages as per the licencing agreement, following ethical and information governance approval	Available for specific academic research uses only	Available for academic and non-profit (e.g. charity, public sector) uses only	Available for limited commercial uses (e.g. relating to a specific domain), in addition to academic and other non-commercial uses	Available for wider commercial uses (in line with ethical and information governance approval), and addition to academic and other non-commercial uses
	Time Lag	Lag between the data being collected and added to the dataset	Approximately 1 year	Approximately 1 month	Approximately 1 week	Effectively real-time data
Value & Interest	Timeliness	Average data access request timeframe	Less than 6 months	Less than 3 months	Less than 1 month	Less than 2 weeks
	Linkages	Ability to link with other datasets	Identifiers to demonstrate ability to link to other datasets	Available linkages outlined and/or List of datasets previously successfully linked provided	List of restrictions on the type of linkages detailed. List of previously successful dataset linkages performed, with navigable links to linked datasets via a Digital Object Identifier (DOI) or Uniform Resource Locator (URL)	Existing linkage with reusable or downstream approvals
Value & Interest	Data Enrichments	Data sources enriched with annotations, image labels, phenomes, derivations, Natural Language Processing (NLP) derived data labels	The data include derived fields, or enriched data.	The data include additional derived fields, or enriched data used by other available data sources.	The derived fields or additional data were generated from, or used by, a peer reviewed algorithm.	The data includes derived fields or enriched data from a national report.

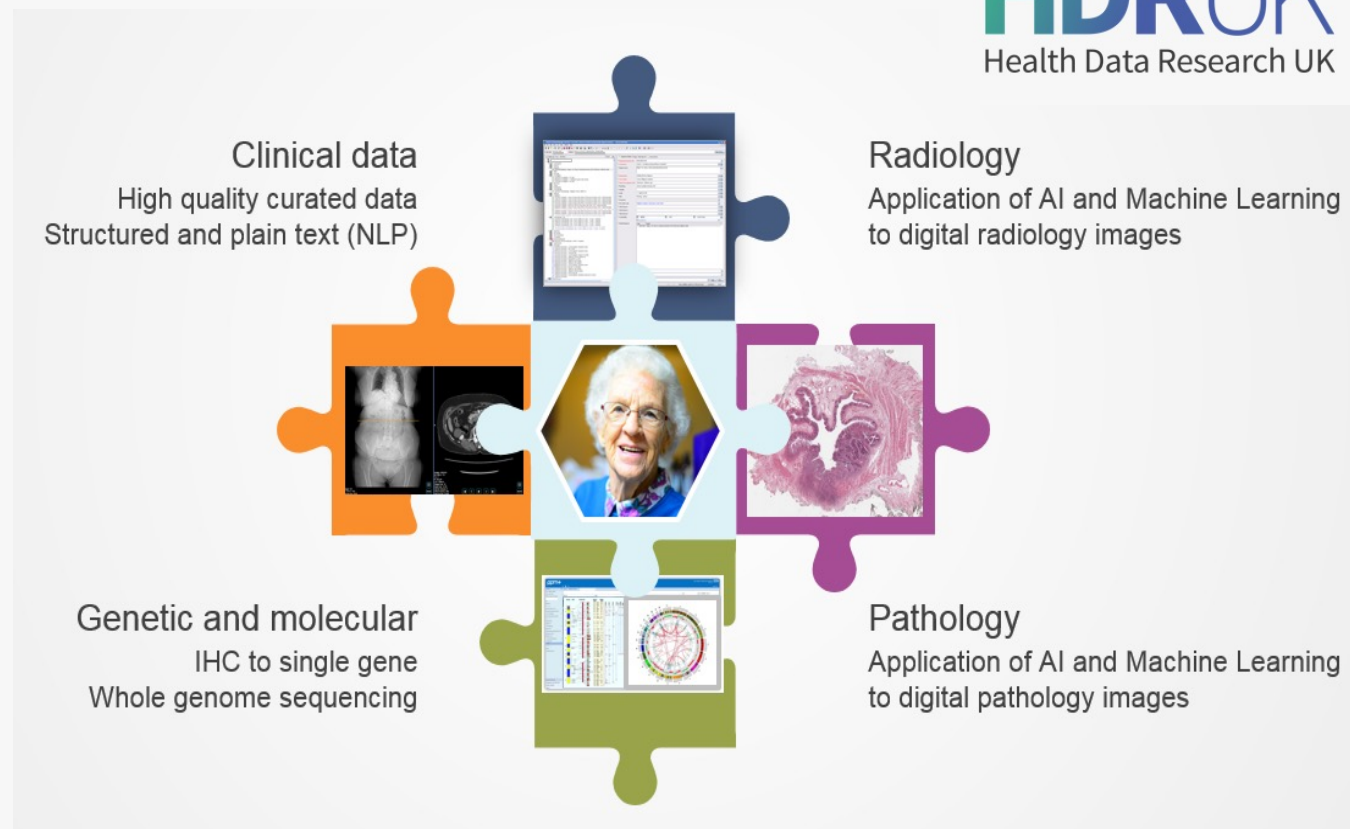
[Development of a data utility framework to support effective health data curation | BMJ Health & Care Informatics](#)

Ben Gordon, Monica Jones, Neil Sebire et al. *BMJ Health Care Informatics* 2021;28:e100303

© Author(s) (or their employer(s)) 2021. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

DATA-CAN: The vision

- **Improve access to existing data**
 - Clinical, academic & commercial
- **Improve quality of data**
 - Better data at the point of care
 - Enhanced curation of current data
- **Improve UK coverage**
 - Four nations of the UK
- **Add new datasets at scale**
 - Genetic/molecular, imaging, pathology
 - Primary care, PROMs



Ensure fair value

NHS & UK
Trusts & CCGs
Patients & public

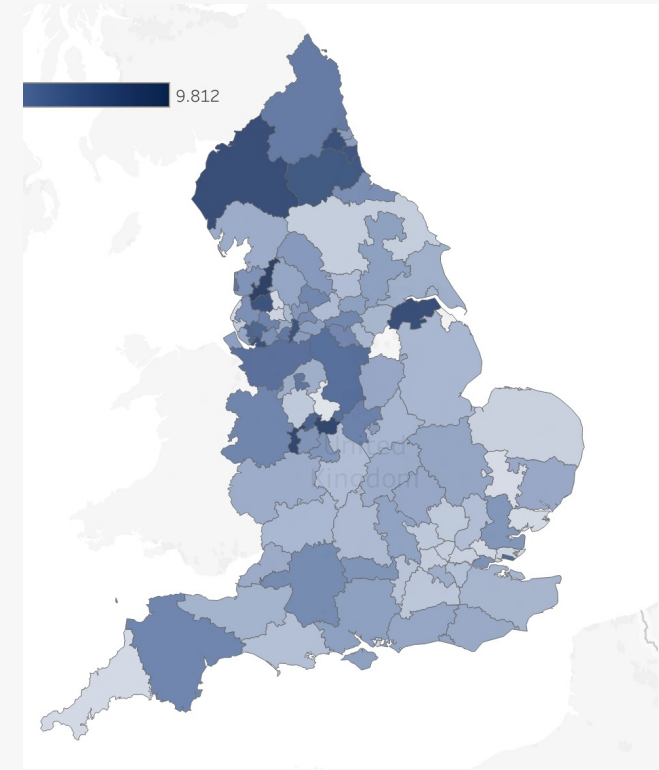
The impact of COVID-19 on cancer, using open standards such as SNOMED CT, in both local and national cancer data



Covid-19 was shown to impact cancer referrals and have a significant impact on projected excess deaths. ¹



Access to the National SDE holding over **12 billion** rows of data and more than **50 million** patients in England

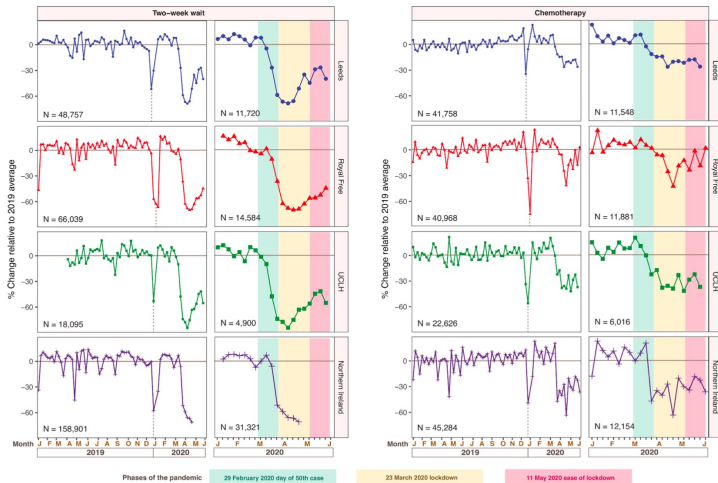


Leeds

Royal Free

UCLH

N.I.



databricks



DELTA LAKE

1. Lai AG, Pasea L, Banerjee A, Jones M, et al. Estimated impact of the COVID-19 pandemic on cancer services and excess 1-year mortality in people with cancer and multimorbidity: near real-time data on cancer care, cancer deaths and a population-based cohort study. *BMJ Open* 2020;10:e043828

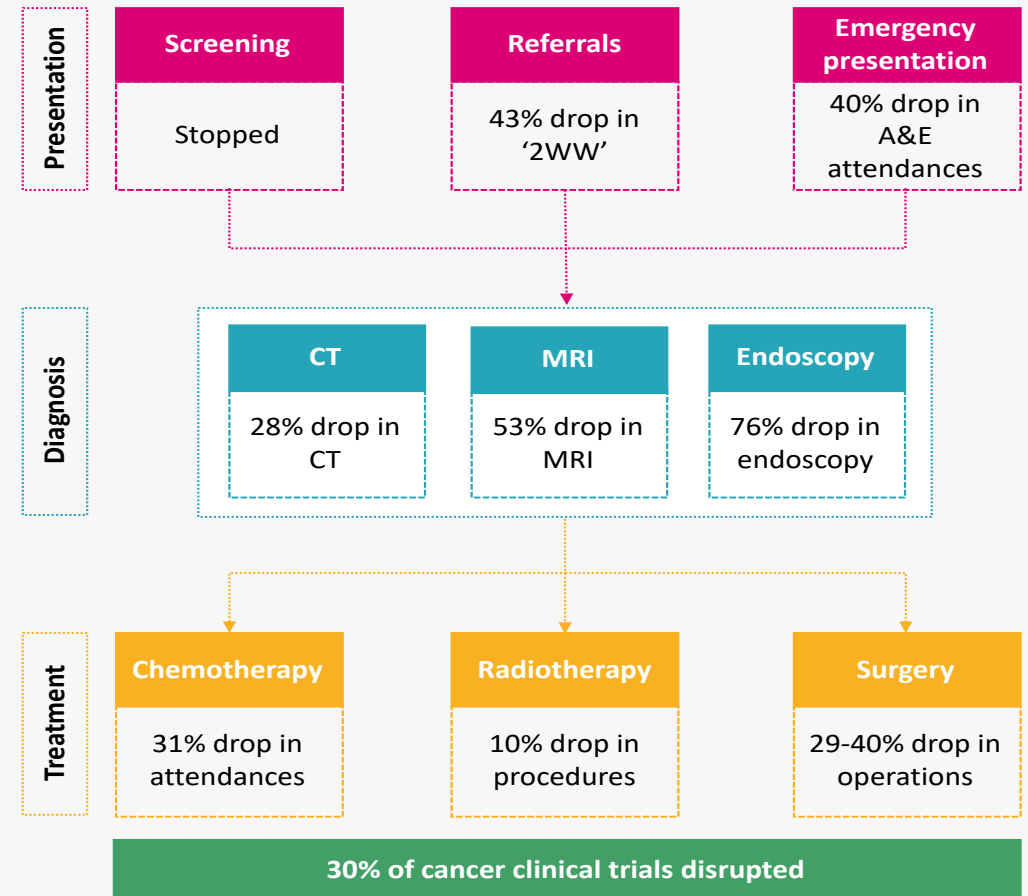
COVID-19 and Cancer: “Real Time” Data analysis

- **DATA-CAN** researchers analysed data from Hospital Trusts across the UK
- Looked at two measures to determine the effect of the pandemic on cancer services:
 - **2 Week Waiting** Time for Cancer Referrals (early warning system for suspicion of cancer) - **7 out of 10 people with suspicion of cancer were not getting referred to cancer specialist services**
 - **Chemotherapy attendances** (proxy measure of the “health” of the cancer treatment service) - **4 out of 10 cancer patients were not getting access to their chemotherapy**
- **First data that drew attention of the government, academia, NHS and the public to the disastrous effect of COVID-19 on cancer services and cancer patients**

Lai AG, ... Jones M, Lawler M, Hemingway H. [Estimated impact of the COVID-19 pandemic on cancer services and excess 1-year mortality in people with cancer and multimorbidity: near real-time data on cancer care, cancer deaths and a population-based cohort study.](#) *BMJ Open.* 2020;

Disruptions to cancer services across the pathway due to COVID-19

- **Disruption was felt across the entire cancer pathway**
- **Presentational Delay, Diagnostic Delay and significant impact on treatment**
- **Significant disruption of cancer research (clinical trials, discovery research, translational research)**

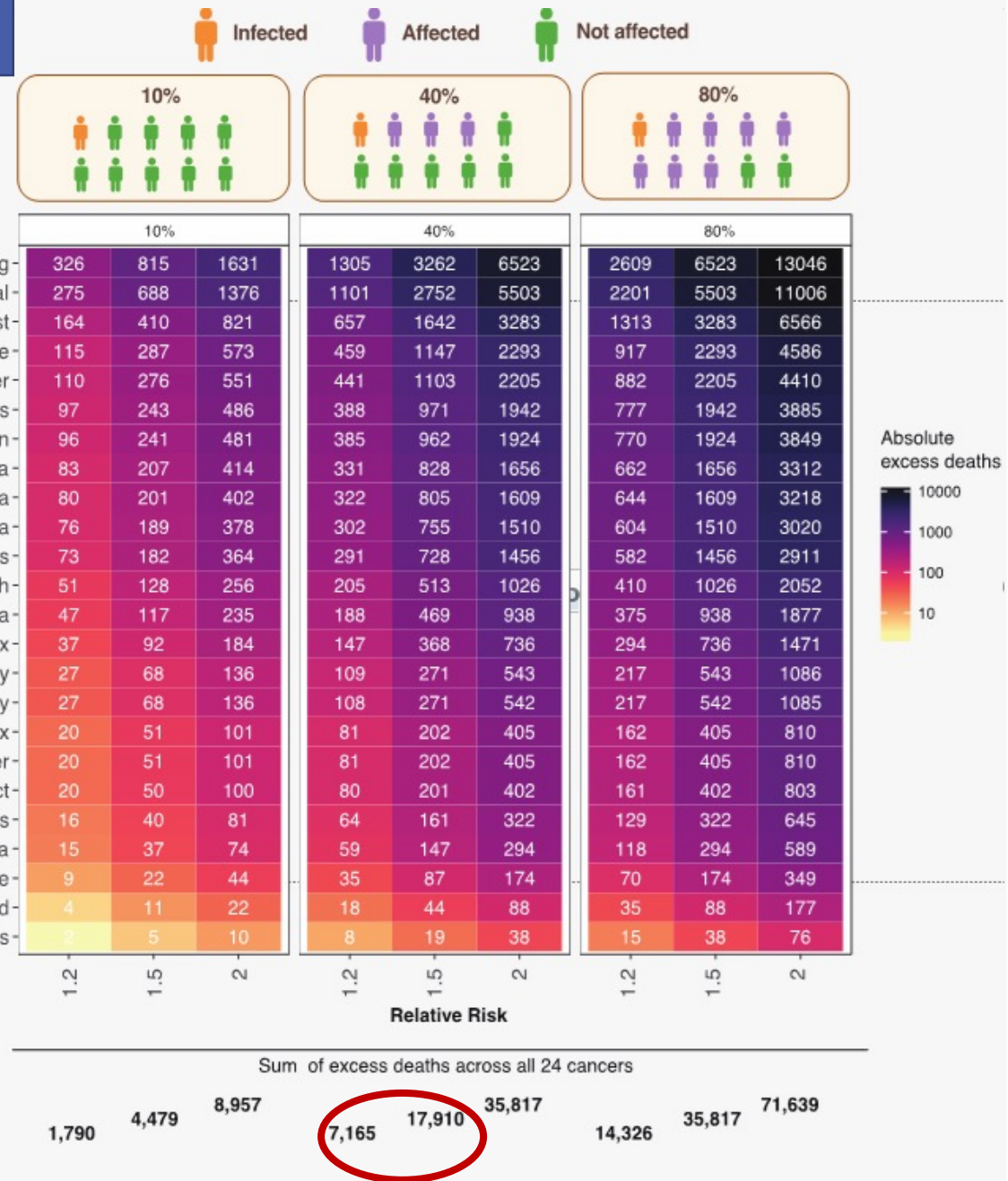


Sud A, Torr B, Loveday C, Jones M, Broggio J, ... McFerran E, Lawler M, Houlston R, Turnbull C. COVID-19 Lockdown and its impact on the two-week wait pathway for suspected cancer (*Lancet Oncology* 2020)

<https://www.carnallfarrar.com/life-sciences/life-sciences-insights/disruption-and-recovery-of-cancer-from-covid-19/>

Covid-19 and Cancer: Excess Mortality

- Looked at **excess mortality** on data from nearly 4M citizens through a **primary/secondary care linked data set**
- **Modelled different scenarios to determine excess deaths due to pandemic**
- **Used Real Time Data from Hospital Trusts to Inform the Model**
- **Significant predicted excess deaths (England) 7,165 – 17,910 excess deaths**

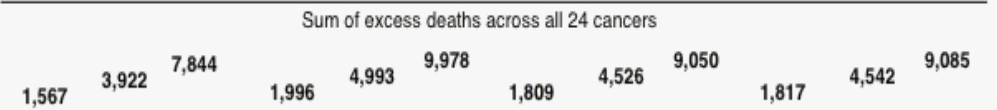
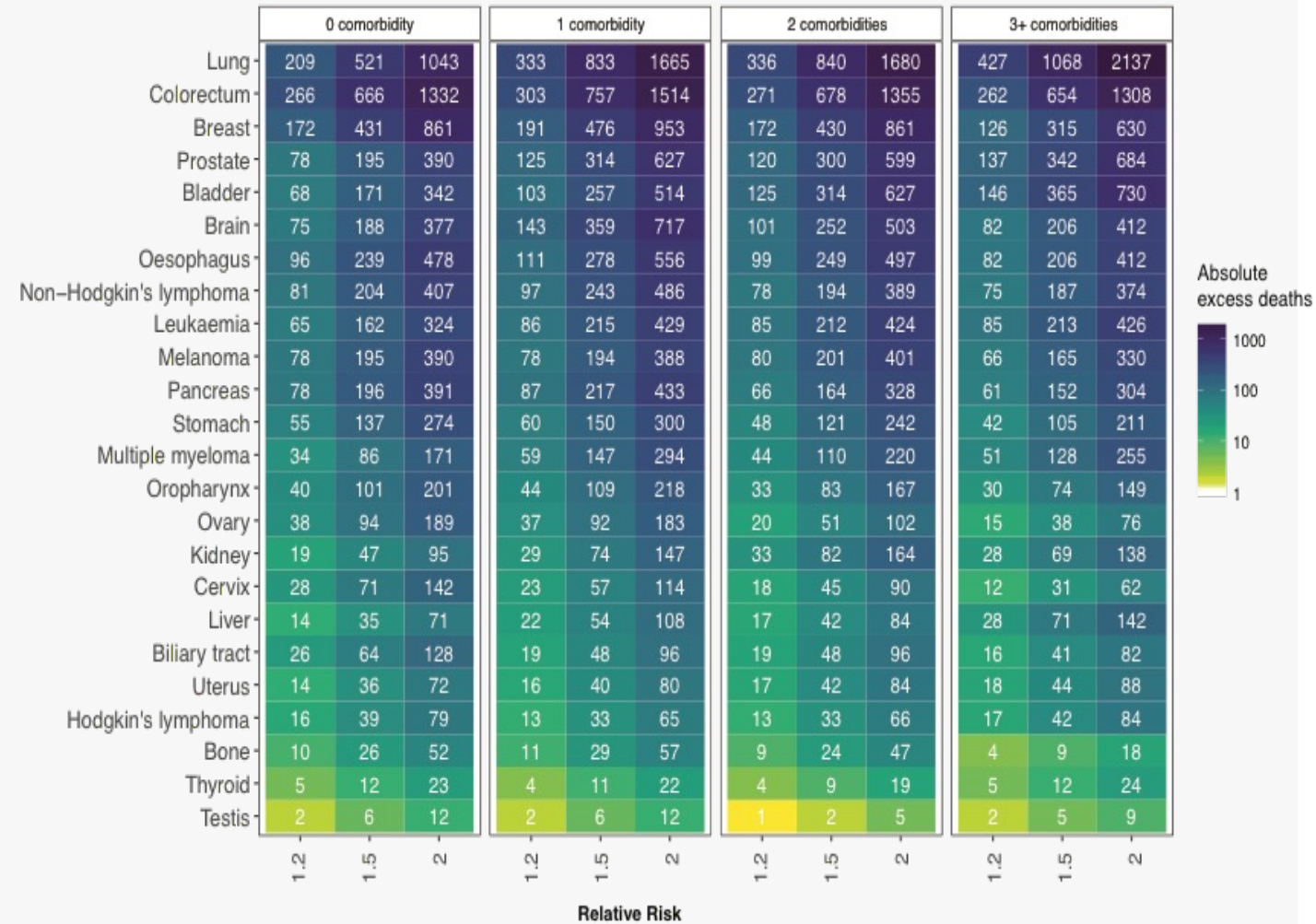


Lai AG, Pasa L, Banerjee A, Hall G, Denaxas S, Chang WH, Katsoulis M, Williams B, Pillay D, Noursadeghi M, Linch D, Hughes D, Forster MD, Turnbull C, Fitzpatrick NK, Boyd K, Foster GR, Enver T, Nafilyan V, Humberstone B, Neal RD, Cooper M, Jones M, Pritchard-Jones K, Sullivan R, Davie C, Lawler M, Hemingway H *BMJ Open*. 2020;

Covid-19 and Cancer: Excess Mortality and Comorbidities

- **Additional risk of excess deaths with at least one additional underlying health condition (comorbidities)**

- **Cardiovascular Disease**
- **Hypertension**
- **Diabetes**
- **COPD**



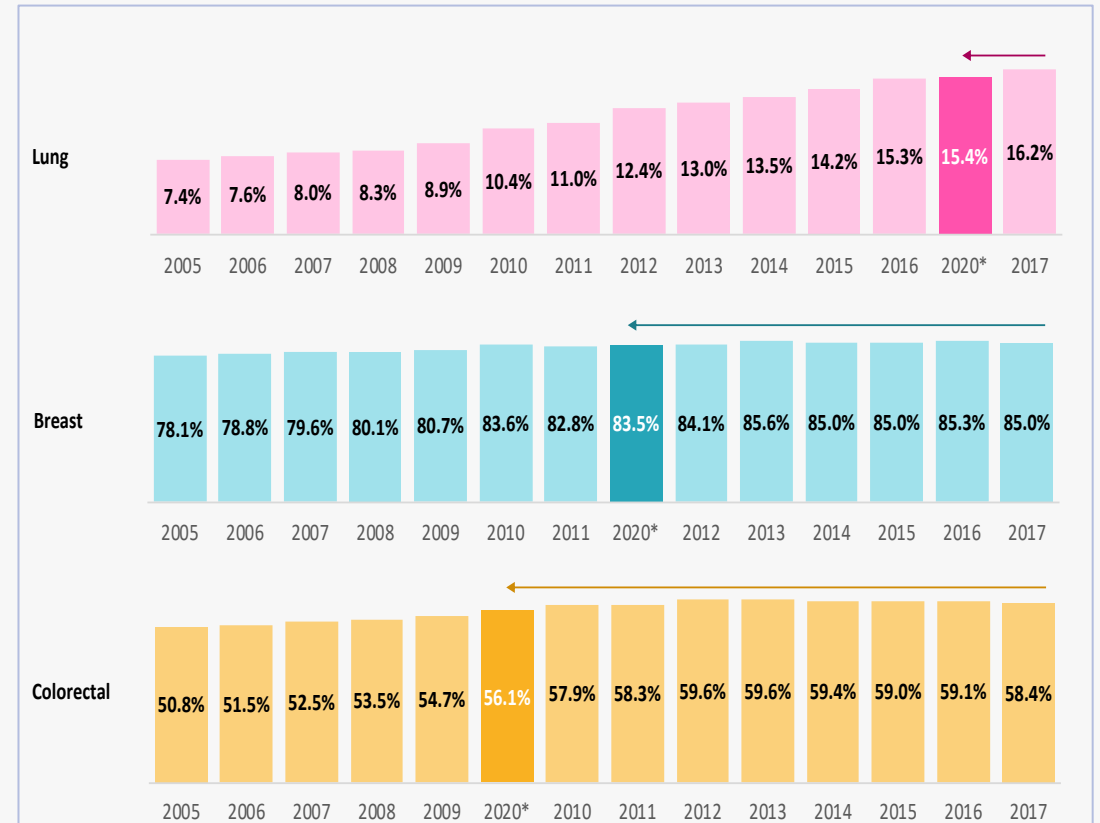
Lai AG, Pasa L, Banerjee A, Hall G, Denaxas S, Chang WH, Katsoulis M, Williams B, Pillay D, Noursadeghi M, Linch D, Hughes D, Forster MD, Turnbull C, Fitzpatrick NK, Boyd K, Foster GR, Enver T, Nafilyan V, Humberstone B, Neal RD, Cooper M, Jones M, Pritchard-Jones K, Sullivan R, Davie C, Lawler M, Hemingway H *BMJ Open*. 2020;

One Step Forwards ... Two Steps Backwards

- Following months of disruption, **stage shift** in detection of cancer i.e. detecting cancer later rather than earlier due to disruptions to the cancer pathway **significantly affected five-year survival**
- For certain cancers e.g. colorectal the **COVID-19 pandemic set us back nearly a decade**

<https://www.carnallfarrar.com/life-sciences/life-sciences-insights/disruption-and-recovery-of-cancer-from-covid-19/>

Five-year survival anticipated (England)



SOURCE: CF analysis, Office for National Statistics, Cancer survival in England, adults diagnosed
* Anticipated 5-year survival for 2020, calculated from 2017 figures

Covid-19 and Cancer: Influence and Impact

- **Results shared pre-submission with governmental/NHS stakeholders including all four CMOs, National Cancer Director and SAGE, contributed to decision to start restoring cancer services**
- **Association of British Pharmaceutical Industries: DATA-CAN should be supported to help “restore standards of care in cancer.”**
- **Research Community – Over 24,000 hits on ResearchGate and 127 citations in further research**
- **Media/Public: Main story Guardian, Telegraph; significant coverage Daily Mail, Financial Times, Sun, Sunday Independent, Wall Street Journal, others including print/online in France, Germany and Spain**
- **Lead story on Sky News; Interviews with BBC, ITV, Channel 4, NBC News, SBS News, multiple local radio**
- **Media Question to Prime Minister on his first press briefing post return from COVID-19 hospitalisation**
- **Featured prominently in [BBC Panorama](#) and [BBC Spotlight](#) programmes on Cancer and COVID**
- **Quoted extensively in report from [Carnall Farrar](#) and the [Institute for Public Policy Research](#)**
- **Won the [Royal College of Physicians \(RCP\)](#) research award [#EPCA2021](#) for original research that has contributed to significant improvements in health outcomes or the quality of patient care**

The impact of COVID-19 on cancer - Analysis

Enumerating the impact of COVID-19 on cancer pathways: a robust evaluation of the NHS England Secure Data Environment (SDE)

DATA-CAN delineated the precise impact of the COVID 19 pandemic on cancer systems and cancer patients. This required access to both historical data (pre-2020) and near real-time data on patients referred with (i) a suspicion of cancer and (ii) those diagnosed with and/or managed for cancer.

The emphasis was on data on cancer referral, cancer diagnosis, cancer treatment and cancer outcomes, examining the impact of age, ethnicity, deprivation, comorbidity and concurrent medication.

WP 2.1 - Indirect impact of COVID-19 on cancer:

WP 2.2 - Influence/associations of cancer on COVID-19 outcomes (such as admissions to hospital, admission to ITU, mechanical ventilation and death):

WP 2.3 - Influence/associations of cancer risk factors on COVID-19 outcomes:

WP 2.4 - Influence/associations of cancer medications on COVID-19 outcomes:

WP 2.5 - Direct impact of COVID-19 disease on cancer disease occurrence, re-occurrence and outcomes in short, medium and long term:

These analyses required access to linked data from the personal demographic service, primary care, hospital emergency, inpatient and outpatient care, intensive care, registered deaths by cause, cancer registries and COVID- 19 laboratory testing. All data are accessed by named, approved researchers in the Cancer SDE within NHS England.

Utilising the power of SNOMED CT to evidence the impact of the pandemic on cancer patients

HDRUK and DATA-CAN used live and historic comorbidity data coded using SNOMED CT to highlight the effect of the COVID-19 pandemic on cancer wait times, treatment attendance and outcomes

Lifesaving researchers: HDRUK and DATA-CAN

Health Data Research UK (HDRUK) is the national health data science institute, with a vision that *“every health and care interaction and research endeavour will be enhanced by access to large scale data and advanced analytics.”*

Previous projects have included using healthcare data to understand causes of deaths of homeless people, understanding suicide risk and identifying why ethnic minority groups were at greater risk from COVID-19.

DATA-CAN is HDRUK’s cancer hub; a unique partnership of NHS organisations, patients, charities, academic institutions, and industry working across all four nations of the United Kingdom. DATA-CAN works to use the nation’s healthcare data that is accessible to researchers and healthcare professionals working with cancer, to improve outcomes. The organisation also contributes to funded cancer research and produces quality cancer datasets for this purpose.



Chief Data Officer at University of Leeds and HDRUK Associate Director, Monica Jones

The impact of COVID-19 on cancer patients

The COVID-19 pandemic was an uncertain and challenging time for the entire world, and many patients across the health and care system were impacted directly or indirectly as a result. One cohort of patients who encountered many challenges across the health and care system were patients affected by cancer, including those with a suspected or existing diagnosis. These patients were identified as being vulnerable, and at a higher risk from having severe symptoms or dying from the virus.

Because of this, NHS cancer services were forced to adopt new operating practices, introduce enhanced safety measures, and review priorities, all of which significantly changed the patient experience.

“Using information recorded using SNOMED CT meant this critical comorbidity data was accessible in real-time from a range of care settings and sources. This information was then used as the basis to highlight how people with cancer were more likely to die from COVID-19, and how their interactions with lifesaving services were affected by the pandemic”

1. Cancer 2WW referrals - Dramatic fall
2. Cancer diagnoses - Significant fall
3. Mode of presentation - Increase emergency, decrease 2WW, loss screening
4. Stage at presentation - Stage shift (more advance disease)
5. Surgery - Decrease in number and complexity
6. Chemotherapy - Decrease in use, especially adjuvant, decrease intensity
7. Radiotherapy - Decease in use, lower numbers of fractions
8. Survival - Negative impact on 30-day, 90-day and 1, 2 & 5 year survival
9. PROMs - Not possible in real time
10. End of life - Shift from hospital and hospice to home

https://nhsengland.kahootz.com/gf2.ti/f/762498/196619525.1/PDF/-/SNOMED%20Implementation_Case%20Studies_HDRUK%20Case%20Study%20v3.0.pdf



Monica Jones FBCS CITP MInstLM

www.linkedin.com/in/monicacmjones

m.c.m.jones@leeds.ac.uk