North Central London
Health and Care
Integrated Care System

NHS
North Central London
Cancer Alliance

Use Cases to support improving data quality and example logic models

Why it matters to get coding right?

Current data capture

A lot of important patient information in GP systems is in unstructured data (free text) or in attachments (letters, reports). The data in this format cannot be processed and analysed via conventional data tools which are reliant on 'coded data'. The impact of this incomplete information ranges from direct care to research.

> *"My breast cancer family history wasn't known about and hence I wasn't referred to a genetics clinic"*
>
> *Primary Care Coding PPIE Workshop 17-Sep-24*

Summary of Problem Statements via END USER

Direct Care

- Restricts case-finding
- Unable to easily red flag patients who may be more at risk for cancer reappearance
- Difficult to find patients who should be on a pre-cancerous condition surveillance programme
- Cannot triage/prioritise those for diagnostics
- Struggle to identify those with a missing diagnosis and chase
- Leads to missed, delayed or inappropriate treatment

Service planning and service evaluation

- Under reporting of cancer cases
- Inaccurate estimates of cancer burden
- Ineffective cancer prevention
- Financial losses for healthcare providers
- Failure to service plan adequately
- Cannot evaluate/ monitor services

R&D/Innovation

- Limited data for prediction modelling
- Lack of evidence about the quality of primary care coding
- Restricted insights into the long term effects of treatment
- Cannot easily flag locally relevant and timely areas for research that will solve a local problem

> *"My cancer treatment (chemo) was not coded so I wasn't identified as high risk and in need of a 3rd covid vaccination."*
>
> *Primary Care Coding PPIE Workshop 17-Sep-24*

Implementation Strategies – to Improve coding

INTERVENTION/STRATEGY - STANDALONE OR COMBINATION		PROS	CONS
Communication	Publicise the coding standards and the rationale in the GP communication channels	Low cost Structures in place	Low rates of behaviour change
Education	Providing training and education for healthcare professionals on cancer coding and ensuring they are aware of the correct code to use in each context - from teaching packs, videos, manuals	Effective Cost defined Low complexity	Can be high cost Labour intensive
Training	Outreach, training inhouse, learning collaboratives, early adopters use, champion use	Effective	High cost Labour intensive Coordination and tracking intensive
Incentives	Work to incentivize the adoption and implementation	Coding quality increases	High cost Behaviour change not retained if incentive stopped
EPR technology (EMIS Web)	Utilise the EPR to support coding through – templates, recommending codes, drop down lists - to help facilitate accurate and consistent use of agreed clinical codes.	Ardens is funded in NCL already Works to an extent	Onboarding support needed Update support needed
Automation	Limited digital products that can work with existing systems to extract coded data across a defined set of use cases	Frees up admin/clin time	Can be clunky/not user friendly
AI	Adopting technology-enabled intervention strategies, such as machine learning applications (NLP)	Very effective Low impact on human workload Less reliant on human and practice behaviours and resources	Unknowns multiple as novel – IG, ownership, running costs, integration

USE CASE 1: Inequalities in cancer care provision

Status Quo

Significant (avoidable) health inequalities exist in the cancer early detection to cancer treatment and those living with cancer. Individuals from more deprived backgrounds, certain ethnic minority groups, or with lower levels of educational attainment experience greater cancer burden and worse outcomes. However, current electronic health records (EHRs) in primary care lack consistent and complete capture of key sociodemographic factors. While ethnicity coding is relatively complete (~86.5%), data on other critical determinants of health—such as carer status, employment, education level, deprivation, and language spoken—are often poorly recorded or missing.

Why is this a problem?

Without complete data:

- We cannot confidently identify or quantify high-risk groups experiencing cancer-related inequalities.
- Reasonable adjustments (e.g., translation services, out-of-hours appointments) cannot be reliably offered to those who need them.
- Screening uptake and early diagnosis efforts remain unequal and poorly targeted.

Opportunities with Better Data

If sociodemographic data were routinely and accurately captured:

- We could identify inequities in access, outcomes, and service use through meaningful analysis.
- Care could be tailored, personalised, and equitable—for example, through targeted outreach or culturally appropriate services.
- Resources could be allocated to close the gap in outcomes across populations.

Who Benefits from This Data?

- **Direct care providers:** GPs, practice nurses
- **System actors:** Integrated Care Boards (ICBs), Cancer Alliances, population health analysts, researchers

Where Is This Data Captured?

Primarily in **primary care** records (EHR systems).

EXAMPLE Logic Model: Improving Equity in Cancer Care Through Better Primary Care Coding

Inputs	Activities	Outputs	Short-term Outcomes	Medium-term Outcomes	Long-term Impact
- Primary care EHR systems (EMIS/SystmOne) - Trained primary care workforce (GPs, nurses, admin staff) - Stakeholder engagement (ICBs, Cancer Alliance, Health Inequality Leads) - Funding/incentives (QOF, enhanced services) - Standardised templates for sociodemographic data - IT support	- Develop and implement standardised templates - Train staff on data capture - Conduct audits and feedback loops - Update data regularly (e.g., at registration, annual reviews) - Use data for targeted outreach - Share use cases to demonstrate impact	- More complete coding of sociodemographic fields (e.g., carer status, employment, education, deprivation, language) - Routine data updates - Segmentation of populations by risk factors	- Better identification of high-risk groups - Ability to tailor screening and early detection programmes - Enable personalised interventions	- Increased screening uptake among underserved populations - Improved access to primary care for high-risk groups - More timely diagnoses in deprived communities	- Reduced inequalities in cancer outcomes (e.g., stage at diagnosis, survival) - Sustainable, equitable, personalised care pathways - Greater equity in cancer services across NCL and wider systems

USE CASE 2: Early cancer detection

Status Quo

Achieving the national ambition of diagnosing 75% of cancers at Stage 1 or 2 by 2028 is critical to improving prognosis and quality of life for patients. One major enabler of this goal is the ability to identify people at higher risk of cancer earlier and more accurately. However, many of the key risk factors that could support this—such as physical characteristics (e.g. BMI), health behaviours (e.g. smoking, alcohol intake), family history of cancer, known genetic mutations (e.g. Lynch syndrome, BRCA1/2), and environmental exposures—are inconsistently or poorly coded. Screening uptake is also inconsistently recorded.

Why This Is a Problem

In the absence of complete and accurate coding:

- High-risk individuals cannot be confidently identified.
- Risk stratification tools either cannot be applied, or they amplify existing data biases.
- Surveillance and preventative programmes are under-utilised for eligible high-risk patients.

Opportunities with Better Data

If risk factor data were systematically coded and updated:

- High-risk cohorts could be proactively identified and invited for earlier screening or surveillance.
- More accurate and equitable cancer risk stratification models could be developed and validated.
- Earlier detection and diagnosis would be more achievable, improving outcomes and reducing treatment burden.

Who Benefits from This Data?

- **Direct care providers:** GPs, nurses, cancer screening coordinators
- **System-level users:** ICB prevention leads, Cancer Alliances, public health teams, clinical researchers

Where Is This Data Captured?

Across both **primary** and **secondary care** electronic health records (EHRs).

EXAMPLE Logic Model: Improving Early Cancer Detection through Better Coding of Risk Factors

Inputs	Activities	Outputs	Short-term Outcomes	Intermediate Outcomes	Long-term Impact
- Primary and secondary care clinical systems - Engagement from primary care teams (GPs, nurses, admin staff) - Cancer Alliance support - Data quality improvement expertise - Risk stratification research expertise - IT infrastructure for coding templates and prompts	- Standardise coding templates for cancer risk factors in EHRs - Educate primary care teams on risk factor documentation - Integrate prompts into routine workflows - Link risk factor data to risk stratification tools - Conduct regular audits on data completeness - Co-develop risk stratification algorithms	- More complete recording of BMI, smoking, alcohol intake, family history, genetic risks, environmental exposures - Dashboards tracking data completeness and risk identification - Deployment of risk stratification tools - Increased referrals of high-risk patients to cancer surveillance	- Better identification of high-risk individuals in primary care - More accurate population-level cancer risk profiles - Enhanced targeting of early diagnosis initiatives	- Increased participation in cancer surveillance among high-risk patients - Earlier cancer detection in vulnerable cohorts - Reduced variation in referrals and diagnosis	- Progress toward 75% early-stage diagnosis target by 2028 - Improved prognosis and survival outcomes - Reduced inequalities in cancer detection and care - Stronger real-world data to inform preventive strategies

USE CASE 3: Cancer care in the community

Status Quo

A significant number of people are living with cancer, yet our understanding of the true cancer burden is limited by incomplete coding. In particular, documentation of cancer treatments (such as chemotherapy, surgery, radiotherapy, immunotherapy), treatment summaries, surveillance protocols, and follow-up plans is very sparse in the records. This gap hinders effective service planning for cancer survivors.

Why This Is a Problem

Without comprehensive treatment data:

- The long-term health impacts of cancer therapies remain obscured.
- Linkages between previous cancer treatments and current health issues are missed, leading to ineffective management.
- Patients eligible for appropriate follow-up tests, immunisations, or interventions may be overlooked, compromising care and safety.

Opportunities with Better Data

Enhanced coding and documentation of cancer treatments would enable:

- Improved identification of patients for targeted follow-up care and surveillance.
- More accurate risk stratification for late effects of treatment.
- Tailored, timely interventions that boost patient safety and overall care quality.

Who Benefits from This Data?

- **Direct Care Providers:** GPs, nurses, care coordinators
- **Service Improvement:** ICB teams, Cancer Alliances, and researchers

Where Is This Data Captured?

Data is captured in both **primary and secondary care** electronic health records.

EXAMPLE Logic Model: Improving Management of People Living with Cancer through Better Treatment and Follow-up Coding

Inputs	Activities	Outputs	Short-term Outcomes	Intermediate Outcomes	Long-term Impact
- Primary and secondary care clinical systems - Engagement from clinical teams (oncologists, GPs, nurses, secondary care coders) - Cancer Alliance leadership - Access to discharge summaries and end-of-treatment documents - IT infrastructure for structured treatment and follow-up coding - Funding and commissioning support for survivorship pathways	- Standardise templates for coding treatments and follow-up - Educate staff on importance of structured treatment history - Embed structured fields in EHRs - Improve transmission of treatment summaries to primary care - Audit coding completeness regularly - Automate follow-up flags based on treatment history	- More complete coding of cancer treatments and follow-up plans - Reliable transfer of treatment summaries across settings - Dashboards tracking survivorship and care needs - Data linkage between treatment history and ongoing care	- Primary care teams can identify patients' prior cancer treatments - Enhanced clinical linkage between past treatments and current symptoms - Better scheduling of appropriate follow-up tests	- Safer, personalised management for cancer survivors - Fewer missed secondary prevention opportunities - Increased proactive care for late effects of treatment	- Better patient safety, quality of life, and outcomes for cancer survivors - More efficient and equitable service planning - Stronger living with and beyond cancer care models aligned with NHS cancer strategies